

Tensors, Differential Forms, and Computer Graphics

ROSS GROGAN-KAYLOR

September 2020

Acknowledgements

- A special word of thanks goes to Josh Davis for supervising my independent study in differential forms.

To do

Checklist for turn in

use “assume the hypotheses of the previous theorem”

ToC

- included unnumbered subsections

Bibliography

- credit Mark where necessary
- mention I independently filled out the details of lot of things after reading brief Wikipedia expositions
- add a note for how each source was used

Defns

- Make sure “if” is not used in any definition, and that “iff” is always used.
- Use “... is said to be ...” for commonly used defns, and use “we say ... is ...” for defn particular to this book

Lin alg

- make sure to say “extend with linearity” when extending with seeming-multilinearity
- mention $\text{sgn}(\sigma)$ well defined
- ~~make sure that this has been done: replace “matrix relative to E and E ” with “matrix relative to E ”~~
- check that there’s enough stars in lin alg. add stars to other review sections?
- ~~fix lin alg injective iff surjective. only works when dimensions the same. look at relevant thms about linear isos, main dim thm, and fix~~
- remove finite-dimensionality from motivated intro and other sections when unnecessary
- make sure “bilinear” said instead of “multilinear” when applicable
- new style: write overall hypotheses for each section. add redundant hypotheses for each theorem in small text. don’t worry about this till ends
- use “has as a basis” instead of “and let E be a basis for V ”?
- say “over the same field” when no field is explicitly mentioned, when necessary
- add matrix addition, etc.

- Do I emphasize the equivalence between primitive matrix and the characterizing property of $[\mathbf{f}(E)]_F$ enough? Maybe a remark on these two ways to interpret things would be good.
- Add direct sums and $\dim(W_1 + W_2)$ theorem to lin alg. Look in Halmos for good proof on direct sum condition.

Proofs

- add $\text{lemma} =$ to proofs
- use ; in thm/defn parentheticals instead of “and”
- no theorem text should be on the same line as any theorem parenthetical title
- read through and make one-to-one and onto vs. injective and surjective consistent
- make sure injectivity and surjectivity checked for all isos.

$\binom{p}{q}$ tensors

- go through and circle things that don’t need finite-dimensionality. then decide how to handle finite-dimensionality: is it best to say V is finite-dimensional in def or not?
- make sure things of the form $\cup_{p,q \in \mathbb{N}} T_q^p(V)$ are used instead of $T_q^p(V)$
- ~~use ϵ^i for arb basis, use ϕ^{e_i} for induced dual basis~~
- check that induced dual bases not assumed where unnecessary in bilinear forms section
- When defining maps on elementary tensors, is the elementary tensor of the form $\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_p \otimes \phi^1 \otimes \dots \otimes \phi^q$?

Exterior powers

- Fix “we won’t define alternating tensor algebra” vs. actually doing so
- Replace alternating tensors with antisymmetric tensors, and add a remark

Misc.

- decide on itemize vs. enumerate
- every \mapsto should have a function over it
- add more subsections, and explanatory text in-between theorems in these subsections
- add text in-between proofs regardless!
- make sure sum parenthetical conventions are consistent
- Are all instances of lists such as $x_1 \dots x_r \dots x_n$ replaced with $x_1 \dots \cancel{x_r} \dots x_n$?
- make sure $\{1, \dots, n\} - i$ is replaced with $\{1, \dots, n\} - \{i\}$ in sum subscripts
- standardize the way I type “pull back”, “pullback”, “pull-back”
- emphasize “for all” as encompassing entire proof, like brackets in programming.
- replace $\widetilde{\mathbf{u}}_i$ with $\widetilde{\mathbf{u}}_i$

Calc

- decide $\nabla_{\mathbf{x}}$ vs. ∇ ; generalization of Leibniz vs. Newton
- double check evaluation conventions such as $\mathbf{x} = \mathbf{x}_0$ and $t = t_0$; use $\mathbf{x} = \mathbf{p}$ instead of $\mathbf{x} = \mathbf{x}_0$
- add technical hypotheses to calc chapter
- proof FTC via Hubbard

Notation

- revamp notation page
- Might delete $\hat{\mathfrak{F}}$ from notation. Also make note about Russian letter, and how I don't actually ever talk about elements of W^{**}
- Define set of permutations S_n somewhere

Manifolds and diff forms

- “ $T_{\mathbf{p}}(\mathbb{R}) \cong \mathbb{R}$ ” and “ $T_{\mathbf{F}(\mathbf{p})}(\mathbb{R})$ ” should be “ $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ ”
- don't mention dimension of manifold when unnecessary
- replace $d\mathbf{x}_{\mathbf{q}}^{-1}$ with $d\mathbf{x}^{-1}|_{\mathbf{q}}$
- ~~replace lower case \mathbf{f} with capital \mathbf{F}~~
- finish up Lee citations
- check theorems against Lee for hypotheses
 - charts need to be: smooth, maybe oriented
- use “WWBOC”, add remark for this convention
- use the word “tangent vector” more
- finish up Lee citations
- when using $\pm \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega)$, do we handle the negative case when necessary? it might be that accidentally I only mention the positive case sometimes
- replace “pairwise nonoverlapping” with “boundaries have measure zero and are pairwise disjoint”, or something like that- look at parameterization thm in Lee to find out what to say

About this book

Main goals

- Stokes' theorem
- Interpretations that go along with differential forms
 - x^i as coordinate function
 - $\frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$ as a tangent vector, $\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} = \frac{\partial}{\partial \mathbf{e}^i} \Big|_{\mathbf{p}}$ as a basis tangent vector
 - algebraic meaning for placeholder in integrand

Tentative prerequisites and reading advice

This book is primarily written for a reader who has experience with the following:

- the content of typical three-course calculus sequence: single-variable differential calculus, single-variable integral calculus, and multivariable calculus (but *not* differential equations),
- introductory linear algebra
- introductory logic and proof writing

A dedicated reader who has only taken the three-course calculus sequence mentioned above can still understand everything in this book with a bit of extra effort. Such a reader should skim Ch. 1 and then master Ch. 2, the review chapter on linear algebra. My advice is to use Ch. 2 as a guide for learning the core theory and to consult an introductory linear algebra textbook, such as any edition of Otto Bretscher's linear algebra book (look it up online) for concrete examples. Two linear algebra textbooks written for a more advanced level are Halmos's *Finite Dimensional Vector Spaces* and Curtis's "introductory" linear algebra book. Be warned: I have found no linear algebra book that satisfactorily explains the matrix with respect to bases of a linear function, matrix-vector products, or matrix-matrix products; even theoretical treatments miss the mark by focusing on the fact that linear functions correspond to matrices (rather than focusing on why this correspondence happens). For these concepts, consult Ch. 2; and be wary when reading about them in other books.

There are two review-style chapters of this book: one on linear algebra and one on calculus. (The chapter on topology could also be considered to be a review chapter, but, as was stated above, I assume the reader has no knowledge of topology). For reasons expanded upon below, the content in the linear algebra review chapter is almost constantly applied throughout this book, as the new ideas of tensors and differential forms are really reorganizations of mathematical structure, and are therefore mostly algebraic. *You should read this chapter even if you have taken introductory linear algebra before!*

On the prominence of algebraic structure

Tensors are result of investigating, generalizing, and reorganizing various abstract algebraic ideas about linear functions. So it is not too surprising that algebraic strategies (like constantly being on the look-out for linear isomorphisms) dominate the theory of tensors.

On the other hand, one might be surprised that similar algebraic lines of thought dominate the study of differential forms. After all, differential forms are supposed to be about calculus- which is about measuring change and accumulating change and smooth surfaces- not algebra, right?

Well, differential forms generalize and reorganize ideas about the structure of calculus. Since differential forms are primarily about reorganization and structure, the content the reader does not yet know is algebraic. However, there is a better reason for the prominence of algebra in the study of differential forms: calculus is really about *local* linear algebra on the “tangent space” (think tangent plane) of an arbitrary point on the surface of interest. Due to this, we will in fact see that a differential form evaluated at a point is actually a special type of tensor.

Other

Theme of algebra: “if it it’s helpful to think of something in such and such way or is helpful to use such and such mnemonic, formalize it! This will generate new insights”

This book uses this maxim as a guiding principle. (e.g. how we approach matrices, doing things in terms of abstract tensors, exterior derivative in analogy to div)

Contents

1	Review of logic, proofs, and functions	13
1.1	Propositions and predicates	13
1.2	Quantificational logic	14
1.3	Implications with sets	15
1.4	Sets	19
1.5	Relations	20
2	Review of linear algebra	23
2.1	Vector spaces, span, and linear independence	23
2.2	Linear functions	27
2.3	The dot product and cross product	38
I	Multilinear algebra and tensors	45
3	A motivated introduction to tensors	49
3.1	Multilinear functions and tensor product spaces	49
3.2	A motivated introduction to $\binom{p}{q}$ tensors	52
3.3	Introduction to dual spaces	55
3.4	$\binom{p}{q}$ tensors	59
4	Bilinear forms, metric tensors, and coordinates of tensors	63
4.1	Bilinear forms and metric tensors	63
4.2	Coordinates of $\binom{p}{q}$ tensors	68
5	Exterior powers, the determinant, and orientation	79
5.1	Exterior powers	79
5.2	The determinant	85
5.3	Orientation of finite-dimensional vector spaces	91
5.4	The cross product	96
II	Calculus and basic topology	97
6	Review of calculus	99
6.1	Single-variable calculus	99
6.2	Multivariable calculus	101
6.3	Technical theorems	104
7	Basic topology	105
7.1	Topological spaces	107
7.2	Continuous functions, homeomorphisms, and compactness	111

III	Differential forms	113
8	Manifolds	115
8.1	Introduction to manifolds	115
8.2	Coordinatizing manifolds	116
8.3	Smooth manifolds	117
8.4	Tangent vectors	119
8.5	Tangent vectors and tangent covectors with coordinates	126
8.6	Vector fields and covector fields	128
8.7	Oriented manifolds and their oriented boundaries	130
8.8	Riemannian manifolds	131
9	Differential forms on manifolds	135
9.1	Differential forms	135
9.2	Integration of differential forms on manifolds	135
9.3	The exterior derivative	142
9.4	The generalized Stokes' theorem	146
9.5	Etc.	149
IV	Computational applications of differential forms	151
10	Discrete differential geometry and computer graphics	153
10.1	Discrete differential geometry	153
10.2	Smoothing with the Laplacian	156
10.3	Surface parameterization	162
V	Appendix	165
11	Different definitions of tensor	167
12	Tensors in physics and engineering	169
13	Review of pedagogy	173

Notation

Here is a list of most of the notation used in this book. Since the concepts that the notation has been designed around have not been introduced yet, do not worry about fully understanding this page on a first read-through. This page will be more helpful later.

- cross out notating
- V and W are vector spaces over a field K . When these spaces are finite-dimensional, we set $\dim(V) = n$ and $\dim(W) = m$.
- \cong is used to denote an isomorphism of vector spaces.
- \mathbf{v} is used for an element of the vector space V , and \mathbf{w} is used for an element of the vector space W or for another element of V .
- $E = \{\mathbf{e}_i\}_{i=1}^n$ is an arbitrary basis for V , and $F = \{\mathbf{f}_i\}_{i=1}^m$ is either another arbitrary basis for V or is an arbitrary basis for W .
- $U = \{\mathbf{u}_i\}_{i=1}^n$ is an orthonormal basis for V .
- $[\mathbf{v}]_E$ denotes the vector that contains the coordinates of $\mathbf{v} \in V$ relative to the basis E .
- $\hat{\mathbf{e}} = \{\hat{\mathbf{e}}_i\}_{i=1}^n$ is the standard basis of K^n (that is, $\hat{\mathbf{e}}_i$ is the tuple of entries from K whose j th entry is 1 when $j = i$ and 0 otherwise), and $\hat{\mathbf{f}} = \{\hat{\mathbf{f}}_i\}_{i=1}^m$ is the standard basis K^m (defined similarly).
- ϕ is an arbitrary element of V^* , and ψ is an arbitrary element of W^* .
- $E^* = \{\epsilon^i\}_{i=1}^n$ is the dual basis to E for V^* , and $F^* = \{\delta^i\}_{i=1}^m$ is either the dual basis to F for W^* or is another dual basis for V^* .
- δ_j^i is the Kronecker delta function defined by $\delta_j^i = 1$ when $i = j$ and $\delta_j^i = 0$ otherwise.
- Φ is an element of V^{**} , and Ψ is an element of W^{**} .
- $E^{**} = \{\Upsilon_i\}_{i=1}^n$ is an arbitrary basis of V^{**} , and $\{\Xi_i\}_{i=1}^m$ is an arbitrary basis of W^{**} .
- \mathbf{f} is a linear function $V \rightarrow W$, and \mathbf{A} is the $m \times n$ matrix of \mathbf{f} with respect to the bases E and F of the finite-dimensional vector spaces V and W .
- $\mathcal{L}(V_1 \times \dots \times V_k \rightarrow W)$ is the vector space of k -linear functions $V_1 \times \dots \times V_k \rightarrow W$. In particular, $\mathcal{L}(V \rightarrow W)$ is the vector space of linear functions $V \rightarrow W$.
- The set of $\binom{p}{q}$ tensors on V is denoted $T_q^p(V)$, and \mathbf{T} is used to denote an element of $T_q^p(V)$.
- The k th exterior power of V is denoted $\Lambda^k(V)$.
- M and N are manifolds.
- The set of differential k -forms on a manifold is denoted $\Omega^k(M)$, and ω is typically used to denote an element of $\Omega^k(M)$.

Covariance and contravariance. As seen above, bolded symbols are used to denote elements of vector spaces (except for elements of V^* , W^* , V^{**} and W^{**}) as well as functions that return elements of vector spaces as output.

Also, upper indices are used for elements of V , V^{**} , W , W^{**} , while lower indices are used for elements of V^* , W^* . The reason for this becomes clear [...] later.

Notation in definitions. The notation $:=$ is used to indicate a definition (this is different than $=$, which indicates an equality obtained through logical reasoning). In definitions, “if and only if” is abbreviated as “iff”.

1

Review of logic, proofs, and functions

1.1 Propositions and predicates

A *proposition* is a statement that is either true or false; an example is “the sky is blue right now”. A *predicate* is a proposition in which variables are used to denote one or more of the entities involved. For example, if S represents “the sky”, then the previous proposition can be turned into the predicate “ S is blue right now”. The distinction between proposition and predicate is not important to remember in practice. The most important idea introduced by both concepts is the idea of formalizing the idea of “true” and “false” statements.

We can additionally denote predicates themselves with letters, and say things such as “let P be a predicate”. When a predicate P is true, we write $P \cong T$. The \cong symbol denotes *logical equality*, and the T denotes “truth”. Similarly, we write $P \cong F$ when P is false.

Logical operators

More complicated predicates can be constructed from simpler predicates. Examples of some more complicated predicates are $((3 > 4) \text{ and } (\text{every rectangle is a square})) \cong F$ and $((5 > -3) \text{ or } (2 > 100)) \cong T$.

There are three fundamental operations on predicates that are used to build more complicated predicates from simpler ones: there are the two-argument (binary) operators *and* and *or* and the one-argument (unary) operator *not*.

The operators *and*, *or*, and *not* act on predicates P and Q as is expressed in the following “truth tables”.

P	Q	$P \text{ and } Q$	P	Q	$P \text{ or } Q$	P	$\text{not } P$
T	T	T	T	T	T	T	F
T	F	F	T	F	T	T	F
F	T	F	F	T	T	F	T
F	F	F	F	F	F		

So, $P \text{ and } Q$ evaluates to true only when both P and Q are true, $P \text{ or } Q$ evaluates to true whenever either of P or Q is true, and $\text{not } P$ evaluates to the “opposite” of P . Note that *or* is not the same as *exclusive or*, also called *xor*, which evaluates to true whenever exactly one of P, Q is true.

By looking at the above truth tables, you can surmise that, for predicates P and Q , we have the following two logical identities, called *DeMorgan’s laws*:

$$\begin{aligned}\text{not } (P \text{ and } Q) &\cong (\text{not } P) \text{ or } (\text{not } Q) \\ \text{not } (P \text{ or } Q) &\cong (\text{not } P) \text{ and } (\text{not } Q).\end{aligned}$$

Sometimes, symbols are used to represent *and*, *or*, and *not*: \wedge denotes *and*, \vee denotes *or*, and either \sim or \neg denotes *not*. We will not use these symbols.

1.2 Quantificational logic

A *predicate-valued function* is a function that returns a different predicate for each input value. (We have not formally defined functions yet). For example, $P(x) = (x > 3)$ is a predicate-valued function which is true for some values of x but false for others. We will sometimes informally refer to predicate-valued functions as *properties*.

Quantifiers

The *universal quantifier* is the symbol \forall ; we read “ $\forall x$ ” as “for all x ”. The *existential quantifier* is the symbol \exists and is read as “there exists”; we read “ $\exists x$ ” as “there exists x such that”. The quantifiers \forall and \exists are used in the following way to create predicates from predicate-valued functions:

$$\begin{aligned}\forall x P(x) & \text{ is the predicate which states “for all } x, P(x) \cong T\text{”} \\ \exists x Q(x) & \text{ is the predicate which states “there exists an } x \text{ such that } Q(x) \cong T\text{”}.\end{aligned}$$

The predicate $\forall x P(x)$ statement is true exactly when $P(x)$ is true for all x , and the predicate $\exists x Q(x)$ is true exactly when $Q(x)$ is true for one or more x . In this sense, \forall is similar to *and* and \exists is similar to *or*. Here’s an example: the predicate $(\forall x x > 3)$ is false, while the predicate $(\exists x x > 3)$ is true¹.

Nested quantifiers

Suppose we have a predicate-valued function that has two inputs, $P(x, y)$. In the last section, applying a quantifier-variable pair to a predicate-valued function produced a predicate. Now, since our predicate-valued function has two inputs, applying any of the four quantifier-variable pairs ($\forall x, \forall y, \exists x, \exists y$) to $P(x, y)$ produces a predicate-valued function. For example, we can define a predicate-valued function $Q(y) \cong \forall x P(x, y)$.

We can repeat this process to obtain a predicate involving *nested quantifiers*. Continuing the example above, we could consider $\forall y Q(y)$, which is the same as $\forall y (\forall x P(x, y))$.

Given a predicate-valued function P of two inputs, the four possible ways to “nest” quantifiers are as follows:

$$\begin{aligned}\forall x \forall y P(x, y) \\ \forall x \exists y P(x, y) \\ \exists x \forall y P(x, y) \\ \exists x \exists y P(x, y).\end{aligned}$$

Always remember that the innermost pair of quantifier with predicate-valued function is a predicate-valued function.

It’s useful to know that when two quantifier-variable pairs are nested and the quantifiers are the same, we have the following commutative property:

$$\begin{aligned}\forall x \forall y P(x, y) & \cong \forall y \forall x P(x, y) \\ \exists x \exists y Q(x, y) & \cong \exists y \exists x Q(x, y).\end{aligned}$$

There is a shorthand notation for situations in which we have two nested quantifiers of the same type:

$$\begin{aligned}\forall x, y P(x, y) & := \forall x \forall y P(x, y) \\ \exists x, y P(x, y) & := \exists x \exists y P(x, y).\end{aligned}$$

¹Technically, these statements aren’t really sensible since we haven’t specified that the x ’s involved are numbers.

Negating quantifiers

The *not* operator applies to predicates constructed with quantifiers as follows:

$$\text{not}(\forall x P(x)) \cong \exists x \text{ not } P(x)$$

$$\text{not}(\exists x P(x)) \cong \forall x \text{ not } P(x).$$

Intuitively, a property doesn't hold true for all objects exactly when that property doesn't hold for one or more of the objects, and a property isn't true for one or more of the objects exactly when it isn't true for all objects.

Nested quantifiers can be negated with this rule, as well. To negate a nested quantifier, just treat the inner quantifier-predicate pair as a predicate-valued function so that the above rules apply. For example, $(\text{not } (\forall x \exists y P(x, y))) \cong (\exists x \forall y \text{ not } P(x, y))$.

Essentially, all of math is expressed using quantifiers and logical operators

We roughly define a *first-order mathematical theory* to consist of

- a list of *axioms*, or assumptions thought of as inherently true, that are expressible by using the quantifiers \forall and \exists on variables (such as x) in conjunction with predicate-valued functions and logical operators
- the collection of all predicates (“facts”) which are logically equivalent to the axioms.

The *Zermelo–Fraenkel set theory* with the *axiom of choice* (abbreviated ZFC, where the “C” is for “choice”) is a commonly used first-order mathematical theory. The axioms of ZFC are relatively complicated, and will not be stated here. The important point is that the axioms of ZFC are stated in accordance to the two bullet points above; they are stated completely in terms of quantifiers, predicate-valued functions, and logical operators derived from *and*, *or*, and *not*.

This may sound a bit esoteric. You may ask, “just how much can we say with ZFC?” The answer is: “a lot”. Essentially all of math (calculus, real analysis, probability, statistics, linear algebra, differential equations, abstract algebra, number theory, topology, differential geometry, etc.) can be expressed in terms of ZFC. Since physics, engineering, and the other sciences are built on top of math, then the math that got humans to the moon can be derived from ZFC.

How can ZFC (and similar theories) do all of this? The answer is by building up abstraction. While mathematical constructions may always reduce down to quantificational logic, we do not in practice explicitly deal in quantificational logic all the time. Instead, sophisticated ideas are expressed by defining mathematical objects using previously defined notions, thinking about these objects in intuitive terms while still keeping the rigorous definition in mind, and proving facts (theorems) about these objects. The two ideas that most fields of math are built upon are the those of *sets*, which are essentially lists, and *functions*, which haven't been formally introduced yet. When you have these two concepts, you can build pretty much any theory.

1.3 Implications with sets

Sets

A *set* is an unordered list of objects. Examples of sets include $S_1 = \{\text{grass}, \text{tree}, -1, \pi\}$, $S_2 = \{0, 2, 4, 6, \dots\}$ and $S_3 = \{0\}$. The *empty set* is the set which contains no objects, and is denoted \emptyset . Sets can contain finitely many objects or infinitely many. S_1 and S_3 are examples of finite sets, and S_2 is an example of an infinite set. Because the order of objects in a set is irrelevant, we have for example that $\{1, 2\} = \{2, 1\}$. Formally, the fact that the order of objects in a set doesn't matter is established by defining sets to be equivalent when they contain precisely the same objects.

Constructing sets

$\{x \mid P(x)\}$ denotes the set of objects x which satisfy the property $P(x)$. The “ \mid ” symbol can be read as “such that”; we read “ $\{x \mid P(x)\}$ ” out loud as “ x such that P of x ”.

When an object x is contained in a set S , we write $x \in S$; the symbol \in is called the *set membership symbol*. “ $x \in S$ ” is read out loud as “ x in S ”.

We define $\{x \in S \mid P(x)\}$ to denote the set $\{x \mid x \in S \text{ and } P(x)\}$. You would read “ $\{x \in S \mid P(x)\}$ ” out loud as “ x in S such that $P(x)$.”

Implications

We now define the *implication* operator \implies , which is another operation on predicates that produces a more complicated predicate. Many authors introduce \implies immediately after *or*, *and*, and *or*; we introduce it now because it is best understood in the context of “for all” statements that involve sets.

This operator is read out loud as “implies”, and is defined as follows:

$$P \implies Q \text{ :}\cong (\text{not } P) \text{ or } Q.$$

The argument before the \implies symbol is referred to as the *hypothesis*, and the argument after the \implies symbol is referred to as the *conclusion*. Here is the truth table for \implies .

P	Q	$P \implies Q$
T	T	T
T	F	F
F	T	T
F	F	T

Beware: just because \implies is read as “implies” does *not* mean that it functions in the way that you might expect. A better name for \implies would be “primitive implies”, since \implies is defined *for the purpose of being used inside a “for all” statement*. “For all” statements that involve \implies in the following way are what correspond to the English language meaning of “implies”:

$$\forall x \, x \in S \implies P(x).$$

We might express the above in English as “ x being in S implies $P(x)$ ”. The key difference between this informal sentence in English and the statement in mathematical notation is that the English sentence lacks a “for all” quantification. So, if the English sentence were truly correct, it would be “For all x , x being in S implies $P(x)$ ”. In the mathematical notation, we are technically considering *all* entities (such as “grass”, the function f defined by $f(x) = 3x^2, -1$, etc.) due to the “ $\forall x$ ”. We restrict our attention to the set S by using \implies , so that it is only possible for the predicate-valued function $(x \in S \implies P(x))$ to be false when $x \in S$.

The following thought experiment helps further illustrate the idea behind \implies . Suppose that you suspect that all squares are rectangles (you are right). To test your belief, you set out to test every single geometric shape, one by one. (Looking at every single geometric shape corresponds to the “ $\forall x$ ” part of the above line. You have no control over the fact that, in full formality, you are always considering *all* objects x). You set $P(x) \cong (x \text{ is a square})$ and $Q(x) \cong (x \text{ is a rectangle})$. First, you inspect a square, and determine that it is indeed a rectangle. (This corresponds to the first row of the truth table; $P(x)$ is true and $Q(x)$ is true for this particular x , so your theory holds, at least so far). Next, you look at a circle. You’re not interested in circles- you’re interested in squares! Knowing whether or not the circle is a rectangle is irrelevant. (This corresponds to the last two rows of the truth table; for any x , whenever $P(x)$ is false, then $P(x) \implies Q(x)$ is true). After an infinite amount of time, you have tested all squares and determined that they are all rectangles, so your theory stands;

that is, the predicate that is the “for all” statement evaluates to true. If there had been a single square that wasn’t a rectangle, then the predicate-valued function inside the “for all” would have evaluated as false for some x . (This corresponds to the second row of the truth table). This would make the entire “for all” statement false.

“Necessary” and “sufficient”

Let P and Q be propositions, and consider $P \implies Q$. Due to the reasoning “if Q happened then P must have happened”, P is said to be a *necessary condition* for Q and Q is said to be a *sufficient condition* for P . In my opinion, this language should only refer to “for all” statements such as $\forall x P(x) \implies Q(x)$, since the connotations of “necessary” and “sufficient” can mislead someone to misunderstand that \implies really only makes sense when used inside a “for all” statement.

Quantifiers with set membership

We define

$$\begin{aligned}\forall x \in S P(x) &:= \forall x x \in S \implies P(x) \\ \exists x \in S P(x) &:= \exists x \in S \text{ and } P(x).\end{aligned}$$

The first line was motivated in the previous section. The second line is probably easier to understand than the first, since it doesn’t involve \implies .

The negations of the above newly defined expressions are what you expect:

$$\begin{aligned}\text{not}(\forall x \in S P(x)) &\cong \exists x \in S \text{ not } P(x) \\ \text{not}(\exists x \in S P(x)) &\cong \forall x \in S \text{ not } P(x).\end{aligned}$$

This is because a slightly more general versions of the above facts hold. (To obtain the above from the below, substitute $P(x) = (x \in S)$ and $Q(x) = P(x)$ into the below).

$$\begin{aligned}\text{not}(\forall x P(x) \text{ and } Q(x)) &\cong \exists x P(x) \text{ and } (\text{not } Q(x)) \\ \text{not}(\exists x P(x) \text{ and } Q(x)) &\cong \forall x P(x) \text{ and } (\text{not } Q(x)).\end{aligned}$$

These more general facts are useful when objects with a particular property are being considered. For example, consider all triangles in the plane. Given a right triangle T , let a_T, b_T , and c_T be the lengths of the sides of T , with c_T being the length of the hypotenuse. The Pythagorean theorem states² ($\forall T$ T is a right triangle and $a_T^2 + b_T^2 = c_T^2$). (Note: this “for all” statement does *not* state that all triangles are right triangles satisfying the Pythagorean theorem! It states that all right triangles satisfy $a_T^2 + b_T^2 = c_T^2$). The negation of the Pythagorean theorem, using the above, is then ($\exists T$ T is a right triangle and $a_T^2 + b_T^2 \neq c_T^2$). Since the Pythagorean theorem is true, this means that it is *not* the case that there is a right triangle T for which $a_T^2 + b_T^2 \neq c_T^2$.

Here’s a proof of the first line of the more general statement; the proof of the second line is similar.

$$\text{not}(\forall x P(x) \text{ and } Q(x)) \cong \forall x \text{ not}(P(x) \text{ and } Q(x)) \cong \forall x \text{ not } P(x) \text{ or not } Q(x) \cong \forall x P(x) \text{ and not } Q(x).$$

The logical equality follows because, for propositions P and Q , we have $(P \text{ or } Q) \cong (P \text{ and } (\text{not } Q))$. This can be checked by truth table; it can also be understood intuitively: “if P or Q happens, and Q doesn’t happen, then P must happen”.

²Because the “for all statement” considers all triangles T , not just the right triangles, the definition of c_T does not make sense for non-right triangles, since non-right triangles don’t have a hypotenuse. We could improve the definition of c_T so it applies to all right triangles and is still the length of the hypotenuse when T is a right triangle. However, it is actually fine for c_T to be undefined for non-right triangles T , because an *and* statement evaluates to false whenever either argument (such as “ T is a right triangle”) is false. So, in the general case, allowing for this so-called *short-circuit* interpretation of *and* allows us to use predicate-valued functions $Q(x)$ that are only defined when $P(x)$ is true.

Common shorthand

- When people write something of the form “ $\forall x P(x) \implies Q(x)$ ”, they mean “ $(\forall x P(x) \implies Q(x)) \cong T$ ”.
- When people write something of the form “ $P(x) \implies Q(x)$ ”, they really should have written “ $\forall x P(x) \implies Q(x)$ ”. An extremely common example of this shorthand is “ $x \in S \implies P(x)$ ”.
 - This shorthand has a verbal equivalent: “If $P(x)$, then $Q(x)$ ”. The verbal equivalent is not considered bad notation, however.
- Combining both shorthand styles is extremely common in proof writing. You will often see a proof that contains a sentence of the form “ $P(x) \implies Q(x) \implies R(x)$ ”.

Converses

Given predicates P and Q , consider the implication $P \implies Q$. The *converse* to this implication is the implication $Q \implies P$. Please note that $(P \implies Q) \cong (Q \implies P)$ is a false statement! On the level of the English language, this means that $(\forall x P(x) \implies Q(x)) \cong (\forall x Q(x) \implies P(x))$ is a false statement. For example, “whenever it rains, there are clouds”, but it is not true that “whenever there are clouds, it rains”!

“If and only if”

We define one more operator on predicates, the *bidirectional implication* operator, denoted \iff . Given predicates P and Q , we define

$$P \iff Q := ((P \implies Q) \text{ and } (Q \implies P)).$$

Here’s the truth table for \iff .

P	Q	$P \iff Q$
T	T	T
T	F	F
F	T	F
F	F	T

The operator \iff is spoken aloud as “if and only if”; this is often abbreviated in writing as “iff”. In the context of the bidirectional implication $P \iff Q$, the implication $P \implies Q$ is referred to as the *forward implication* and its converse $Q \implies P$ is referred to as the *reverse implication*. Note that \iff is symmetric in the sense that $(P \iff Q) \cong (Q \iff P)$.

The English language interpretation of \iff is similar to to the English language interpretation of \implies : the mathematical version of what someone really means when they say “ $P(x)$ if and only if $Q(x)$ ” is $\forall x P(x) \iff Q(x)$. Again, we see the difference between full mathematical formalism and language is that language often omits the “for all”.

Many theorems in math state that a certain “if and only if” predicate is true. More specifically, such theorems usually state that some property of an object manifests if and only if another property of that object also manifests. So, when you see a theorem about an *equivalent condition* or an *equivalent definition*, there is an “if and only if” statement at play. All definitions in math are also automatically “if and only if” statements. There is, however, the common misleading convention of writing definitions using the word “if” (for example: “We say x has property $P(x)$ if $Q(x)$ ”; the “if” should really be an “iff”).

A neat fact is that \cong is the same operator on predicates as \iff . This is because \cong and \iff both have the same truth tables.

The contrapositive and proof by contradiction

For predicates P and Q , the following logical identity is true:

$$(P \implies Q) \cong ((\text{not } Q) \implies (\text{not } P)).$$

The right-hand side is called the *contrapositive* of the left-hand side.

You could verify this identity by using truth tables. Here is a nicer proof that uses the facts $(P \text{ or } Q) \cong (Q \text{ or } P)$ and $Q \cong \text{not}(\text{not } Q)$.

$$(P \implies Q) \cong ((\text{not } P) \text{ or } Q) \cong (\text{not}(\text{not } Q) \text{ or } (\text{not } P)) \cong ((\text{not } Q) \implies (\text{not } P)).$$

Proof by contradiction

Suppose we want to prove $P \implies Q$. One way to do so is to use *proof by contradiction*. Proof by contradiction proceeds as follows. Assume P is true, and suppose that Q is false. Then if, as a direct result of supposing Q to be false, we reach a logical impossibility, such as $1 = 0$, we know Q must be true.

Formally, proof by contradiction is an application of the contrapositive. The first step in a proof by contradiction of writing “assume P is true” serves no formal mathematical purpose, and is really just a reminder of the statement that will be contradicted. The next step, which is the first step that formally matters, is to assume $(\text{not } Q)$; this corresponds to the hypothesis of the contrapositive. Lastly, the contradiction (such as $1 = 0$) achieved at the end of the proof is actually always logically equivalent to $(\text{not } P)$ when the full context is considered. Proof by contradiction is just a more verbose way of proving the contrapositive, $((\text{not } Q) \implies (\text{not } P))$.

For some proofs, using the contrapositive in its raw form is most clear; for others, using the verbal format of proof by contradiction is more clear.

1.4 Sets

Set equality, subsets

Let S and T be sets. We define S and T to be *equal* iff $\forall x \ x \in S \iff x \in T$.

We say T is a *subset* of S iff $\forall x \in S \ x \in T$. When T is a subset of S , we write $T \subseteq S$. Note that, for all sets S , we have $\emptyset \subseteq S$ and $S \subseteq S$. When $T \subseteq S$ and $T \neq S$, we write $T \subsetneq S$. (Some authors write $T \subset S$ for this condition, but this is confusing because other authors write $T \subset S$ to mean $T \subseteq S$. Avoid the notation $T \subset S$).

Common sets

We define notation for many common infinite sets.

$$\mathbb{N} := \text{“the natural numbers”} = \{0, 1, 2, 3, \dots\}$$

$$\mathbb{Z} := \text{“the integers”} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

$$\mathbb{Q} := \text{“the rational numbers”} = \left\{ \frac{n}{m} \mid n, m \in \mathbb{Z} \right\}$$

$$\mathbb{R} := \text{“the real numbers”} = \{\text{all limits of sequences of rational numbers, roughly speaking}\}$$

$$\mathbb{C} := \text{“the complex numbers”} = \{a + b\sqrt{-1} \mid a, b \in \mathbb{R}\}$$

Indexing sets

Let S be a set. An *indexing set* of S is a set I that is thought of as “labeling” the elements S . Any set can be technically be used as an indexing set, though, we tend to use sets of numbers as indexing sets. The set of elements in S indexed by I is written as

$$\{x_\alpha \in S \mid \alpha \in I\}.$$

We use a Greek letter such as “ α ” when the size of the indexing set is unspecified; an indexing set can be either finite or infinite. When the indexing set is finite, we use a normal Roman letter such as “ i ”.

Union and intersection

We define the following operations on sets.

There is no universal set in ZFC

It seems natural that there would be a “set of all sets”; a set which contains every other set, but does not contain itself. However, this is not possible in ZFC. Suppose there does exist such a universal set U with $U \subsetneq U$. But since U contains all sets, U must contain itself, so $U \subseteq U$. This is a *contradiction*;

Other versions of set theory can handle

Set difference and set complement

DeMorgan’s laws for sets

Ordered pairs

Cartesian product

1.5 Relations

inverse relations

inverse functions \Leftrightarrow bijection

Equivalence relations

Definition 1.1. (Equivalence relation).

Definition 1.2. (Quotient set).

Functions

Latin meanings of injective and surjective.

Definition 1.3. (Function).

Functions are also commonly called *maps*, or *mappings*.

domain, codomain, range

Definition 1.4. (Uniqueness).

$\exists!x P(x) \cong (\exists x_0 P(x_0))$ and $(\forall x P(x) \implies x = x_0)$.

Remark 1.5. (Well-definedness and uniqueness).

What well-definedness refers to

Any property which is unique is the output of a well-defined function

Definition 1.6. (One-to-one, injection).

Definition 1.7. (Onto, surjection).

Definition 1.8. (Bijection).

left-inverse, right-inverse, relation to -jectivities

addition of fns, scaling of fns?, image-set

preimages (not neccess fns)

Definition 1.9. (Inverse function).

Theorem 1.10. (Invertible iff bijection).

Every one-to-one function has an invertible restriction.

Cardinality of sets

It was mentioned earlier that sets can be finite or infinite. We formally define

Theorem 1.11. (Left inverse and right inverse implies two-sided inverse). Let X and Y be sets, and let $f : X \rightarrow Y$ be a function. A *left-inverse* of f is a function $\ell : Y \rightarrow X$ such that $\ell \circ f = I_X$, where I_X is the identity on X . A *right-inverse* of f is a function $r : Y \rightarrow X$ such that $f \circ r = I_Y$, where I_Y is the identity on Y .

If $f : X \rightarrow Y$ has a left-inverse ℓ and a right-inverse r , then they must be equal, and we denote them by $f^{-1} := \ell = r$.

Proof. Let ℓ and r be left- and right- inverses of f , respectively. Then by associativity of function composition, $\ell \circ f \circ r = \ell \circ I_Y = I_X \circ r$. Therefore $\ell = r$. \square

Definition 1.12. (Proof by induction).

Defn?

2

Review of linear algebra

Advice for reading this chapter. This chapter is a comprehensive review of introductory linear algebra, minus “eigenstuff” and determinants. If you have previous experience with linear algebra, you might want to skim this chapter. To help with this, particularly important ideas are marked with stars: ★.

Pedagogy of this chapter. The approach we will use in relating linear functions to their corresponding matrices is a little unconventional, for the better! I have not seen the notion of “primitive matrix,” the definition of the matrix $[\mathbf{f}(E)]_F$, nor the definition of the function $\mathbf{f}_{E,F}$ presented elsewhere. However, I truly believe that the concepts are made clearer by the introduction of this notation.

This review chapter does not present “eigenstuff.” There are two core concepts taught in an introductory linear algebra class that do not appear in this chapter. The first of these core concepts, determinants, is treated extensively in Ch. 5. The second core concept, “eigenstuff,” is not treated because knowing it is not necessary for understanding the main content of this book. Traditional linear algebra texts explain “eigenstuff” quite well- consult one of the linear algebra texts mentioned in the preface (“About this book”) if you are interested.

Notation for covariance and contravariance is not used in this chapter. The use of both upper and lower indices to distinguish between “covariant” and “contravariant” will not be used in the following chapter of linear algebra review to prevent confusion. Only lower indices will be used. (If you don’t know what “covariant” or “contravariant” means, that is to be expected. Covariance and contravariance are explained later).

2.1 Vector spaces, span, and linear independence

Definition 2.1. (Field). Consider a tuple $(K, +, \cdot)$, where K is a set, $+: K \times K \rightarrow K$ is thought of as the “addition operation on K ,” and \cdot is thought of as the “multiplication operation on K .” We call $(K, +, \cdot)$ a *field* iff it satisfies the following:

1. (Requirements on $+$).
 - 1.1. (Closure under $+$). For all $c_1, c_2 \in K$, $c_1 + c_2 \in K$.
 - 1.2. (Existence of additive identity). There exists $0 \in K$ such that for all $c \in K$, $c + 0 = c$.
 - 1.3. (Closure under additive inverses). For all $c \in K$ there exists $-c \in K$ such that $c + (-c) = 0$.
 - 1.4. (Associativity of $+$). For all $c_1, c_2, c_3 \in K$, $(c_1 + c_2) + c_3 = c_1 + (c_2 + c_3)$.
 - 1.5. (Commutativity of $+$). For all $c_1, c_2 \in K$, $c_1 + c_2 = c_2 + c_1$.
2. (Requirements on \cdot).
 - 2.1. (Closure under \cdot). For all $c_1, c_2 \in K$, $c_1 \cdot c_2 \in K$.
 - 2.2. (Existence of multiplicative identity). There exists $1 \in K$ such that for all $k \in K$, $1 \cdot k = k = k \cdot 1$.
 - 2.3. (Associativity of \cdot). For all $c_1, c_2, c_3 \in K$, $(c_1 \cdot c_2) \cdot c_3 = c_1 \cdot (c_2 \cdot c_3)$.

2.4. (Closure under multiplicative inverses). For all $k \in K, k \neq 0$, there exists $\frac{1}{k} \in K$ such that $k \cdot \frac{1}{k} = 1 = \frac{1}{k} \cdot k$.

2.5. (Commutativity of \cdot). For all $c_1, c_2 \in K$, $c_1 \cdot c_2 = c_2 \cdot c_1$.

3. ($+$ distributes over \cdot). For all $c_1, c_2, c_3 \in K$, $(c_1 + c_2) \cdot c_3 = c_1 \cdot c_3 + c_2 \cdot c_3$.

(Equivalently, a field can be defined as an integral domain that is closed under multiplicative inverses, or as a commutative division ring).

In practice, we simply say that “ K is a field” to mean “ $(K, +, \cdot)$ is a field” when the definitions of the binary functions \cdot and $+$ are clear from context.

Remark 2.2. (Field). It’s not necessary to memorize all the conditions for a field. Just remember that a field is “a set in which one can add, subtract, multiply, and divide.” (Though, this doesn’t work when the field is finite).

Definition 2.3. (\star Vector space over a field \star). Consider a tuple $(V, K, +, \cdot)$, where V is a set, K is a field, $\cdot : K \times V \rightarrow V$ is thought of as the “scaling of a vector” operation, and $+$: $V \times V \rightarrow V$ is thought of as the “vector addition” operation. We say that $(V, K, +, \cdot)$ is a *vector space* iff \cdot and $+$ satisfy the following conditions:

1. $(V, +)$ is a *commutative group*. This means that conditions 1.1 through 1.5 must hold.

1.1. (Closure under $+$). For all $\mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{v}_1 + \mathbf{v}_2 \in V$.

1.2. (Existence of additive identity). There exists $\mathbf{0} \in V$ such that for all $\mathbf{v} \in V$, $\mathbf{v} + \mathbf{0} = \mathbf{v}$.

1.3. (Closure under additive inverses). For all $\mathbf{v} \in V$ there exists $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

1.4. (Associativity of $+$). For all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V$, $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$.

1.5. (Commutativity of $+$). For all $\mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$.

2. (Scalar-vector compatibility). For all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $c \in K$, $c(\mathbf{v}_1 + \mathbf{v}_2) = c\mathbf{v}_1 + c\mathbf{v}_2$.

3. (Scalar-vector compatibility). For all $\mathbf{v} \in V$ and $c_1, c_2 \in K$, $(c_1 + c_2)\mathbf{v} = c_1\mathbf{v} + c_2\mathbf{v}$.

4. (Scalar-vector compatibility). For all $\mathbf{v} \in V$ and $c_1, c_2 \in K$, $c_2(c_1\mathbf{v}) = c_2c_1\mathbf{v}$.

In practice, we say that “ V is a vector space over K ” to mean “ $(V, K, \cdot, +)$ is a vector space” when the definitions of the binary functions \cdot and $+$ are clear from context. We often also don’t refer to the field K , and just say “let V be a vector space.”

Elements of vector spaces are often called “vectors,” and elements of the field K are often called “scalars.”

Remark 2.4. (\emptyset is not a vector space). The empty set \emptyset is not a vector space over any field, because it contains no additive identity ($\mathbf{0}$).

Definition 2.5. (\star Vector subspace \star). If V, W are vector spaces over K and $W \subseteq V$, then W is a *vector subspace* of V .

Span and linear independence

Let V be a vector space over a field K , and consider $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$.

Definition 2.6. (\star Linear combination \star). A (*finite*) *linear combination* of $\mathbf{v}_1, \dots, \mathbf{v}_k$ is a vector $\mathbf{v} \in V$ of the form

$$\mathbf{v} = \sum_{i=1}^k c_i \mathbf{v}_i,$$

where the c_i ’s are some scalars in K .

Definition 2.7. (\star Span of vectors \star). We define the *span* of $\mathbf{v}_1, \dots, \mathbf{v}_k$ to be the set of all the finite linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_k$. That is,

$$\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k) := \left\{ \sum_{i=1}^k c_i \mathbf{v}_i \mid c_1, \dots, c_k \in K \right\}.$$

Definition 2.8. (\star Linear independence of vectors, intuitive version \star). When $k > 1$, we say that $\mathbf{v}_1, \dots, \mathbf{v}_k$ are *linearly independent* iff there is no \mathbf{v}_i , $i \in \{1, \dots, k\}$, contained in the span of any sublist of $\mathbf{v}_1, \dots, \mathbf{v}_k$.

Remark 2.9. (\star). The above definition does not apply to a “list” of just one vector, \mathbf{v}_1 .

Definition 2.10. (\star Linear independence of vectors \star). When $k \geq 1$, we say $\mathbf{v}_1, \dots, \mathbf{v}_k$ are *linearly dependent* iff there exist scalars $c_1, \dots, c_k \in K$ not all 0 such that

$$\sum_{i=1}^k c_i \mathbf{v}_i = \mathbf{0}.$$

We say $\mathbf{v}_1, \dots, \mathbf{v}_k$ are *linearly independent* iff they are not linearly dependent. That is, $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent iff the only choice of c_1, \dots, c_k for which the above equation holds is $c_1 = \dots = c_k = 0$.

Remark 2.11. (\star). With this new more general definition of linear independence, we see that a “list” of one vector, \mathbf{v}_1 , is linearly independent iff $\mathbf{v}_1 \neq \mathbf{0}$.

Theorem 2.12. (\star Equivalence of definitions of linear independence \star). When $k > 1$, the more recent definition of linear independence is equivalent to the intuitive definition of linear independence. (Proof left as exercise).

Bases and dimension of vector spaces

Definition 2.13. (\star Finite-dimensional vector space \star). A vector space V is *finite-dimensional* iff it is spanned by a finite set of vectors, and *infinite-dimensional* iff this is not the case.

Definition 2.14. (\star Basis, dimension of a vector space \star). Let V be a vector space. A set of vectors $E = \{\mathbf{e}_i \mid i \in I\}$ that both spans V and is linearly independent is called a *basis* of V . When V is finite-dimensional, then the *dimension* of V , denoted $\dim(V)$, is defined to be the number of basis vectors in a basis of V .

Definition 2.15. (\star Standard basis for K^n \star). Consider K^n as a vector space over K . We define the *standard basis* of K^n to be the basis $\hat{\mathcal{E}} = \{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n\}$, where the j th entry of $\hat{\mathbf{e}}_i$ is 1 when $i = j$ and 0 otherwise.

Remark 2.16. (Why not define dimensionality in terms of bases?). It is tempting to define a finite-dimensional vector space as one that has a finite basis. This definition would be equivalent to the one we’ve put in place as far as finite-dimensional vector spaces are concerned, but it becomes problematic for infinite-dimensional vector spaces. If we take the Axiom of Choice to be false, then not all vector spaces spanned by an infinite number of vectors have a basis. Therefore, it is best for an infinite-dimensional vector space to be one spanned by an infinite number of vectors rather than one that has as an infinite basis. If we took the later definition of “infinite-dimensional,” then, assuming the Axiom of Choice is false, not all vector spaces spanned by an infinite number of vectors would be classified as “infinite-dimensional”!

Remark 2.17. (\star 0-dimensional spaces \star). If we use the convention that the “empty sum” is the additive identity, $\mathbf{0}$, then the only basis of the vector space $\{\mathbf{0}\}$ is the empty set, \emptyset . Therefore $\{\mathbf{0}\}$ is the only 0-dimensional vector space.

Theorem 2.18. Every finite-dimensional vector space has a basis.

Proof. Take a spanning set of the vector space, and remove vectors until it becomes linearly independent to produce a basis. \square

Remark 2.19. The statement “every vector space, including infinite-dimensional vector spaces, has a basis” is equivalent to the Axiom of Choice.

Lemma 2.20. (Linear dependence lemma). Let V and W be vector spaces over a field K . If $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ and $\mathbf{w}_1, \dots, \mathbf{w}_\ell \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ with $\ell > k$, then the \mathbf{w}_i ’s are linearly dependent.

Proof. We prove the lemma by induction on k .

Base case. Assume $\mathbf{v}_1 \in V$ and $\mathbf{w}_1, \dots, \mathbf{w}_\ell \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$, where $\ell > k$. Then $\mathbf{w}_1 = c_1 \mathbf{v}_1, \dots, \mathbf{w}_\ell = c_\ell \mathbf{v}_1$ for some $c_1, \dots, c_\ell \in K$. Since $c_2 \mathbf{w}_1 - c_1 \mathbf{w}_2 + 0 \cdot \mathbf{w}_3 + 0 \cdot \mathbf{w}_4 + \dots + 0 \cdot \mathbf{w}_\ell = \mathbf{0}$ is a nontrivial linear combination, the \mathbf{w}_i ’s are linearly dependent.

Induction step. Suppose if $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ and $\mathbf{w}_1, \dots, \mathbf{w}_\ell \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ with $\ell > k$, then the \mathbf{w}_i ’s are linearly dependent. We will show this statement on k is true for $k + 1$.

Let $\mathbf{v}_1, \dots, \mathbf{v}_{k+1} \in V$ and assume $\mathbf{w}_1, \dots, \mathbf{w}_\ell \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1})$, where $\ell > k + 1$. Then for each i , $\mathbf{w}_i = c_{i1} \mathbf{v}_1 + \dots + c_{i, k+1} \mathbf{v}_{k+1}$ for some c_{ij} ’s.

We may assume that for some i , we have $c_{i, k+1} \neq 0$. (If this isn’t the case and $c_{i, k+1} = 0$ for all i , then the \mathbf{w}_i ’s are in the span of $\mathbf{v}_1, \dots, \mathbf{v}_k$, and we’re done by the induction hypothesis).

Since there is some i for which we can divide by $c_{i, k+1} \neq 0$, we see $\mathbf{v}_{k+1} \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_\ell)$; explicitly write out the linear combination to see this. Then, since each $\mathbf{w}_i \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1})$, and $\mathbf{v}_{k+1} \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_\ell)$, then each $\mathbf{w}_i \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_\ell)$; again write out the linear combinations to verify.

Now we define, for this proof, a primitive notion of the *projection* of a vector $\mathbf{v}_1 \in V$ onto another vector $\mathbf{v}_2 \in V$: we define $\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)$ to be the unique vector $\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2) := c_2 \mathbf{v}_2$ for which $\mathbf{v}_1 = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2$.

For each i we have $(\mathbf{w}_i - \text{proj}(\mathbf{w}_i \rightarrow \mathbf{w}_\ell)) \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$. Note that $\{\mathbf{w}_i - \text{proj}(\mathbf{w}_i \rightarrow \mathbf{w}_\ell) \mid i = 1, \dots, \ell\}$ is a linearly dependent set, since $\mathbf{w}_\ell - \text{proj}(\mathbf{w}_\ell \rightarrow \mathbf{w}_\ell) = \mathbf{0}$.

Claim: $\{\mathbf{w}_i - \text{proj}(\mathbf{w}_i \rightarrow \mathbf{w}_\ell) \mid i = 1, \dots, \ell - 1\}$ is also a linearly dependent set. To prove the claim, note that we assumed in the induction step that $\ell > k + 1$, so $\ell - 1 > k$; apply the induction hypothesis to this last inequality.

Since the claim is true, there exist nontrivial c_i ’s such that $c_1(\mathbf{w}_1 - \text{proj}(\mathbf{w}_1 \rightarrow \mathbf{w}_\ell)) + \dots + c_{\ell-1}(\mathbf{w}_{\ell-1} - \text{proj}(\mathbf{w}_{\ell-1} \rightarrow \mathbf{w}_\ell)) = \mathbf{0}$. Since, for each i , $\text{proj}(\mathbf{w}_i \rightarrow \mathbf{w}_\ell) = q_i \mathbf{w}_\ell$ for some $q_i \in F$, we can distribute the c_i ’s over the $\mathbf{w}_i - \text{proj}(\mathbf{w}_i \rightarrow \mathbf{w}_\ell)$, combine the $c_i q_i \mathbf{w}_\ell$ terms, and obtain an equation of the form $d_1 \mathbf{w}_1 + \dots + d_\ell \mathbf{w}_\ell + d_{\ell+1} \mathbf{w}_{\ell+1} = \mathbf{0}$ for some scalars d_i . As the c_i ’s are not all zero, then the d_i ’s must also not all be zero. This means the \mathbf{w}_i ’s are linearly dependent. \square

Lemma 2.21. (Linear dependence lemma contrapositive). Let V and W be vector spaces over a field K . Then if $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ are linearly independent and $\mathbf{w}_1, \dots, \mathbf{w}_\ell \in V$ with $k > \ell$, then some $\mathbf{v}_i \notin \text{span}(\mathbf{w}_1, \dots, \mathbf{w}_\ell)$.

Proof. Take the contrapositive of the previous lemma; then swap \mathbf{v}_i ’s with \mathbf{w}_i ’s and k with ℓ . \square

Theorem 2.22. (Uniqueness of dimension for finite-dimensional vector spaces). The dimension of a finite-dimensional vector space is well defined; a finite-dimensional vector space cannot have two different dimensions.

Proof. Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ be two bases for a finite-dimensional vector space. We show that E and F must contain the same number of vectors.

Suppose for contradiction that one of E, F contained more vectors than the other; without loss of generality, say F contains more vectors than E . Then, since E is a basis, each \mathbf{f}_i is in the span of $\mathbf{e}_1, \dots, \mathbf{e}_n$. But $m > n$, so, by the linear dependence lemma (Lemma 2.20), the vectors in F are linearly dependent. This is a contradiction because F is a basis. \square

2.2 Linear functions

Let V, W be vector spaces over a field K .

Definition 2.23. (\star Linear function \star). A function $\mathbf{f} : V \rightarrow W$ is *linear* iff for any basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of V and for all scalars $v_1, \dots, v_n \in K$,

$$\mathbf{f}(v_1\mathbf{e}_1 + \dots + v_k\mathbf{e}_k) = v_1\mathbf{f}(\mathbf{e}_1) + \dots + v_k\mathbf{f}(\mathbf{e}_k).$$

That is, \mathbf{f} is a linear function iff it preserves the decomposition of any input vector expressed “relative to the basis E .”

Equivalently, \mathbf{f} is linear iff, for all $\mathbf{v}, \mathbf{w} \in V$ and $c \in K$,

$$\begin{aligned}\mathbf{f}(\mathbf{v} + \mathbf{w}) &= \mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w}) \\ \mathbf{f}(c\mathbf{v}) &= c\mathbf{f}(\mathbf{v}).\end{aligned}$$

Another equivalent condition is

$$\mathbf{f}(c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k) = c_1\mathbf{f}(\mathbf{v}_1) + \dots + c_k\mathbf{f}(\mathbf{v}_k),$$

for all $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ and $c_1, \dots, c_k \in K$.

Linear functions are most commonly called “linear transformations” in traditional linear algebra books.

Remark 2.24. Every linear algebra book I have read defines a linear function $\mathbf{f} : V \rightarrow W$ to be one for which $\mathbf{f}(\mathbf{v} + \mathbf{w}) = \mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w})$ and $\mathbf{f}(c\mathbf{v}) = c\mathbf{f}(\mathbf{v})$. These conditions for linearity might be the “best” because they do not require a basis (and are therefore guaranteed to apply when V is infinite-dimensional) but I don’t think they are the best starting point for intuition.

Theorem 2.25. Linear functions $\mathbb{R}^n \rightarrow \mathbb{R}^n$ fix the origin and keep parallel lines parallel.

Proof. Let \mathbf{f} be a linear function $\mathbb{R}^n \rightarrow \mathbb{R}^n$.

First we show \mathbf{f} sends $\mathbf{0}$ to itself. We have $\mathbf{f}(\mathbf{0}) = \mathbf{f}(0 \cdot \mathbf{0}) = 0 \cdot \mathbf{f}(\mathbf{0}) = \mathbf{0}$.

Now we show \mathbf{f} sends parallel lines to parallel lines. Consider two parallel lines described by $\mathbf{r}_1(t) = \mathbf{v}_0 + t\mathbf{v}$ and $\mathbf{r}_2(t) = \mathbf{w}_0 + t\mathbf{v}$. Then $\mathbf{f}(\mathbf{r}_1(t)) = \mathbf{f}(\mathbf{v}_0) + t\mathbf{f}(\mathbf{v})$ and $\mathbf{f}(\mathbf{r}_2(t)) = \mathbf{f}(\mathbf{w}_0) + t\mathbf{f}(\mathbf{v})$. These transformed lines are parallel because they have the same direction vector, $\mathbf{f}(\mathbf{v})$. \square

Kernel and image of a linear function

Let V, W be vector spaces, and let $\mathbf{f} : V \rightarrow W$ be a linear function.

Definition 2.26. The *kernel* of \mathbf{f} is $\ker(\mathbf{f}) := \mathbf{f}^{-1}(\mathbf{0}) = \{\mathbf{v} \in V \mid \mathbf{f}(\mathbf{v}) = \mathbf{0}\}$. The *image* of V is $\text{im}(\mathbf{f}) := \mathbf{f}(V) = \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V \text{ s.t. } \mathbf{w} = \mathbf{f}(\mathbf{v})\}$.

Definition 2.27. (Rank of a linear function). The *rank* of \mathbf{f} is defined to be $\dim(\mathbf{f}(V))$, the dimension of the image of \mathbf{f} .

Theorem 2.28. (Kernel and image are subspaces). The kernel of \mathbf{f} is a vector subspace of V and the image of \mathbf{f} is a vector subspace of W . (Proof left as exercise).

Theorem 2.29. (One-to-one linear functions have trivial kernels). \mathbf{f} is one-to-one iff $\mathbf{f}^{-1}(\mathbf{0}) = \{\mathbf{0}\}$.

As $\{\mathbf{0}\}$ is the smallest (in the sense of set-containment) kernel possible for a linear function, we say that \mathbf{f} has a *trivial* kernel iff $\mathbf{f}^{-1}(\mathbf{0}) = \{\mathbf{0}\}$.

Proof. We use the contrapositive and prove that \mathbf{f} has a nontrivial kernel iff it is not one-to-one.

\mathbf{f} has a nontrivial kernel \iff there is a nonzero $\mathbf{v} \in V$ for which $\mathbf{f}(\mathbf{v}) = \mathbf{0} \iff$ for any $\mathbf{v}_1 \in V$ we have $\mathbf{f}(\mathbf{v}_1 + \mathbf{v}) = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}) = \mathbf{f}(\mathbf{v}_1) + \mathbf{0} = \mathbf{f}(\mathbf{v}_1) \iff \mathbf{f}$ is not one-to-one. \square

Remark 2.30. The idea here is that vectors in the preimage of some $\mathbf{w} \in W$ “differ by an element of the kernel.” You could prove the following fact to formalize this: $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{f}^{-1}(\mathbf{w})$ for some $\mathbf{w} \in f(V)$ if and only if $\mathbf{v}_1 = \mathbf{v}_2 + \mathbf{v}$, where $\mathbf{v}_2 \in V$ and $\mathbf{v} \in \mathbf{f}^{-1}(\mathbf{0})$.

Theorem 2.31. (Main dimension theorem). When V is finite-dimensional, then $\mathbf{f}^{-1}(\mathbf{0})$ and $\mathbf{f}(V)$ are also finite-dimensional, and we have $\dim(\mathbf{f}(V)) = \dim(V) - \dim(\ker(V))$.

Also, if $\mathbf{f}^{-1}(\mathbf{0})$ and $\mathbf{f}(V)$ are finite-dimensional, then V must be finite-dimensional, and the same relationship with dimensions holds.

This result is commonly called the *rank-nullity theorem*.

Proof. We prove the first part of the theorem (before “Also”).

If V is finite-dimensional, then $\mathbf{f}^{-1}(\mathbf{0})$ is also finite-dimensional since $\mathbf{f}^{-1}(\mathbf{0}) \subseteq V$. Choose a basis $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ for the kernel. Then add vectors to this basis so that it becomes $\{\mathbf{e}_1, \dots, \mathbf{e}_k, \mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$, a basis for V . Since $\dim(\mathbf{f}^{-1}(\mathbf{0})) = k$ and $\dim(V) = n$, we want to show $\dim(\mathbf{f}(V)) = \dim(V) - \dim(\mathbf{f}^{-1}(\mathbf{0})) = n - k$.

Suppose $\mathbf{v} = c_1\mathbf{e}_1 + \dots + c_n\mathbf{e}_n \in V$. Then for linear $\mathbf{f} : V \rightarrow W$, $\mathbf{f}(\mathbf{v}) = c_1\mathbf{f}(\mathbf{e}_1) + \dots + c_k\mathbf{f}(\mathbf{e}_k) + c_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + \dots + c_n\mathbf{f}(\mathbf{e}_n)$. Since $\mathbf{e}_1, \dots, \mathbf{e}_k \in \mathbf{f}^{-1}(\mathbf{0})$, this simplifies to $\mathbf{f}(\mathbf{v}) = \mathbf{f}(\mathbf{e}_{k+1}) + \dots + c_n\mathbf{f}(\mathbf{e}_n)$.

Therefore, any $\mathbf{w} \in \mathbf{f}(V)$ is in the span of $\{\mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$. We will show that $\{\mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$ is a basis for $\mathbf{f}(V)$. Once know this, then, since there are $n - k$ of these vectors, we have shown $\dim(\mathbf{f}(V)) = \dim(V) - \dim(\mathbf{f}^{-1}(\mathbf{0})) = n - k$, which is what we want.

So, it remains to show $\{\mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$ is a linearly independent set. Suppose for the sake of contradiction it's linearly dependent, i.e., that $d_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + \dots + d_n\mathbf{f}(\mathbf{e}_n) = \mathbf{0}$ for some d_i 's not all zero. By the linearity of \mathbf{f} , this is equivalent with $\mathbf{f}(d_{k+1}\mathbf{e}_{k+1} + \dots + d_n\mathbf{e}_n) = \mathbf{0}$ for some d_i 's not all zero. Thus $d_{k+1}\mathbf{e}_{k+1} + \dots + d_n\mathbf{e}_n \in \mathbf{f}^{-1}(\mathbf{0}) = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_k)$, which means $d_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + \dots + d_n\mathbf{f}(\mathbf{e}_n) = q_1\mathbf{e}_1 + \dots + q_k\mathbf{e}_k$ for some d_i 's and q_i 's not all zero. Then $d_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + \dots + d_n\mathbf{f}(\mathbf{e}_n) = q_1\mathbf{e}_1 + \dots + q_k\mathbf{e}_k$ for some d_i 's and q_i 's not all zero, i.e., $-(q_1\mathbf{e}_1 + \dots + q_k\mathbf{e}_k) + d_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + \dots + d_n\mathbf{f}(\mathbf{e}_n) = \mathbf{0}$ for some d_i 's and q_i 's not all zero. But, $\mathbf{e}_1, \dots, \mathbf{e}_n$ is a basis for V , so this cannot happen. Thus $\mathbf{f}(\mathbf{e}_{k+1}), \dots, \mathbf{f}(\mathbf{e}_n)$ are linearly independent. \square

Linear isomorphisms

Definition 2.32. (\star Linear isomorphism \star).

Let V, W be vector spaces over a field K . If $\mathbf{f} : V \rightarrow W$ is a bijective linear function, then it is called a *linear isomorphism* or an *isomorphism (of vector spaces)*.

When we have a linear isomorphism $\mathbf{f} : V \rightarrow W$, then, roughly speaking, all elements in V “interact” in the same way as their corresponding elements in W , so V and W are in some sense the same vector space. Specifically, the “interaction” among elements of V is mirrored by an “interaction” among elements of W as follows. If $\mathbf{f}(\mathbf{v}_1) = \mathbf{w}_1, \mathbf{f}(\mathbf{v}_2) = \mathbf{w}_2$, and c_1, c_2 are scalars in K , then, because \mathbf{f} is linear, \mathbf{f} sends the vector $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \in V$ to $c_1\mathbf{w}_1 + c_2\mathbf{w}_2 \in W$.

When we have an isomorphism between vector spaces V and W , we write $V \cong W$. (Note: “being isomorphic” is an equivalence relation). In this scenario, we also often say that an element $\mathbf{v} \in V$ can be *identified* with an element $\mathbf{w} \in W$.

Note, the previous theorem states that if linear function $\mathbf{f} : V \rightarrow W$ is a linear function, then \mathbf{f} is automatically an isomorphism if it is one-to-one or onto.

Lemma 2.33. (Only invertible linear functions preserve linear independence).

Let V be a finite-dimensional vector space, and let $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ be linearly independent. Consider a linear function $\mathbf{f} : V \rightarrow V$. Then

$$\begin{aligned} (\mathbf{v}_1, \dots, \mathbf{v}_k \text{ are linearly independent}) &\implies (\mathbf{f}(\mathbf{v}_1), \dots, \mathbf{f}(\mathbf{v}_n) \text{ are linearly independent}) \\ &\text{if and only if} \\ &\mathbf{f} \text{ is one-to-one} \end{aligned}$$

Proof.

(\Leftarrow). Suppose that \mathbf{f} is one-to-one and that $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent. We need to show that $\mathbf{f}(\mathbf{v}_1), \dots, \mathbf{f}(\mathbf{v}_k)$ are linearly independent. Since $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent, then the only choice of c_i 's for which $c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k = \mathbf{0}$ is the choice of all c_i 's being 0. Apply \mathbf{f} to both sides to obtain $c_1\mathbf{f}(\mathbf{v}_1) + \dots + c_k\mathbf{f}(\mathbf{v}_k) = \mathbf{0}$ only when $c_i = 0$ for all i . Therefore $\mathbf{f}(\mathbf{v}_1), \dots, \mathbf{f}(\mathbf{v}_k)$ are linearly independent.

(\Rightarrow). Suppose that if $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent, then $\mathbf{f}(\mathbf{v}_1), \dots, \mathbf{f}(\mathbf{v}_k)$ are linearly independent. We need to show \mathbf{f} is one-to-one; it suffices to show that \mathbf{f} has a trivial kernel. Let $\mathbf{v} \in \mathbf{f}^{-1}(\mathbf{0})$, so $\mathbf{f}(\mathbf{v}) = \mathbf{0}$. We want to show $\mathbf{v} = \mathbf{0}$. Since V is finite-dimensional, there is a basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for V . Then $\mathbf{f}(\mathbf{v}) = \mathbf{f}\left(\sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{e}_i\right) = \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{f}(\mathbf{e}_i) = \mathbf{0}$. Since E is a basis, it is a linearly independent set, and $([\mathbf{v}]_E)_i = 0$ for all i is the only solution to $\sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{f}(\mathbf{e}_i) = \mathbf{0}$. Therefore $([\mathbf{v}]_E)_i = 0$ for all i , so $\mathbf{v} = \mathbf{0}$. \square

Lemma 2.34. (Linear isomorphisms provide bases).

Let V and W be finite-dimensional vector spaces. If $\mathbf{f} : V \rightarrow W$ is a linear isomorphism, then any basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for V provides the basis $\mathbf{f}(E) = \{\mathbf{f}(\mathbf{e}_1), \dots, \mathbf{f}(\mathbf{e}_n)\}$ for W . Therefore, finite-dimensional vector spaces that are isomorphic must have the same dimension.

Proof. Since \mathbf{f} is surjective, then for all $\mathbf{w} \in W$ there exists a $\mathbf{v} \in V$ for which $\mathbf{f}(\mathbf{v}) = \mathbf{w}$. Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V . Then for all $\mathbf{w} \in W$ we have $\mathbf{w} = \mathbf{f}\left(\sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{e}_i\right) = \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{f}(\mathbf{e}_i)$. Thus, the $\mathbf{f}(\mathbf{e}_i)$ span W . Since E is a basis and \mathbf{f} has a trivial kernel (it is injective), then the $\mathbf{f}(\mathbf{e}_i)$ are linearly independent because injective linear functions preserve the linear independence of vectors (see the previous lemma). Thus, $\{\mathbf{f}(\mathbf{e}_1), \dots, \mathbf{f}(\mathbf{e}_n)\}$ is a basis for W . \square

Theorem 2.35. (Finite-dimensional vector spaces of the same dimension are isomorphic).

Let V and W be finite-dimensional vector spaces. Then there exists a linear isomorphism $V \rightarrow W$ iff $\dim(V) = \dim(W)$.

Proof. The previous lemma showed the forward direction of the iff. We show the reverse direction. We show that V is isomorphic to the vector space $K^{\dim(V)}$ over K . Then we have $W \cong K^{\dim(W)} \cong K^{\dim(V)} \cong V$ because $\dim(V) = \dim(W)$.

Choose a basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for V , and let $F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ be a basis for $K^{\dim(V)}$. We define a linear isomorphism on basis vectors by $\mathbf{e}_i \mapsto \mathbf{f}_i$. The explicit checks of injectivity and surjectivity are left to the reader. (To show injectivity, we show this map has a trivial kernel. This follows from the fact that E is a basis. Surjectivity follows from linearity). \square

Remark 2.36. While the previous theorem shows that all finite-dimensional vector spaces of the same dimension are isomorphic to each other, it is possible for a linear function between two finite-dimensional vector spaces of the same dimension to fail to be a linear isomorphism.

Theorem 2.37. (A linear function of finite-dimensional vector spaces of the same dimension is one-to-one iff it is onto).

Let V and W be finite dimensional vector spaces with same dimension, $\dim(V) = \dim(W)$, and let $\mathbf{f} : V \rightarrow W$ be a linear function. Then \mathbf{f} is one-to-one iff \mathbf{f} is onto.

Therefore, to check if a linear function $V \rightarrow W$ is a linear isomorphism, it suffices to check either injectivity or surjectivity.

Proof. First, we use the contrapositive to show that if $\mathbf{f} : V \rightarrow W$ is linear, then \mathbf{f} is one-to-one iff $\dim(V) = \dim(\mathbf{f}(V))$: \mathbf{f} is not one-to-one $\iff \mathbf{f}$ has a nontrivial kernel $\iff \dim(\mathbf{f}^{-1}(\mathbf{0})) > 0 \iff \dim(\mathbf{f}(V)) < \dim(V) \iff \dim(\mathbf{f}(V)) \neq \dim(V)$. The second to last logical equivalence uses the main dimension theorem. The last logical equivalence follows because $\dim(\mathbf{f}(V)) > \dim(V)$ is impossible, which is also the case because of the main dimension theorem.

Thus, \mathbf{f} is one-to-one iff $\dim(V) = \dim(\mathbf{f}(V))$. To complete the proof, we show that if \mathbf{f} is one-to-one, $\dim(V) = \dim(W)$, and $\dim(V) = \dim(\mathbf{f}(V))$, then \mathbf{f} is a linear isomorphism.

So, assume these hypotheses are the case. Since $\dim(\mathbf{f}(V)) = \dim(W)$, it is straightforwardly shown that $\mathbf{f}(V) \supseteq W$. Then $\mathbf{f}(V) \subseteq W$ and $W \subseteq \mathbf{f}(V)$, so $\mathbf{f}(V) = W$, which means \mathbf{f} is onto, i.e., \mathbf{f} is a linear isomorphism. \square

Definition 2.38. (\star Natural linear isomorphism \star).

Roughly speaking, a linear isomorphism is said to be “natural” if it does not depend on a choice of basis. This definition of “natural” is not completely technically correct, but it will suffice for our purposes, because the converse (any linear isomorphism which depends on a choice of basis is unnatural) *is* true. To read more about what “natural” really means, look up “category theory” online.

Inverse of a linear function

Recall the definition of an inverse function (Definition 1.9). Let V and W be vector spaces; we will consider an invertible linear function $\mathbf{f} : V \rightarrow W$.

Theorem 2.39. (The inverse of a linear function is a linear function). Consider an invertible linear function $\mathbf{f} : V \rightarrow W$. The inverse \mathbf{f}^{-1} is also a linear function.

Proof. Since invertible functions are bijections (see Theorem 1.10), then \mathbf{f} is a linear isomorphism. We need to show that $\mathbf{f}^{-1}(\mathbf{v} + \mathbf{w}) = \mathbf{f}^{-1}(\mathbf{v}) + \mathbf{f}^{-1}(\mathbf{w})$ and $\mathbf{f}^{-1}(c\mathbf{v}) = c\mathbf{f}^{-1}(\mathbf{v})$.

Consider $\mathbf{f}^{-1}(\mathbf{v} + \mathbf{w})$. Since \mathbf{f} is onto, then $\mathbf{v} = \mathbf{f}(\mathbf{v}')$ and $\mathbf{w} = \mathbf{f}(\mathbf{w}')$ for some $\mathbf{v}', \mathbf{w}' \in V$. Therefore $\mathbf{f}^{-1}(\mathbf{v} + \mathbf{w}) = \mathbf{f}^{-1}(\mathbf{f}(\mathbf{v}') + \mathbf{f}(\mathbf{w}')) = \mathbf{f}^{-1}(\mathbf{f}(\mathbf{v}' + \mathbf{w}')) = \mathbf{v}' + \mathbf{w}' = \mathbf{f}^{-1}(\mathbf{v}) + \mathbf{f}^{-1}(\mathbf{w})$.

A similar argument works for showing $\mathbf{f}^{-1}(c\mathbf{v}) = c\mathbf{f}^{-1}(\mathbf{v})$. \square

Theorem 2.40. (Dimensions of vector spaces must be equal when a linear function between them is invertible).

Let V and W be finite-dimensional. If $\mathbf{f} : V \rightarrow W$ is invertible, then $\dim(V) = \dim(W)$.

Proof. As was noted in the proof of the previous theorem, every invertible linear function is a linear isomorphism. Since \mathbf{f} is one-to-one, $\mathbf{f}^{-1}(\mathbf{0}) = \{\mathbf{0}\}$ and so $\dim(\mathbf{f}^{-1}(\mathbf{0})) = 0$. Using that \mathbf{f} is onto with the main dimension theorem (Theorem 2.31), we have $\dim(\mathbf{f}(V)) = \dim(W) = \dim(V) - \dim(\mathbf{f}^{-1}(\mathbf{0}))$, so $\dim(W) = \dim(V)$. \square

Definition 2.41. (Identity matrix).

Definition 2.42. (Inverse matrix).

Systems of linear equations with matrices

Reading this subsection is entirely unnecessary for the remaining content in this book. A treatment of systems of linear equations has only been provided to emphasize that linear algebra should not be introduced as a consequence of studying systems of linear equations. Instead, systems of linear equations should be viewed as an application of linear algebra.

Theorem 2.43. (Row operations are linear transformations).

Remark 2.44. (Column operations).

Lemma 2.45. (The rows of an invertible matrix form a basis for K^n).

Suppose \mathbf{A} is an $n \times n$ matrix with entries in a field K . If \mathbf{A} is invertible, then \mathbf{A} has linearly independent columns, i.e., the columns of \mathbf{A} form a basis for K^n . If \mathbf{A} is invertible, then \mathbf{A}^\top is also invertible (with $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$). Therefore, if \mathbf{A} is invertible, then the rows of \mathbf{A} form a basis for K^n .

Lemma 2.46. (A matrix is invertible iff it is a product of elementary matrices).

Theorem 2.47. (A matrix is invertible iff its RREF is the identity matrix).

Proof.

(\implies).

(\impliedby).

[...]

row operations correspond to invertible linear functions. since columns are LI iff invertible and (only) linear functions preserve LI, RREF must be identity iff invertible. \square

Matrices and coordinatization

★ Coordinates relative to a basis ★

Let V be a finite-dimensional vector space over a field K , and let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V .

Definition 2.48. (Coordinates of a vector relative to a basis). Given a vector $\mathbf{v} \in V$, we define $[\mathbf{v}]_E$ to be the vector in $K^{\dim(V)}$ that stores the *coordinates of \mathbf{v} relative to the basis E* . Formally, $[\mathbf{v}]_E$ is the tuple of scalars

$$[\mathbf{v}]_E := \begin{pmatrix} ([\mathbf{v}]_E)_1 \\ \vdots \\ ([\mathbf{v}]_E)_n \end{pmatrix} \in K^n$$

for which

$$\mathbf{v} = \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{e}_i.$$

Note, we are guaranteed that such scalars exist because E is a basis for V , so E in particular spans V .

Definition 2.49. (Linear function acting on a list of vectors). Consider a linear function $\mathbf{f} : V \rightarrow K^m$. We define the notation

$$\mathbf{f}(E) := \begin{pmatrix} \mathbf{f}(\mathbf{e}_1) & \dots & \mathbf{f}(\mathbf{e}_n) \end{pmatrix}.$$

Note that when \mathbf{f} is invertible and E is a basis of V , then the columns of the matrix $\mathbf{f}(E)$ are a basis of K^m , since invertible linear functions preserve linear independence (see Theorem 2.33).

We also define $\mathbf{E} := \mathbf{I}_V(E)$, where \mathbf{I}_V is the identity on V , since

$$\mathbf{I}_V(E) = \begin{pmatrix} \mathbf{e}_1 & \dots & \mathbf{e}_n \end{pmatrix}.$$

Theorem 2.50. Taking coordinates relative to a basis is an invertible linear operation. Put differently, $[\cdot]_E$ is an invertible linear function.

Proof. For linearity, we show that $[\mathbf{v}_1 + \mathbf{v}_2]_E = [\mathbf{v}_1]_E + [\mathbf{v}_2]_E$ and that $[c\mathbf{v}]_E = c[\mathbf{v}]_E$.

$$\begin{aligned} [\mathbf{v}_1 + \mathbf{v}_2]_E &= \left[\left(\sum_{i=1}^n ([\mathbf{v}_1]_E)_i \mathbf{e}_i + \sum_{i=1}^n ([\mathbf{v}_2]_E)_i \mathbf{e}_i \right) \right]_E \\ &= \left[\left(\sum_{i=1}^n \left[\left(([c_1 \mathbf{v}_1]_E)_i + ([\mathbf{v}_2]_E)_i \right) \mathbf{e}_i \right] \right) \right]_E = \begin{pmatrix} ([\mathbf{v}_1]_E)_1 + ([\mathbf{v}_2]_E)_1 \\ \vdots \\ ([\mathbf{v}_1]_E)_m + ([\mathbf{v}_2]_E)_m \end{pmatrix} = \begin{pmatrix} ([\mathbf{v}_1]_E)_1 \\ \vdots \\ ([\mathbf{v}_1]_E)_m \end{pmatrix} + \begin{pmatrix} ([\mathbf{v}_2]_E)_1 \\ \vdots \\ ([\mathbf{v}_2]_E)_m \end{pmatrix} \\ &= [\mathbf{v}_1]_E + [\mathbf{v}_2]_E. \end{aligned}$$

Now we show $[c\mathbf{v}]_E = c[\mathbf{v}]_E$. If $[\mathbf{v}]_E = \begin{pmatrix} ([\mathbf{v}]_E)_1 \\ \vdots \\ ([\mathbf{v}]_E)_n \end{pmatrix}$, then $\mathbf{v} = \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{e}_i$, so $c\mathbf{v} = c \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{e}_i =$

$\sum_{i=1}^n c([\mathbf{v}]_E)_i \mathbf{e}_i$. Thus by definition of $[\cdot]_E$ we see $[\cdot]_E(c\mathbf{v}) = c[\mathbf{v}]_E$. Therefore $[\cdot]_E$ is linear.

$[\cdot]_E$ is invertible because it has a trivial kernel (see 2.29): if $[\cdot]_E(\mathbf{v}) = \mathbf{0}$, then the coordinates of \mathbf{v} relative to E are all zero, so $\mathbf{v} = \mathbf{0}$. \square

Matrices as representative of linear functions

Let V be a finite-dimensional vector space over a field K with a basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$.

Derivation 2.51. (\star Primitive matrix of a linear function $V \rightarrow K^m$ \star).

The fundamental idea behind this theorem is the definition of a linear function. Recall from Definition 2.23 that the action of a linear function on any vector is determined by the function does to a basis.

To start, consider a linear function $\mathbf{f} : V \rightarrow K^m$. Then from the definition of $[\cdot]_E$ (see Definition 2.48) we have $\mathbf{v} = \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{e}_i$, so

$$\mathbf{f}(\mathbf{v}) = \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{f}(\mathbf{e}_i).$$

This is just an expression of the fact that linear functions are completely determined by what they do to a set of basis vectors.

Why not just specify what \mathbf{f} is by storing the transformed basis vectors? This exactly what we will do. We define the *matrix-vector product* between a *matrix* $\mathbf{A} = (a_{ij})$, which is a two-dimensional

grid of scalars from K whose ij entry is denoted a_{ij} , and a *column vector* $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in K^{\dim(V)}$.

$$\mathbf{A}\mathbf{c} = \begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} := \sum_{i=1}^n c_i \mathbf{a}_i$$

Note that the \mathbf{a}_i are column vectors in K^m , so the matrix \mathbf{A} is indeed a grid of scalars from K . This definition was contrived so that the action of \mathbf{f} on a vector \mathbf{v} is expressed with such a matrix-vector product:

$$\mathbf{f}(\mathbf{v}) = \begin{pmatrix} \mathbf{f}(\mathbf{e}_1) & \dots & \mathbf{f}(\mathbf{e}_n) \end{pmatrix} \begin{pmatrix} ([\mathbf{v}]_E)_1 \\ \vdots \\ ([\mathbf{v}]_E)_n \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{e}_1) & \dots & \mathbf{f}(\mathbf{e}_n) \end{pmatrix} [\mathbf{v}]_E.$$

Note that because \mathbf{f} maps into K^m , each $\mathbf{f}(\mathbf{e}_i)$ is a column vector in K^m . So the matrix in the above expression, containing $\mathbf{f}(\mathbf{e}_i)$ as its i th column, is grid of scalars- just as was the \mathbf{A} in the definition of matrix-vector product.

Now we see that, *after choosing a basis E for V* , a linear function $\mathbf{f} : V \rightarrow K^m$ corresponds to its so-called *primitive matrix relative to the basis E* ,

$$\begin{pmatrix} \mathbf{f}(\mathbf{e}_1) & \dots & \mathbf{f}(\mathbf{e}_n) \end{pmatrix}.$$

(More on what “primitive” refers to follows soon). We say that the matrix is expressed *relative to E* because the column vectors in the matrix depend on the choice of E ; the i th column of the “primitive” matrix of \mathbf{f} relative to E is $\mathbf{f}(\mathbf{e}_i)$.

If \mathbf{A} denotes the primitive matrix relative to E of the linear function $\mathbf{f} : V \rightarrow K^m$, then \mathbf{A} has the characterizing property

$$\mathbf{f}(\mathbf{v}) = \mathbf{A}[\mathbf{v}]_E$$

Derivation 2.52. (\star Matrix of a linear function relative to bases \star).

What about the above approach is “primitive”? Well, it is “primitive” in the sense that it works for linear functions $V \rightarrow K^m$, but not for linear functions $V \rightarrow W$, where W is another finite-dimensional vector space. This is because the output of a linear function mapping into an arbitrary finite-dimensional vector space such as W isn’t necessarily a tuple of scalars and could be something like a polynomial.

Given a linear function $\mathbf{f} : V \rightarrow W$, we can still produce a matrix from \mathbf{f} . Let F be a basis for W ; then $[\cdot]_F \circ \mathbf{f} : V \rightarrow K^{\dim(W)}$ is a linear function because a composition of linear functions is also a linear function (prove this fact as an exercise). We will use the primitive matrix of $[\cdot]_F \circ \mathbf{f} : V \rightarrow K^{\dim(W)}$ relative to E :

$$\left(([\cdot]_F \circ \mathbf{f})(\mathbf{e}_1) \quad \dots \quad ([\cdot]_F \circ \mathbf{f})(\mathbf{e}_n) \right) = \left([\mathbf{e}_1]_F \quad \dots \quad [\mathbf{e}_n]_F \right)$$

We call this matrix *the matrix of \mathbf{f} relative to the bases E and F* . This matrix is $[\mathbf{f}(E)]_F$, where we have used Definition 2.49 to define $[\mathbf{f}(E)]_F := [\cdot]_F(\mathbf{f}(E))$. That is,

$$\boxed{[\mathbf{f}(E)]_F := \left([\mathbf{e}_1]_F \quad \dots \quad [\mathbf{e}_n]_F \right)}$$

Still looking at the characterization of the matrix of a linear function $V \rightarrow K^m$ from above, we see that $[\mathbf{f}(E)]_F$ must satisfy the characterizing property

$$\boxed{[\mathbf{f}(\mathbf{v})]_F = [\mathbf{f}(E)]_F [\mathbf{v}]_E}$$

Note, the right-hand side of the above is a matrix-vector product.

In words, the matrix of \mathbf{f} with respect to E and F expresses the action of \mathbf{f} by converting an input vector \mathbf{v} to its coordinatization in $K^{\dim(V)}$, mapping this coordinatization into the vector space W , and then applying a final coordinatization to return a vector in $K^{\dim(W)}$.

Definition 2.53. (Matrix of a linear function $V \rightarrow V$ relative to one basis).

Let V be a vector space with basis E , and consider a linear function $\mathbf{f} : V \rightarrow V$. The matrix $[\mathbf{f}(E)]_E$ of \mathbf{f} relative to E and E is called *the matrix of \mathbf{f} relative to E* .

Remark 2.54. (Matrix-vector product pedagogy). All the linear algebra texts I have read always present the relationship between linear functions and matrices in the following way: first define matrices as grids of scalars (often in the context of systems of linear equations), then define a linear function as satisfying the second condition of Definition 2.23, and then prove that each linear function has a matrix. This is bad pedagogy; there should be no need to conjecture and prove that a matrix-vector product corresponds to the action of a linear function, because this fact is apparent from Derivation 2.51.

It’s important to emphasize that definition of matrix-vector product is therefore really no more than a definition. We decided that writing a grid of scalars next to a vector in the specific way of Derivation 2.51 should produce the output of Derivation 2.51 because this is what formalizes the correspondence between linear functions and the lists of their transformed basis vectors. Now, it is true that this definition of matrix-vector product leads to somewhat complicated formulas for the i th entry of a matrix-vector product and for the ij entry of a matrix-matrix product (these formulas will be presented in Theorems 2.61 and 2.62). Upon seeing these index-notation formulas, remember that they are consequences, and are not “just the way things are”!

Remark 2.55. (Primitive matrix as special case of matrix relative to bases). The primitive matrix of a linear function $\mathbf{f} : V \rightarrow K^m$ relative to E is the matrix $[\mathbf{f}(E)]_{\hat{\mathbf{e}}} = \mathbf{f}(E)$ of $\mathbf{f} : V \rightarrow K^m$ relative to the bases E and $\hat{\mathbf{e}}$, where $\hat{\mathbf{e}}$ is the standard basis for K^m , and where $\mathbf{f}(E)$ is the matrix resulting from a linear function $V \rightarrow K^m$ acting on a list of vectors (see Definition 2.49). The key fact here is

$$\text{that } \left[\begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} \right]_{\hat{\mathbf{e}}} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} \text{ for } \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} \in K^m.$$

Theorem 2.56. (Matrices with respect to bases are also primitive matrices).

The previous remark discussed how matrices with respect to bases can be regarded as primitive matrices. This theorem shows the converse of the last remark; we will see that every matrix with respect to bases can be viewed as a primitive matrix.

Let V and W be finite-dimensional vector spaces with bases E and F , respectively. We know from Derivation 2.52 that every linear function $\mathbf{f} : V \rightarrow W$ corresponds to the matrix $[\mathbf{f}(E)]_F$ of \mathbf{f} relative to E and F , where $[\mathbf{f}(E)]_F$ is characterized by the equation $[\mathbf{f}(\mathbf{v})]_F = [\mathbf{f}(E)]_F[\mathbf{v}]_E$. Rephrase this equation as $([\cdot]_F \circ \mathbf{f})(\mathbf{v}) = [\mathbf{f}(E)]_F[\mathbf{v}]_E$ and set $\mathbf{c} = [\mathbf{v}]_E$ to obtain

$$([\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1})(\mathbf{c}) = [\mathbf{f}(E)]_F \mathbf{c}.$$

Now we define $\mathbf{f}_{E,F} : K^{\dim(V)} \rightarrow K^{\dim(W)}$ by

$$\boxed{\mathbf{f}_{E,F} := [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}}$$

so that the above rephrases as

$$\mathbf{f}_{E,F}(\mathbf{c}) = [\mathbf{f}(E)]_F \mathbf{c} = [\mathbf{f}(E)]_F [\mathbf{c}]_{\hat{\mathbf{e}}},$$

where $\hat{\mathbf{e}}$ is the standard basis for K^n . Here, we have used the fact that for $\mathbf{c} \in K^n$, we have $[\mathbf{c}]_{\hat{\mathbf{e}}} = \mathbf{c}$.

Looking at the characterizing property of a primitive matrix (see the last box of Derivation 2.51), we see that $[\mathbf{f}(E)]_F$ is the primitive matrix of $\mathbf{f}_{E,F} : K^{\dim(V)} \rightarrow K^{\dim(W)}$ relative to $\hat{\mathbf{e}}$. Recalling from the previous remark that we can notate the primitive matrix of $\mathbf{f}_{E,F}$ relative to $\hat{\mathbf{e}}$ by $\mathbf{f}_{E,F}(\hat{\mathbf{e}})$, this fact can be stated as

$$\boxed{\underbrace{\mathbf{f}_{E,F}(\hat{\mathbf{e}})}_{\text{primitive matrix of } \mathbf{f}_{E,F} \text{ relative to } \hat{\mathbf{e}}} = \underbrace{[\mathbf{f}(E)]_F}_{\text{matrix of } \mathbf{f} \text{ relative to } E \text{ and } F}}$$

Remark 2.57. ($\mathbf{f}_{E,F}$ as an induced function).

Consider the context of the previous theorem. We can additionally understand $\mathbf{f}_{E,F}$ to be the “induced” linear function for which this diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{\mathbf{f}} & W \\ [\cdot]_E \downarrow & & \downarrow [\cdot]_F \\ K^{\dim(V)} & \xrightarrow{\mathbf{f}_{E,F}} & K^{\dim(W)} \end{array}$$

To say the diagram “commutes” is just another way of saying $\mathbf{f}_{E,F} = [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}$. We think of $\mathbf{f}_{E,F}$ as being *induced* by the choice of \mathbf{f} .

Definition 2.58. (\star Composition of linear functions that map into K^m and $K^p \star$).

Let V be a finite-dimensional vector space over a field K , and consider linear functions $\mathbf{f} : V \rightarrow K^m$ and $\mathbf{g} : K^m \rightarrow K^p$. Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V , $\hat{\mathbf{e}} = \{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_m\}$ be the standard basis for K^m , and $\hat{\mathbf{f}} = \{\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_p\}$ be the standard basis for K^p .

The composition $\mathbf{g} \circ \mathbf{f} : V \rightarrow K^p$ is also a linear function (prove this as an exercise). Since $\mathbf{g} \circ \mathbf{f}$ is a linear function $V \rightarrow K^p$, its primitive matrix relative to E (see Derivation 2.51) is

$$\left((\mathbf{g} \circ \mathbf{f})(\mathbf{e}_1) \quad \dots \quad (\mathbf{g} \circ \mathbf{f})(\mathbf{e}_n) \right) = \left(\mathbf{g}(\mathbf{f}(\mathbf{e}_1)) \quad \dots \quad \mathbf{g}(\mathbf{f}(\mathbf{e}_n)) \right).$$

But $\mathbf{f} : V \rightarrow K^m$ and $\mathbf{g} : K^m \rightarrow K^p$, so \mathbf{f} and \mathbf{g} also have primitive matrices. The primitive matrix of \mathbf{f} relative to E is $[\mathbf{f}(E)]_{\hat{\mathbf{e}}} = \mathbf{f}(E)$ and the primitive matrix of \mathbf{g} relative to $\hat{\mathbf{e}}$ is $[\mathbf{g}(\hat{\mathbf{e}})]_{\hat{\mathbf{g}}} = \mathbf{g}(\hat{\mathbf{e}})$ (see Remark 2.55). The above then becomes

$$\left(\mathbf{g}(\hat{\mathbf{e}})\mathbf{f}(\mathbf{e}_1) \quad \dots \quad \mathbf{g}(\hat{\mathbf{e}})\mathbf{f}(\mathbf{e}_n) \right) = \left(\mathbf{g}(\hat{\mathbf{e}})(\mathbf{f}(E))_1 \quad \dots \quad \mathbf{g}(\hat{\mathbf{e}})(\mathbf{f}(E))_n \right)$$

where $(\mathbf{f}(E))_i$ is the i th column of $\mathbf{f}(E)$.

Given an $m \times n$ matrix \mathbf{A} and a $p \times m$ matrix \mathbf{B} , we define \mathbf{BA} to be the matrix

$$\boxed{\mathbf{BA} := \begin{pmatrix} \mathbf{Ba}_1 & \dots & \mathbf{Ba}_n \end{pmatrix}}$$

so that

$$\boxed{(\text{primitive matrix of } \mathbf{g} \circ \mathbf{f} \text{ relative to } E) = \mathbf{g}(\hat{\mathbf{e}})\mathbf{f}(E)}$$

We call \mathbf{BA} the *matrix-matrix* product of \mathbf{B} and \mathbf{A} . So, the right-hand side of the most recent equation is the matrix-matrix product of $\mathbf{g}(\hat{\mathbf{e}})$ and $\mathbf{f}(E)$.

Note that $(\mathbf{BA})\mathbf{v} = \mathbf{B}(\mathbf{A}\mathbf{v})$ because $(\mathbf{g} \circ \mathbf{f})(\mathbf{v}) = \mathbf{g}(\mathbf{f}(\mathbf{v}))$.

Remark 2.59. (Compatibility of matrices for matrix-matrix products). The composition of two linear functions is only defined when the output space of one is the entire input space of the other. Thus, the matrix-matrix product \mathbf{BA} of an $m \times n$ matrix with an $r \times s$ matrix \mathbf{B} is only defined when $r = n$.

Theorem 2.60. (Matrix-matrix products of linear functions mapping into finite-dimensional vector spaces).

Let V, W, Y be finite-dimensional vector spaces, with bases E, F, G , respectively, and let $\mathbf{f} : V \rightarrow W$ and $\mathbf{g} : W \rightarrow Y$ be linear functions. We will use the previous definition to produce a matrix relative to bases for the linear function $\mathbf{g} \circ \mathbf{f} : V \rightarrow Y$.

The matrix of $\mathbf{g} \circ \mathbf{f}$ relative to E and G is the same as the primitive matrix for $(\mathbf{g} \circ \mathbf{f})_{E,G}$ relative to E (see Theorem 2.56). We will therefore compute this later matrix. We have

$$(\mathbf{g} \circ \mathbf{f})_{E,G} = ([\cdot]_G \circ \mathbf{g} \circ [\cdot]_{F^{-1}}) \circ ([\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}) = \mathbf{g}_{F,G} \circ \mathbf{f}_{E,F}$$

Note that $\mathbf{f}_{E,F} : K^{\dim(V)} \rightarrow K^{\dim(W)}$, $\mathbf{g}_{F,G} : K^{\dim(W)} \rightarrow K^{\dim(Y)}$, and $(\mathbf{g} \circ \mathbf{f})_{E,G} : K^{\dim(V)} \rightarrow K^{\dim(Y)}$, so we are in the situation of the previous definition. Thus, the primitive matrix of $(\mathbf{g} \circ \mathbf{f})_{E,G}$ relative to E is the matrix-matrix product of the primitive matrix of $\mathbf{g}_{F,G}$ relative to F and the primitive matrix of $\mathbf{f}_{E,F}$ relative to E :

$$(\text{primitive matrix of } (\mathbf{g} \circ \mathbf{f})_{E,G} \text{ relative to } E) = [\mathbf{g}(F)]_G [\mathbf{f}(E)]_F.$$

Therefore

$$\boxed{(\text{matrix of } \mathbf{g} \circ \mathbf{f} \text{ relative to } E \text{ and } G) = [\mathbf{g}(F)]_G [\mathbf{f}(E)]_F}$$

Theorem 2.61. (\star i th entry of matrix-vector product \star).

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix with entries in a field K and let $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in K^n$ be a

column vector. Referring to the definition of matrix-vector product in Derivation 2.51, we see the matrix-vector product \mathbf{Ac} has the following i th entry:

$$(\mathbf{A}\mathbf{c})_i = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = c_1 \begin{pmatrix} a_{1,1} \\ \vdots \\ a_{m,1} \end{pmatrix} + \cdots + c_n \begin{pmatrix} a_{1,n} \\ \vdots \\ a_{m,n} \end{pmatrix} = \begin{pmatrix} c_1 a_{1,1} + \cdots + c_n a_{1,n} \\ \vdots \\ c_n a_{m,1} + \cdots + c_n a_{m,n} \end{pmatrix}.$$

Therefore,

$$\boxed{(\mathbf{A}\mathbf{c})_i = (i\text{th row of } \mathbf{A}) \cdot \mathbf{c}}$$

Here $\cdot : K^n \times K^n \rightarrow K$ denotes the *dot product* of vectors in K^n , defined by

$$\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \cdot \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix} = c_1 d_1 + \cdots + c_n d_n.$$

Since the dot product must take two column vectors as input, what we technically mean by “ i th row of \mathbf{A} ” in the boxed equation is “column vector that contains entries of i th row of \mathbf{A} .”

Section [...] of this Appendix discusses the intuition behind the dot product.

Theorem 2.62. (\star ij entry of matrix-matrix product \star).

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix with entries in a field K and $\mathbf{B} = (b_{ij})$ be an $n \times p$ matrix with entries in K . Then the ij entry of the matrix-matrix product $\mathbf{B}\mathbf{A}$ can be computed using the definition of matrix-matrix product (Theorem 2.60) and the previous theorem, which gives a formula for the i th entry of a vector:

$$(\mathbf{B}\mathbf{A})_{ij} = \mathbf{B} \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} \mathbf{B}\mathbf{a}_1 & \cdots & \mathbf{B}\mathbf{a}_n \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \cdot \mathbf{a}_1 & \cdots & \mathbf{b}_1 \cdot \mathbf{a}_n \\ \vdots & & \vdots \\ \mathbf{b}_m \cdot \mathbf{a}_1 & \cdots & \mathbf{b}_m \cdot \mathbf{a}_n \end{pmatrix}.$$

Here \mathbf{a}_i is the i th column of \mathbf{A} and \mathbf{b}_i is the i th row of \mathbf{B} . So we get

$$\boxed{(\mathbf{B}\mathbf{A})_{ij} = (i\text{th row of } \mathbf{B}) \cdot (j\text{th column of } \mathbf{A})}$$

Similarly as in the previous theorem, what we mean by “ i th row of \mathbf{B} ” in the boxed equation is “column vector that contains entries of i th row of \mathbf{B} .”

Change of basis

Theorem 2.63. (\star Change of basis for vectors \star).

Let V be a finite-dimensional vector space with bases E and F . We will now discover how to relate $[\mathbf{v}]_E$ to $[\mathbf{v}]_F$.

To start, consider the special case $V = K^n$. Let \mathbf{E} and \mathbf{F} be the matrices with i th columns \mathbf{e}_i and \mathbf{f}_i , respectively. Then for $\mathbf{c} \in K^n$ we have $\mathbf{F}[\mathbf{c}]_F = \mathbf{c} = \mathbf{E}[\mathbf{c}]_E$, so $\mathbf{c} = \mathbf{F}^{-1}\mathbf{E}[\mathbf{c}]_E$. By the definition of matrix-matrix multiplication (Definition 2.58), we have

$$\mathbf{F}^{-1}\mathbf{E} = \mathbf{F}^{-1} \begin{pmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{pmatrix} = \begin{pmatrix} \mathbf{F}^{-1}\mathbf{e}_1 & \cdots & \mathbf{F}^{-1}\mathbf{e}_n \end{pmatrix} = \begin{pmatrix} [\mathbf{e}_1]_F & \cdots & [\mathbf{e}_n]_F \end{pmatrix} = [\cdot]_F(E) := [\mathbf{E}]_F,$$

where we have defined the notation $[\mathbf{E}]_F := [\cdot]_F(E)$. (The notation $[\cdot]_F(E)$ was defined in Definition 2.49).

Therefore $\mathbf{c} = [\mathbf{E}]_F[\mathbf{c}]_E$. We now generalize this result to one that holds when V is an arbitrary finite-dimensional vector space.

The matrix $[\cdot]_F(E) = [\mathbf{E}]_F$ was the centerpiece of the above argument. To generalize, we notice that the primitive matrix of $[\cdot]_F$ relative to E is $[\cdot]_F(E) = [\mathbf{E}]_F$. The characterizing property of primitive matrices (see the very end of Derivation 2.51) then implies that for any $\mathbf{v} \in V$, we have

$$[\mathbf{v}]_F = [\mathbf{E}]_F [\mathbf{v}]_E$$

It's also worth noting that since $[\mathbf{E}]_F = [\mathbf{I}_V(E)]_F$, then $[\mathbf{E}]_F$ is the matrix of the identity \mathbf{I}_V on V . So, the above can be restated as

$$[\cdot]_F = [\mathbf{I}_V(E)]_F \circ [\cdot]_E.$$

This equation is not of much practical use, but it does give more insight; it is a good sanity check that the identity on V is involved in changing bases, since representing a vector with different bases does not change the vector itself.

Theorem 2.64. (\star). Let V be a vector space. The identity function $\mathbf{I}_V : V \rightarrow V$ on V satisfies $(\mathbf{I}_V)_{E,F}^{-1} = (\mathbf{I}_V)_{F,E}$. As a corollary, we have $[\mathbf{E}]_F^{-1} = [\mathbf{F}]_E$.

Proof. Given any bases E, F of V , Theorem 2.56 defines $\mathbf{f}_{E,F} := [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}$. Therefore $(\mathbf{I}_V)_{E,F} = [\cdot]_F \circ [\cdot]_E^{-1}$. Since the definition of $\mathbf{f}_{E,F}$ holds for any two bases of V , we can switch E and F to obtain $(\mathbf{I}_V)_{F,E} = [\cdot]_E \circ [\cdot]_F^{-1}$. The claim follows.

We obtain the corollary by starting with $(\mathbf{I}_V)_{E,F}^{-1} = (\mathbf{I}_V)_{F,E}$ and taking the primitive matrices of each side, relative to E and F , respectively. \square

Theorem 2.65. (Change of basis for linear functions). Let V and W be finite-dimensional vector spaces. Let E, G be bases of V , let F, H be bases of W , and consider a linear function $\mathbf{f} : V \rightarrow W$. Then $\mathbf{f}_{E,F}$ and $\mathbf{f}_{G,H}$ are related by

$$\mathbf{f}_{G,H} = [\cdot]_H \circ [\cdot]_{F^{-1}} \circ \mathbf{f}_{E,F} \circ [\cdot]_E \circ [\cdot]_{G^{-1}}.$$

This is because $\mathbf{f}_{E,F}$ was defined as $\mathbf{f}_{E,F} := [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}$. (See Theorem 2.56). But $[\cdot]_H \circ [\cdot]_{F^{-1}} = (\mathbf{I}_W)_{F,H}$ and $[\cdot]_E \circ [\cdot]_{G^{-1}} = (\mathbf{I}_V)_{G,F}$, so

$$\mathbf{f}_{G,H} = (\mathbf{I}_W)_{F,H} \circ \mathbf{f}_{E,F} \circ (\mathbf{I}_V)_{G,F}.$$

We now translate the above equation into a statement about primitive matrices. Since the primitive matrix of a composition of functions is the product of matrices taken relative to the appropriate bases (see Theorem 2.60), we have

$$[\mathbf{f}(G)]_H = [\mathbf{I}_W(F)]_H [\mathbf{f}(E)]_F [\mathbf{I}_V(G)]_F = [\mathbf{F}]_H [\mathbf{f}(E)]_F [\mathbf{G}]_F = [\mathbf{F}]_H [\mathbf{f}(E)]_F [\mathbf{F}]_G^{-1}.$$

The last equality follows from the previous theorem.

Theorem 2.66. (Change of basis for linear functions for a common special case).

Consider the context of the previous theorem. In the special but common case when $E = G$, and $F = H$, we have

$$[\mathbf{f}(E)]_E = [\mathbf{F}]_F [\mathbf{f}(E)]_E [\mathbf{F}]_F^{-1}.$$

I have never ever actually seen the previous theorem used (or even stated). The theorem that has just been stated is what people refer to when they speak of changing the bases of a linear function's matrix.

Theorem 2.67. (★ Change of basis in terms of basis vectors ★).

Let V be a finite-dimensional vector space with bases E and F . By the definition of $[\cdot]_F$, we have

$$\mathbf{f}_i = \sum_{j=1}^n ([\mathbf{f}_i]_E)_j \mathbf{e}_j = \sum_{j=1}^n ([\mathbf{F}]_E)_{ji} \mathbf{e}_j$$

In the last equality, we have used that $[\mathbf{f}_i]_E$ is the i th column of $[\mathbf{F}]_E$.

Remark 2.68. (On the order of proving change of basis theorems). Most linear algebra texts first prove the previous theorem and use it to show a version of the first equation in the box of Theorem 2.63. This approach for proving Theorem 2.63 was not used because it involves quite a bit more matrix algebra than the approach supplied in this text. However, it good to know that these theorems are equivalent.

2.3 The dot product and cross product

The dot product on \mathbb{R}^n and the cross product on \mathbb{R}^3 are almost never explained satisfactorily.

There are two common pedagogical problems with the dot product. The most common problem is defining the dot product as $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i$ and then showing that this initial definition implies $\mathbf{v}_1 \cdot \mathbf{v}_2 = \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$ by using the law of cosines. This is the wrong way of doing things for two reasons: firstly, there is much more motivation (such as further investigation of vector projections or the physical concept of work) for defining $\mathbf{v}_1 \cdot \mathbf{v}_2 := \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$ and then proving $\mathbf{v}_1 \cdot \mathbf{v}_2 = \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i$, rather than starting with $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i$. Secondly, the law of cosines gives no intuition. There is a much better way to prove that $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i$ implies $\mathbf{v}_1 \cdot \mathbf{v}_2 = \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$, which we present. The second common problem arises when an author *does* decide to start with the pedagogically correct definition, $\mathbf{v}_1 \cdot \mathbf{v}_2 := \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$. Authors will prove that $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i$ follows as a result of the law of cosines. Again, using the law of cosines gives no intuition. Instead, $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i$ should be proved by showing and then using the bilinearity of the dot product. The law of cosines should never be used in proving the equivalence of the two dot product formulas. When the equivalence between these two formulas is shown correctly, the law of cosines can be shown as a consequence.

The cross product also comes with two common pedagogical problems. The first is that the complicated algebraic formula for the cross product is rarely explained. The second problem is that the “right hand rule” is never explicitly formalized. One common “explanation” for the right hand rule goes as follows: “you can use a ‘left hand rule’ if you want to, but then you’ll have to account for a minus sign”. This is a true statement, but it only relates the “right hand rule” with the “left hand rule”- it does not explain the fundamental reason why a right hand rule or left hand rule would emerge in the first place. The “right hand rule” is really a consequence of the concept of *orientation*, which is discussed in Chapter 5.

The dot product on \mathbb{R}^n

Definition 2.69. (Length of a vector in \mathbb{R}^n). Let V be an n -dimensional vector space, and let $\hat{\mathbf{e}}$ be the standard basis for \mathbb{R}^n . In analogy to the Pythagorean theorem, we define the *length* of a vector $\mathbf{v} \in \mathbb{R}^n$ to be $\|\mathbf{v}\| := \sqrt{\sum_{i=1}^n ([\mathbf{v}]_{\hat{\mathbf{e}}})_i^2}$.

Definition 2.70. (Unit vector hat notation). We define $\wedge : \mathbb{R}^n \rightarrow \mathbb{R}^n$ to be the function $\wedge(\mathbf{v}) = \frac{\mathbf{v}}{\|\mathbf{v}\|}$, and denote $\wedge(\mathbf{v}) := \hat{\mathbf{v}}$.

Definition 2.71. (Unsigned angle between vectors in \mathbb{R}^2).

Let vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ have the same length $r = \|\mathbf{v}_1\| = \|\mathbf{v}_2\|$, and consider the circle that results when the initial points of $\mathbf{v}_1, \mathbf{v}_2$ coincide. Further consider the smaller of the two arc lengths s enclosed between \mathbf{v}_1 and \mathbf{v}_2 . We define the *unsigned angle* θ between \mathbf{v}_1 and \mathbf{v}_2 to be the ratio $\theta := \frac{s}{r}$. Note that we use the descriptor “unsigned” because $\theta \geq 0$.

Definition 2.72. (\perp operator on \mathbb{R}^2). We define $\perp: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ to be the rotation that rotates a vector \mathbf{v} counterclockwise by $\frac{\pi}{2}$ radians. Specifically, the primitive matrix of \perp relative to E is

$$\begin{pmatrix} \cos(\frac{\pi}{2}) & -\sin(\frac{\pi}{2}) \\ \sin(\frac{\pi}{2}) & \cos(\frac{\pi}{2}) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Note, this definition makes $\{\mathbf{v}, \mathbf{v}_\perp\}$ positively oriented. We denote $\perp(\mathbf{v}) := \mathbf{v}_\perp$.

Remark 2.73. When we consider $\wedge: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and $\perp: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, we have $\perp \circ \wedge = \wedge \circ \perp$, so writing $\hat{\mathbf{v}}_\perp$ is unambiguous when $\mathbf{v} \in \mathbb{R}^2$.

Definition 2.74. (Orthogonal linear function on \mathbb{R}^n).

We say that a linear function $\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is *orthogonal* iff it preserves length and angle, i.e., iff

- $\|\mathbf{v}\| = \|\mathbf{f}(\mathbf{v})\|$ for all $\mathbf{v} \in \mathbb{R}^n$
- the angle between \mathbf{v}_1 and \mathbf{v}_2 is equal to the angle between $\mathbf{f}(\mathbf{v}_1)$ and $\mathbf{f}(\mathbf{v}_2)$ for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$.

Example 2.75. Rotations are orthogonal linear functions.

Definition 2.76. (Vector projection). Let V be a vector space over K , and consider vectors $\mathbf{v}_1, \mathbf{v}_2, (\mathbf{v}_2)_\perp \in V$.

The *vector projection* of \mathbf{v}_1 onto \mathbf{v}_2 is the unique vector $\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2) := (v_1)_\parallel \hat{\mathbf{v}}_2$ such that $\mathbf{v}_1 = (v_1)_\parallel \hat{\mathbf{v}}_2 + (v_1)_\perp (\hat{\mathbf{v}}_2)_\perp$, where $(v_1)_\parallel, (v_1)_\perp \in K$.

Remark 2.77. Note that, on \mathbb{R}^2 , we have $\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2) = \text{proj}(\mathbf{v}_1 \rightarrow \hat{\mathbf{v}}_2)$ because $\hat{\hat{\mathbf{v}}}_2 = \hat{\mathbf{v}}_2$.

Lemma 2.78. (Rotated projection is projection of rotated vectors). Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$. If \mathbf{f} is a rotation, then $\|\mathbf{f}(\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2))\| = \|\text{proj}(\mathbf{f}(\mathbf{v}_1) \rightarrow \mathbf{f}(\mathbf{v}_2))\|$.

Proof. We have $\mathbf{v}_1 = (v_1)_\parallel \hat{\mathbf{v}}_2 + (v_1)_\perp (\hat{\mathbf{v}}_2)_\perp$, so $\mathbf{f}(\mathbf{v}_1) = (v_1)_\parallel \mathbf{f}(\hat{\mathbf{v}}_2) + (v_1)_\perp \mathbf{f}((\hat{\mathbf{v}}_2)_\perp)$. The claim follows if we show (1) that $\mathbf{f}(\hat{\mathbf{v}}_2) = \widehat{\mathbf{f}(\mathbf{v}_2)}$ and (2) that $\mathbf{f}((\hat{\mathbf{v}}_2)_\perp) = \widehat{\mathbf{f}(\mathbf{v}_2)_\perp}$. (1) is true because rotations are length-preserving. (2) is true because rotations commute with each other and because rotations are length-preserving. \square

Lemma 2.79. (Length of a projection). Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$. Then the length of the projection of \mathbf{v}_1 onto \mathbf{v}_2 is $\|\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)\| = \|\mathbf{v}_1\| \cos(\theta)$, where θ is the unsigned angle between \mathbf{v}_1 and \mathbf{v}_2 .

Proof. The lemma holds in the special case when $\mathbf{v}_2 = \hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; draw a right triangle to see this.

For the general case, consider the rotation \mathbf{f} that satisfies $\mathbf{f}(\hat{\mathbf{v}}_2) = \hat{\mathbf{e}}_1$, that is, $\mathbf{f}(\mathbf{v}_2) = \|\mathbf{v}_2\| \hat{\mathbf{e}}_1$. Then because rotations are length-preserving and with use of the previous lemma, $(v_1)_\parallel = \|\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)\| = \|\mathbf{f}(\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2))\| = \|\text{proj}(\mathbf{f}(\mathbf{v}_1) \rightarrow \mathbf{f}(\mathbf{v}_2))\|$. This is the same as $\|\text{proj}(\mathbf{f}(\mathbf{v}_1) \rightarrow \|\mathbf{v}_2\| \hat{\mathbf{e}}_1)\| = \|\text{proj}(\mathbf{f}(\mathbf{v}_1) \rightarrow \hat{\mathbf{e}}_1)\| = \|\mathbf{f}(\mathbf{v}_1)\| \cos(\phi)$, where ϕ is the unsigned angle between $\mathbf{f}(\mathbf{v}_1)$ and $\hat{\mathbf{e}}_1$. We have $\|\mathbf{f}(\mathbf{v}_1)\| = \|\mathbf{v}_1\|$ because rotations are length-preserving, and $\phi = \theta$, where θ is the unsigned angle from \mathbf{v}_1 to \mathbf{v}_2 , because rotations preserve angle. Therefore $\|\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)\| = \|\mathbf{v}_1\| \cos(\theta)$. \square

Definition 2.80. (Geometric dot product on \mathbb{R}^2). The *geometric dot product on \mathbb{R}^2* is the function $\cdot: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$\mathbf{v}_1 \cdot \mathbf{v}_2 := \|\mathbf{v}_2\| \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2).$$

Why do we care about the geometric dot product? The primary reason is that $\mathbf{v}_1 \cdot \mathbf{v}_2 = \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)$ when $\|\mathbf{v}_2\| = 1$, so investigating the geometric dot product can tell us more about vector projections (we will indeed find that the dot product tells us something about projections in Theorem 2.98!). In

physics, the dot product is also used to define the work done by a force \mathbf{F} along a displacement $\Delta\mathbf{x}$: (work done by \mathbf{F}) $:= \mathbf{F} \cdot \Delta\mathbf{x}$.

Since $\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2) = \|\mathbf{v}_1\| \cos(\theta)$, where θ is the unsigned angle from \mathbf{v}_1 to \mathbf{v}_2 , the geometric dot product can also be written as

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta).$$

Remark 2.81. Vectors in \mathbb{R}^2 are perpendicular when their geometric dot product is zero, since $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ implies $\theta = \frac{\pi}{2}$.

Lemma 2.82. (Projection onto a vector is a linear function).

Let V be a vector space over K , and let $\mathbf{v}_1, \mathbf{v}_2 \in V$. Then the map $\mathbf{v}_1 \mapsto \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)$ is linear.

Proof. Define $\mathbf{f} : V \rightarrow K$ by $\mathbf{f}(\mathbf{v}_1) = \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)$. We show $\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2)$ and $\mathbf{f}(c\mathbf{v}_1) = c\mathbf{f}(\mathbf{v}_1)$.

$$\begin{aligned} \mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) &= \mathbf{f}\left(\left((v_1)_{||}\hat{\mathbf{v}}_2 + (v_1)_{\perp}(\hat{\mathbf{v}}_2)_{\perp}\right) + \left((v_2)_{||}\hat{\mathbf{v}}_2 + (v_2)_{\perp}(\hat{\mathbf{v}}_2)_{\perp}\right)\right) \\ &= \mathbf{f}\left(\left((v_1)_{||} + (v_2)_{||}\right)\hat{\mathbf{v}}_2 + \left((v_1)_{\perp} + (v_2)_{\perp}\right)(\hat{\mathbf{v}}_2)_{\perp}\right) \\ &= \text{proj}\left(\left[\left((v_1)_{||} + (v_2)_{||}\right)\hat{\mathbf{v}}_2 + \left((v_1)_{\perp} + (v_2)_{\perp}\right)(\hat{\mathbf{v}}_2)_{\perp}\right] \rightarrow \mathbf{v}_2\right) \\ &= \left((v_1)_{||} + (v_2)_{||}\right)\hat{\mathbf{v}}_2 = (v_1)_{||}\hat{\mathbf{v}}_2 + (v_2)_{||}\hat{\mathbf{v}}_2 = \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2) + \text{proj}(\mathbf{v}_2 \rightarrow \mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2). \end{aligned}$$

$$\begin{aligned} \mathbf{f}(c\mathbf{v}_1) &= \mathbf{f}\left(c\left((v_1)_{||}\hat{\mathbf{v}}_2 + (v_1)_{\perp}(\hat{\mathbf{v}}_2)_{\perp}\right)\right) = \mathbf{f}\left(c(v_1)_{||}\hat{\mathbf{v}}_2 + c(v_1)_{\perp}(\hat{\mathbf{v}}_2)_{\perp}\right) = \text{proj}\left(\left(c(v_1)_{||}\hat{\mathbf{v}}_2 + c(v_1)_{\perp}(\hat{\mathbf{v}}_2)_{\perp}\right) \rightarrow \mathbf{v}_2\right) \\ &= c(v_1)_{||}\hat{\mathbf{v}}_2 = c\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2) = c\mathbf{f}(\mathbf{v}_1). \end{aligned}$$

□

Theorem 2.83. (The geometric dot product on \mathbb{R}^2 is a bilinear function).

The geometric dot product on \mathbb{R}^2 is a bilinear function. That is, $(\mathbf{v}_1, \mathbf{v}_2) \mapsto \mathbf{v}_1 \cdot \mathbf{v}_2$ is linear in the argument \mathbf{v}_1 when \mathbf{v}_2 is fixed, and is linear in the argument \mathbf{v}_2 when \mathbf{v}_1 is fixed.

Proof. The geometric dot product is symmetric, so it suffices to show that it is a linear function in either argument; it suffices to show that $\mathbf{f} : V \rightarrow K$ defined by $\mathbf{f}(\mathbf{v}_1) = \mathbf{v}_1 \cdot \mathbf{v}_2$ is a linear function. Well, $\mathbf{f}(\mathbf{v}_1) = \|\mathbf{v}_2\| \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)$, where $\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)$ is linear in \mathbf{v}_1 . Therefore, since \mathbf{f} is the result of scaling a linear function by $\|\mathbf{v}_2\|$, it too is a linear function. □

Remark 2.84. Because projection in \mathbb{R}^2 can now be defined in terms of the geometric dot product on \mathbb{R}^2 , we can note that $\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)$ is linear in $\mathbf{v}_1, \mathbf{v}_2$ when $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$.

Lemma 2.85. Consider $\mathbf{v} \in \mathbb{R}^2$ with $\|\mathbf{v}\| = 1$. Then, in applying the formula $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2 \cos(\theta)$, we have

$$\begin{aligned} \mathbf{v} \cdot \mathbf{v} &= 1 \\ \mathbf{v} \cdot \mathbf{v}_{\perp} &= 0. \end{aligned}$$

Derivation 2.86. (Algebraic dot product on \mathbb{R}^2). We can now derive an “algebraic” formula for the dot product, using its bilinearity (Theorem 2.83) together with the previous lemma.

If V is a finite-dimensional vector space over a field K with a basis $E = \{\mathbf{e}_i\}_{i=1}^n$, then a bilinear function $B : V \times V \rightarrow K$ satisfies

$$B(\mathbf{v}_1, \mathbf{v}_2) = \sum_{i=1}^n \sum_{j=1}^n ([\mathbf{v}_1]_E)_i ([\mathbf{v}_2]_E)_i B(\mathbf{e}_i, \mathbf{e}_j).$$

To see this, use bilinearity to “expand” each argument of B ; do so one argument at a time. The geometric dot product is a bilinear function $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$, so the above implies

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \sum_{i=1}^2 \sum_{j=1}^2 ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i \delta_{ij} = \sum_{i=1}^2 ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i.$$

Therefore

$$\boxed{\mathbf{v}_1 \cdot \mathbf{v}_2 = ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_1 ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_1 + ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_2 ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_2}$$

Lemma 2.87. (Orthogonal linear functions on \mathbb{R}^2 preserve algebraic dot product).

Let $\mathbf{f} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be an orthogonal linear function. Then \mathbf{f} preserves the algebraic dot product on \mathbb{R}^2 : $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{f}(\mathbf{v}_1) \cdot \mathbf{f}(\mathbf{v}_2)$.

Proof. By definition, orthogonal linear functions on \mathbb{R}^2 preserve length. We show that the dot product $\mathbf{v}_1 \cdot \mathbf{v}_2$ depends on the lengths $\|\mathbf{v}_1\|$ and $\|\mathbf{v}_2\|$; we claim that the following equation holds:

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \frac{1}{2} (\|\mathbf{v}_1 + \mathbf{v}_2\|^2 - (\|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2)).$$

We now prove this equation. Note that $\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$ for $\mathbf{v} \in \mathbb{R}^2$. So $\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = (\mathbf{v}_1 + \mathbf{v}_2) \cdot (\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{v}_1 \cdot \mathbf{v}_1 + 2\mathbf{v}_1 \cdot \mathbf{v}_2 + \mathbf{v}_2 \cdot \mathbf{v}_2 = \|\mathbf{v}_1\|^2 + 2\mathbf{v}_1 \cdot \mathbf{v}_2 + \|\mathbf{v}_2\|^2$. Solve for $\mathbf{v}_1 \cdot \mathbf{v}_2$ to obtain the above equation.

Since the algebraic dot product is a function of length, which is preserved by \mathbf{f} , then the algebraic dot product is also preserved by \mathbf{f} . \square

Theorem 2.88. (Algebraic dot product on \mathbb{R}^2 formula implies geometric dot product formula on \mathbb{R}^2).

We’ve used the bilinearity of the geometric dot product to prove the algebraic dot product formula. Now we show that we can derive the geometric dot product formula from the algebraic dot product formula. More specifically if $\hat{\mathbf{e}}$ is the standard basis of \mathbb{R}^2 , defining $\cdot : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ by $\mathbf{v}_1 \cdot \mathbf{v}_2 = ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_1 ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_1 + ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_2 ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_2$ implies $\mathbf{v}_1 \cdot \mathbf{v}_2 = \|\mathbf{v}_2\| \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2) = \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$, where θ is the unsigned angle between \mathbf{v}_1 and \mathbf{v}_2 .

Proof. Consider $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$, and let \mathbf{f} be the rotation satisfying $\mathbf{f}(\hat{\mathbf{v}}_2) = \hat{\mathbf{e}}_1$, that is, $\mathbf{f}(\mathbf{v}_2) = \|\mathbf{v}_2\| \hat{\mathbf{e}}_1$.

Orthogonal linear functions on \mathbb{R}^2 preserve the algebraic dot product on \mathbb{R}^2 (see the previous lemma), so

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{f}(\mathbf{v}_1) \cdot \mathbf{f}(\mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) \cdot \|\mathbf{v}_2\| \hat{\mathbf{e}}_1 = \begin{pmatrix} ([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathbf{e}}})_1 \\ ([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathbf{e}}})_2 \end{pmatrix} \cdot \begin{pmatrix} \|\mathbf{v}_2\| \\ 0 \end{pmatrix} = \|\mathbf{v}_2\| ([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathbf{e}}})_1. \quad \text{We have}$$

$$([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathbf{e}}})_1 = \text{proj}(\mathbf{f}(\mathbf{v}_1) \rightarrow \hat{\mathbf{e}}_1) = \text{proj}(\mathbf{f}(\mathbf{v}_1) \rightarrow \mathbf{f}(\hat{\mathbf{v}}_2)) = \text{proj}(\mathbf{f}(\mathbf{v}_1) \rightarrow \mathbf{f}(\mathbf{v}_2)) = \text{proj}(\mathbf{f}(\mathbf{v}_1) \rightarrow \widehat{\mathbf{f}(\mathbf{v}_2)}) \\ = \text{proj}(\mathbf{f}(\mathbf{v}_1) \rightarrow \mathbf{f}(\mathbf{v}_2)) = \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2), \text{ where the last equality is by Lemma 2.78.}$$

Therefore $\mathbf{v}_1 \cdot \mathbf{v}_2 = \|\mathbf{v}_2\| \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)$, which is the definition of the geometric dot product on \mathbb{R}^2 . \square

Remark 2.89. Most proofs of the above theorem use the law of cosines. I personally do not find the law of cosines intuitive, and believe it is best seen as a consequence of the equivalence between the geometric and algebraic dot product formulas. We prove the law of cosines in this way in Theorem 2.96.

Definition 2.90. (Dot product on \mathbb{R}^n). Now that we have motivated the algebraic dot product on \mathbb{R}^2 by proving the previous theorem, we have a sensible way to define a dot product on \mathbb{R}^n . We can't do this by generalizing the geometric dot product on \mathbb{R}^2 because it is not clear how to define the concept of “angle” in \mathbb{R}^n .

Let $\hat{\mathbf{e}}$ be the standard basis of \mathbb{R}^n . We define $\cdot : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i$$

Theorem 2.91. (Dot product on \mathbb{R}^n as matrix-matrix product). $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1^\top \mathbf{v}_2$. (Proof left as exercise).

Theorem 2.92. (Length in \mathbb{R}^n). We can notice that in \mathbb{R}^3 , the length of a vector $\mathbf{v} \in \mathbb{R}^3$ expressed relative to the standard basis $\hat{\mathbf{e}} = \{\hat{\mathbf{e}}_i\}_{i=1}^3$ is $\sqrt{\sum_{i=1}^3 ([\mathbf{v}]_{\hat{\mathbf{e}}})_i^2} = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

This motivates us to define the length of a vector $\mathbf{v} \in \mathbb{R}^n$ to be $\|\mathbf{v}\| := \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

Definition 2.93. (Angle in \mathbb{R}^n). The dot product on \mathbb{R}^2 satisfies $\mathbf{v}_1 \cdot \mathbf{v}_2 = \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$, so $\theta = \cos^{-1} \left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} \right)$. We define angle in \mathbb{R}^n in analogy to the dot product on \mathbb{R}^2 . The angle between vectors in \mathbb{R}^n is $\theta := \cos^{-1} \left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} \right)$, where the dot product here is the dot product on \mathbb{R}^n .

The angle $\theta := \cos^{-1} \left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} \right)$ does indeed satisfy the condition $\theta \in [0, 2\pi)$. In Theorem 4.16, we prove this in a more general setting.

Remark 2.94. (Geometric dot product on \mathbb{R}^n).

With the previous definition of angle in \mathbb{R}^n , we have $\mathbf{v}_1 \cdot \mathbf{v}_2 = \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$ for $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, which looks like the formula for the geometric dot product on \mathbb{R}^2 .

Definition 2.95. (Orthogonality of vectors in \mathbb{R}^n). We say that vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ are *orthogonal* iff the angle in \mathbb{R}^n between \mathbf{v}_1 and \mathbf{v}_2 is $\frac{\pi}{2}$. So, orthogonality is a generalized notion of perpendicularity. Thus, $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ are orthogonal iff $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Theorem 2.96. (Law of cosines in \mathbb{R}^n).

Consider vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$. We can interpret $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_1 - \mathbf{v}_2$ as the oriented side lengths of a triangle; then, the angle θ between \mathbf{v}_2 and \mathbf{v}_1 is the angle opposite to the side \mathbf{u} .

The “law of cosines” is the fact that $\|\mathbf{v}_1 - \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 - 2\|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$. Note that by using $\theta = 0$, we recover the Pythagorean theorem.

The only reason this theorem was included was to demonstrate the point of Remark 2.89.

Proof. $\|\mathbf{v}_1 - \mathbf{v}_2\|^2 = (\mathbf{v}_1 - \mathbf{v}_2) \cdot (\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{v}_1 \cdot \mathbf{v}_1 - 2\mathbf{v}_1 \cdot \mathbf{v}_2 + \mathbf{v}_2 \cdot \mathbf{v}_2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 - 2\|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$. \square

Remark 2.97. The above theorem reveals that the algebraic dot product on \mathbb{R}^2 can also be discovered as an orthogonality condition between vectors. When $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ are orthogonal, they form a right triangle, so Pythagorean theorem gives $\|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 = \|\mathbf{v}_1 - \mathbf{v}_2\|^2$. Use $\mathbf{v}_1 = \sqrt{\sum_{i=1}^2 (([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i)^2}$ to discover that we must have $([\mathbf{v}_1]_{\hat{\mathbf{e}}})_1 + ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_2 = 0$.

Theorem 2.98. (Vector projection in terms of algebraic dot product).

Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$. We can express the projection of \mathbf{v}_1 onto \mathbf{v}_2 in terms of the dot product $\mathbf{v}_1 \cdot \mathbf{v}_2$. (Recall Definition 2.76 for the definition of vector projection).

Recall from Lemma 2.79 that $\|\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)\| = \|\mathbf{v}_1\| \cos(\theta)$. Since $\mathbf{v}_1 \cdot \mathbf{v}_2 = \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$, then $\|\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)\| = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|}$. Thus

$$\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|} \hat{\mathbf{v}}_2 = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2.$$

Theorem 2.99. (Vector projection in \mathbb{R}^n is a bilinear function).

The geometric dot product on \mathbb{R}^2 is a bilinear function. That is, $(\mathbf{v}_1, \mathbf{v}_2) \mapsto \text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)$ is linear in the argument \mathbf{v}_1 when \mathbf{v}_2 is fixed, and is linear in the argument \mathbf{v}_2 when \mathbf{v}_1 is fixed.

Proof. This follows from the fact that vector projection in \mathbb{R}^n can be expressed by using the dot product on \mathbb{R}^n , which is a bilinear function.

In general, vector projection in any finite-dimensional vector space *with inner product* is a bilinear function, because an inner product is by definition a bilinear function. We will discuss inner products and vector spaces with inner product in Chapter 4. \square

Cross product

See [...]

Part I

Multilinear algebra and tensors

Preview of differential forms

Explain diff forms on an intuitive level to someone who's just taken multivariable calc. When a new concept comes up, mention its name and link to the relevant theorems in the appropriate chapter.

3

A motivated introduction to tensors

There are two key ideas that we must formalize before we define what a “tensor” is.

One of the ideas is that of a “multilinear element”. Recall that elements of vector spaces (vectors) can be thought of as “linear elements” because linear functions respect the decomposition of vectors. After defining the notion of *multilinear function*, we will see there is a corresponding notion of “multilinear element”, and that these multilinear elements are elements of *tensor product spaces*. There are two main contribution of tensor product spaces to the overarching theory of tensors: tensor product spaces formalize the structure of how “multilinear things” behave, and they allow multilinear functions to be treated as linear functions. (Tensor product spaces do not account for the entire theory of tensors, though, even though the name might make you think this. One more key idea, described below, is required).

The other idea is to think of linear functions as vectors (as elements of vector spaces). This is achieved by decomposing linear functions into linear combinations of simpler linear functions. Most introductory linear algebra classes approach this idea by proving the fact that the set of $m \times n$ matrices form a vector space. We take this idea and run with it to obtain the theorem which underlies the definition of a “ $\binom{p}{q}$ tensor”.

3.1 Multilinear functions and tensor product spaces

Definition 3.1. (Multilinear function).

Let V_1, \dots, V_k, W be vector spaces over a field K . We say a function $\mathbf{f} : V_1 \times \dots \times V_k \rightarrow W$ is a *k-linear function* iff for all $\mathbf{v}_1 \in V_1, \dots, \mathbf{v}_i \in V_i, \dots, \mathbf{v}_k \in V_k$, the function $\mathbf{f}_i : V_i \rightarrow W$ defined by $\mathbf{f}_i(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_k) = \mathbf{f}(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_k)$ is linear. In other words, \mathbf{f} is *k-linear* iff it is “linear in each argument”.

When k is clear from the context, *k-linear functions* are called *multilinear functions*. A 2-linear function is called a *bilinear function*.

Example 3.2. (Examples of multilinear functions). The dot product on \mathbb{R}^n is a bilinear function on $\mathbb{R}^n \times \mathbb{R}^n$. If you have encountered the determinant before, you might recall that it is a multilinear function.

Definition 3.3. (Vector space of multilinear functions).

If V_1, \dots, V_k, W are vector spaces over a field K , then we use $\mathcal{L}(V_1 \times \dots \times V_k \rightarrow W)$ to denote the vector space over K formed by the set of *k-linear functions* $V_1 \times \dots \times V_k \rightarrow W$ under the operations of function addition and function scaling. In particular, $\mathcal{L}(V_i \rightarrow W)$ denotes the set of linear functions $V_i \rightarrow W$. (The proof that $\mathcal{L}(V_1 \times \dots \times V_k \rightarrow W)$ is indeed a vector space is left as an exercise).

Elements of a vector space can be considered to be “linear elements” because their decompositions relative to a basis are respected by linear functions (see Definition 2.23). We have just been introduced to the notion of a multilinear function. A natural question is then, “what is a reasonable definition of ‘multilinear element’?” We will see that elements of tensor product spaces are “multilinear elements”.

Definition 3.4. (Tensor product space).

Let V_1, \dots, V_k be finite-dimensional vector spaces over a field K . The *tensor product space* $V_1 \otimes \dots \otimes V_k$ is defined to be the vector space over K whose elements are from the set

$$\left\{ (\mathbf{v}_1, \dots, \mathbf{v}_k) \mid \mathbf{v}_i \in V_i, i \in \{1, \dots, k\} \right\}.$$

We write $\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k$ to mean $(\mathbf{v}_1, \dots, \mathbf{v}_k)$. The elements $\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k$ are also subject to addition and scalar multiplication operators defined as follows:

$$\begin{aligned} & \mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_{i-1} \otimes \mathbf{v}_{i1} \otimes \mathbf{v}_{i+1} \dots \otimes \mathbf{v}_k \\ & \quad + \\ & \mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_{i-1} \otimes \mathbf{v}_{i2} \otimes \mathbf{v}_{i+1} \dots \otimes \mathbf{v}_k \\ & \quad = \\ & \mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_{i-1} \otimes (\mathbf{v}_{i1} + \mathbf{v}_{i2}) \otimes \mathbf{v}_{i+1} \dots \otimes \mathbf{v}_k \end{aligned}$$

and

$$\begin{aligned} & c(\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_{i-1} \otimes \mathbf{v}_i \otimes \mathbf{v}_{i+1} \dots \otimes \mathbf{v}_k) \\ & \quad = \\ & \mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_{i-1} \otimes (c\mathbf{v}_i) \otimes \mathbf{v}_{i+1} \dots \otimes \mathbf{v}_k. \end{aligned}$$

These operations were contrived to be such that the “comma in disguise” \otimes appears to be a multilinear function. We did this because we want elements of tensor product spaces to be “multilinear elements”.

When the context is clear, we will refer to elements of tensor product spaces as “tensors”.

Remark 3.5. (Tensor terminology). Some authors use the word “tensor” to mean “ $\binom{p}{q}$ tensor”. (We have not defined $\binom{p}{q}$ tensors yet, but we will in Definition 3.29). We will use the word “tensor” to either mean an element of a tensor product space or a $\binom{p}{q}$ tensor, but we only do this when the meaning is clear from context.

Definition 3.6. (Elementary tensor). Let V_1, \dots, V_k be vector spaces, and consider the tensor product space $V_1 \otimes \dots \otimes V_k$. An element of $V_1 \otimes \dots \otimes V_k$ that is of the form $\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k$ is called an *elementary tensor*. Intuitively, an elementary tensor is an element that is *not* a linear combination of two or more other nonzero tensors. An element of $V_1 \otimes \dots \otimes V_k$ that is not an elementary tensor is called a *nonelementary tensor*.

Theorem 3.7. (Associativity of tensor product).

Let V_1, V_2, V_3 be vector spaces. Then there are natural isomorphisms

$$(V_1 \otimes V_2) \otimes V_3 \cong V_1 \otimes V_2 \otimes V_3 \cong V_1 \otimes (V_2 \otimes V_3).$$

That is, these spaces are “the same”, since an element of one can “naturally” be identified as an element of the other. (See Definition 2.32 for a discussion of linear isomorphisms). These identifications are “natural” in the sense that they do not depend on a choice of basis (see Definition 2.38).

Proof. Since an isomorphism of vector spaces is a linear map, it is enough to define an isomorphism on elementary tensors and “extend with linearity”. To construct these isomorphisms, we will recall the definition of a tensor product space as a quotient space, so that elementary tensors of $(V_1 \otimes V_2) \otimes V_3$ are of the form $((\mathbf{v}_1, \mathbf{v}_2), \mathbf{v}_3)$, elementary tensors of $V_1 \otimes V_2 \otimes V_3$ are of the form $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$, and elementary tensors of $V_1 \otimes (V_2 \otimes V_3)$ are of the form $(\mathbf{v}_1, (\mathbf{v}_2, \mathbf{v}_3))$. For the first isomorphism, we send

$((\mathbf{v}_1, \mathbf{v}_2), \mathbf{v}_3) \mapsto (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$, and for (reverse of) the second isomorphism, we send $(\mathbf{v}_1, (\mathbf{v}_2, \mathbf{v}_3)) \mapsto (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. We leave it to the reader to check that these maps are indeed linear and injective; surjectivity follows from fact that these maps “extend with linearity”. When extending with linearity, it will be necessary to use the fact that \otimes (that is, the outermost comma) appears to be a multilinear function. \square

Theorem 3.8. (Basis and dimension of a tensor product space).

Let V_1, \dots, V_k be finite-dimensional vector spaces with bases E_1, \dots, E_k , respectively, where $E_i = \{\mathbf{e}_{i1}, \dots, \mathbf{e}_{in_i}\}$, and where $\dim(V_i) = n_i$. Then $V_1 \otimes \dots \otimes V_k$ is a $n_1 n_2 \dots n_k$ dimensional vector space with basis

$$\{\mathbf{e}_{i1} \otimes \dots \otimes \mathbf{e}_{in_i} \mid i \in \{1, \dots, k\}\}.$$

Proof. This set spans $V_1 \otimes \dots \otimes V_k$ by definition of $V_1 \otimes \dots \otimes V_k$. To show linear independence, assume that $\sum_{i=1}^n c_{i1} \dots c_{in} \mathbf{e}_{i1} \otimes \dots \otimes \mathbf{e}_{in_i}$ is the zero tensor. We must show that all of the $c_{i1} \dots c_{in}$ ’s are 0.

come back later

Note, This is the case if and only if some $\mathbf{e}_{ij} = \mathbf{0}$ (prove this as an exercise). \square

Theorem 3.9. (Universal property of the tensor product).

This theorem formalizes the notion that multilinear functions preserve the decomposition of multilinear elements. More precisely, it states that a multilinear function uniquely corresponds to a linear function on a tensor product space, which is a function that preserves the decomposition of an element of a tensor product space.

We state the theorem now. Let V_1, V_2, W be vector spaces, and let $\mathbf{f} : V_1 \times V_2 \rightarrow W$ be a bilinear function. Then there exists a linear function $\mathbf{h} : V_1 \otimes V_2 \rightarrow W$ with $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ that uniquely depends on \mathbf{f} , and where $\mathbf{g} : V_1 \times V_2 \rightarrow V_1 \otimes V_2$.

Proof. First we send $(\mathbf{v}_1, \mathbf{v}_2) \xrightarrow{\mathbf{g}} \mathbf{v}_1 \otimes \mathbf{v}_2$ and then $\mathbf{v}_1 \otimes \mathbf{v}_2 \xrightarrow{\mathbf{h}} \mathbf{f}(\mathbf{v}_1, \mathbf{v}_2)$, where we impose that \mathbf{h} be linear. (Note, requiring that \mathbf{h} is linear implies that $\mathbf{h}(\mathbf{T})$ is indeed defined when \mathbf{T} is a nonelementary tensor, since defining how \mathbf{h} acts on elementary tensors is enough to determine how \mathbf{h} acts on any tensor). We have $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ when we restrict both sides so that they only apply to “elementary” vectors $(\mathbf{v}_1, \mathbf{v}_2) \in V_1 \times V_2$. Using the bilinearity of \mathbf{f} and the seeming-bilinearity of \otimes , we can “extend” this statement to a statement that applies to any vector $(\mathbf{v}_1, \mathbf{v}_2) \in V_1 \times V_2$. Thus $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$. The composition map \circ is well-defined, so $\mathbf{h} = \mathbf{f} \circ \mathbf{g}^{-1}$ is uniquely determined. \square

Theorem 3.10. (Multilinear functions are naturally identified with linear functions on tensor product spaces).

Let V_1, \dots, V_k, W be vector spaces. Then the vector space of multilinear functions $V_1 \times \dots \times V_k \rightarrow W$ is naturally isomorphic to the vector space of linear functions $V_1 \otimes \dots \otimes V_k \rightarrow W$:

$$\mathcal{L}(V_1 \times \dots \times V_k \rightarrow W) \cong \mathcal{L}(V_1 \otimes \dots \otimes V_k \rightarrow W).$$

Proof. We prove the theorem for the case $k = 2$, and show $\mathcal{L}(V_1 \times V_2 \rightarrow W) \cong \mathcal{L}(V_1 \otimes V_2 \rightarrow W)$. The general result follows by using induction with the associativity of the Cartesian product \times of sets and the tensor product \otimes of vector spaces.

To construct a linear isomorphism $\mathcal{L}(V_1 \times V_2 \rightarrow W) \mapsto \mathcal{L}(V_1 \otimes V_2 \rightarrow W)$, we send $\mathbf{f} \in \mathcal{L}(V_1 \times V_2 \rightarrow W) \mapsto \mathbf{h} = \mathbf{f} \circ \mathbf{g}^{-1}$, where \mathbf{g} and \mathbf{h} were defined in the proof of Theorem 3.9. We already know this map is an injection because \mathbf{h} is uniquely determined by \mathbf{f} (see the proof of Theorem 3.9). It is a surjection because, given \mathbf{h} , we can choose \mathbf{f} so that $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ (this is the condition that \mathbf{h} uniquely satisfies); choose \mathbf{f} so that $\mathbf{f}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{h}(\mathbf{v}_1 \otimes \mathbf{v}_2)$. It is linear because, given vector spaces Y, Z, W , the map \circ which composes linear functions, $\circ : \mathcal{L}(Y \rightarrow Z) \times \mathcal{L}(Z \rightarrow W) \rightarrow \mathcal{L}(Y \rightarrow W)$, is a bilinear map. (Check this fact for yourself. The consequences of this are explored in [...]). Therefore the map $\mathcal{L}(V_1 \times V_2 \rightarrow W) \mapsto \mathcal{L}(V_1 \otimes V_2 \rightarrow W)$ described above is a linear isomorphism. \square

3.2 A motivated introduction to $\binom{p}{q}$ tensors

Now we will discover the theorem which generalizes the two key notions (thinking of linear functions as vectors and “multilinear elements”) discussed at the beginning of the chapter. Since we now have familiarity with the first key idea, “accidentally” discovering and formalizing the second idea as we go is hopefully not too ambitious.

The theorem we will discover is that when V and W are finite-dimensional vector spaces, there is a natural isomorphism $\mathcal{L}(V \rightarrow W) \cong W^* \otimes V$, where V^* is the *dual vector space* to V . We can see that the two key ideas (the first being thinking of linear functions as vectors and the second being “multilinear elements”) are represented in this theorem with formal notation: the theorem includes the dual space V^* , which (we will see) indicates that thinking of linear functions as vectors is involved, and it also includes the tensor product \otimes , which indicates that multilinear structure is involved.

To begin this discovery, let V and W be finite-dimensional vector spaces over a field K with bases $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$, respectively, and consider a linear transformation $\mathbf{f} : V \rightarrow W$. We will analyze \mathbf{f} by considering its matrix relative to E and F . This matrix, as is the case with any matrix, is a weighted sum of matrices with a 1 in only one entry and 0's in all other entries. For example, a 3×2 matrix (a_{ij}) is expressed with a weighted sum of this style as

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} = a_{11} \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} + a_{12} \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} + a_{21} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} + a_{22} \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} + a_{31} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} + a_{32} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Let $\hat{\mathbf{e}} = \{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n\}$ be the standard basis of $K^n = K^{\dim(V)}$ and let $\hat{\mathbf{f}} = \{\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_m\}$ be the standard basis of $K^m = K^{\dim(W)}$. So, in the example, $\hat{\mathbf{e}} = \{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ and $\hat{\mathbf{f}} = \{\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \hat{\mathbf{f}}_3\}$. The first “big leap” is to notice that the $m \times n$ matrix with ij entry 1 and all other entries 0 is $\hat{\mathbf{f}}_i \hat{\mathbf{e}}_j^\top$, where $\hat{\mathbf{f}}_i \hat{\mathbf{e}}_j^\top$ is the product of a $m \times 1$ matrix with a $1 \times n$ matrix (see Theorem 2.62). This means that the above 3×2 matrix can be expressed as

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} = a_{11} \hat{\mathbf{f}}_1 \hat{\mathbf{e}}_1^\top + a_{12} \hat{\mathbf{f}}_1 \hat{\mathbf{e}}_2^\top + a_{21} \hat{\mathbf{f}}_2 \hat{\mathbf{e}}_1^\top + a_{22} \hat{\mathbf{f}}_2 \hat{\mathbf{e}}_2^\top + a_{31} \hat{\mathbf{f}}_3 \hat{\mathbf{e}}_1^\top + a_{32} \hat{\mathbf{f}}_3 \hat{\mathbf{e}}_2^\top = \sum_{\substack{i \in \{1,2,3\} \\ j \in \{1,2\}}} a_{ij} \hat{\mathbf{f}}_i \hat{\mathbf{e}}_j^\top.$$

Therefore, the matrix of \mathbf{f} relative to E and F is of the form

$$\sum_{\substack{i \in \{1, \dots, n\} \\ j \in \{1, \dots, m\}}} a_{ij} \hat{\mathbf{f}}_i \hat{\mathbf{e}}_j^\top,$$

for some $a_{ij} \in K$.

What we have done is decompose the matrix of \mathbf{f} relative to E and F relative to the basis $\{\hat{\mathbf{e}}_i \hat{\mathbf{f}}_j^\top\}$ of $m \times n$ matrices. This choice of basis for the space of $m \times n$ matrices stems from our choice of the standard bases $\hat{\mathbf{e}}$ and $\hat{\mathbf{f}}$ for K^n and K^m . Nothing is stopping us from using different bases, however. Suppose $G = \{\mathbf{g}_1, \dots, \mathbf{g}_n\}$ is a basis for K^n and $H = \{\mathbf{h}_1, \dots, \mathbf{h}_m\}$ is a basis for K^m . Then $\{\mathbf{g}_i \mathbf{h}_j^\top\}$ is also a basis of the vector space of $m \times n$ matrices, so the matrix of \mathbf{f} relative to E and F is of the form

$$\sum_{\substack{i \in \{1, \dots, n\} \\ j \in \{1, \dots, m\}}} b_{ij} \mathbf{g}_i \mathbf{h}_j^\top,$$

for some $b_{ij} \in K$.

We now convert this discussion of matrices into a discussion about the linear functions they represent. We started with the matrix (a_{ij}) of a linear function \mathbf{f} relative to bases. But what linear functions do the matrices in the above weighted sum represent?

Consider one of the matrices in the weighted sum, $\mathbf{g}_i \mathbf{h}_j^\top$. Initially, we may be tempted to directly investigate the linear function represented by $\mathbf{g}_i \mathbf{h}_j^\top$. This will work, but we can be even more specific; $\mathbf{g}_i \mathbf{h}_j^\top$ is a matrix-matrix product, so it corresponds to a composition of linear functions (see Theorem 2.58 and possibly Theorem 2.60). Asking “to what linear functions do \mathbf{g}_i and \mathbf{h}_j^\top correspond?” will prove fruitful.

Linear functions are composed from right to left, so we will first consider \mathbf{h}_j^\top . The linear function $K^n \rightarrow K$ represented by the $1 \times n$ matrix \mathbf{h}_j^\top is the function $\phi_{\mathbf{h}_j}$ defined by $\phi_{\mathbf{h}_j}(\mathbf{c}) = \mathbf{h}_j^\top \mathbf{c}$. Note that the image of $\phi_{\mathbf{h}_j}$ is the field K , which is a 1-dimensional vector space. So $\phi_{\mathbf{h}_j}$ is a rank-1 linear map (see 2.27).

Now we consider the $m \times 1$ matrix \mathbf{g}_i . In the matrix-matrix product, \mathbf{g}_i is written to the left of \mathbf{h}_j^\top , so it must accept a scalar as input. The linear map $K \rightarrow K^m$ represented by \mathbf{g}_i is thus $\mathbf{g}_i(c) = c\mathbf{g}_i$, where we have used \mathbf{g}_i on the left hand side to denote the linear map represented by \mathbf{g}_i . Note, the image of the map $\mathbf{g}_i : K \rightarrow K^m$ is $\text{span}(\mathbf{g}_i)$, which is 1-dimensional; \mathbf{g}_i is also a rank-1 linear map.

The matrix-matrix product $\mathbf{g}_i \mathbf{h}_j^\top$ then corresponds to the linear function $\mathbf{g}_i \circ \phi_{\mathbf{h}_j}$, where \mathbf{g}_i again denotes the linear map $K \rightarrow K^m$ defined by $\mathbf{g}_i(c) = c\mathbf{g}_i$. Note that the composition $\mathbf{g}_i \circ \phi_{\mathbf{h}_j}$ of is also a rank-1 linear map.

Overall, we have shown that the matrix with respect to bases of a linear function $\mathbf{f} : V \rightarrow W$ can be expressed as a linear combination of the (primitive (see Derivation 2.51)) matrices that represent the linear maps $\mathbf{g}_i \circ \phi_{\mathbf{h}_j}$. Therefore, the linear function \mathbf{f} is a linear combination of the linear functions $\mathbf{g}_i \circ \phi_{\mathbf{h}_j}$:

$$\mathbf{f} = \sum_{\substack{i \in \{1, \dots, n\} \\ j \in \{1, \dots, m\}}} b_{ij}(\mathbf{g}_i \circ \phi_{\mathbf{h}_j}),$$

for the same $b_{ij} \in K$ as before.

At the beginning of this discussion, we chose bases E and F for V and W , but this is actually not necessary. We can formulate a version of the above statement that does not depend on a choice of basis.

A more abstract statement of the above is that any linear function $V \rightarrow W$ is a sum of rank-1 linear functions. (While bases were chosen to show this result, this statement of the result does not depend on the choice of bases). To recover the particular format of the above basis-dependent result, we use this fact in conjunction with the following theorem.

Theorem 3.11. Let V and W be vector spaces. Any rank-1 linear function $V \rightarrow W$ can be expressed as $\mathbf{w} \circ \phi$, for some $\mathbf{w} \in W$ and some linear function $\phi : V \rightarrow K$, where $\mathbf{w} : K \rightarrow W$ is the linear map defined by $\mathbf{w}(c) = c\mathbf{w}$.

Proof. Let \mathbf{f} be a rank-1 linear function $V \rightarrow W$. Then the image of \mathbf{f} is $\mathbf{f}(V) = \text{span}(\mathbf{w})$ for some $\mathbf{w} \in W$, so, for all $\mathbf{v} \in V$, $\mathbf{f}(\mathbf{v}) = c\mathbf{w}$ for some $c \in K$.

Define $\phi(\mathbf{v}) = d$, where d is the unique scalar in K such that $\mathbf{f}(\mathbf{v}) = d\mathbf{w}$. Define $\mathbf{w}(c) = c\mathbf{w}$.

With these definitions, then for all $\mathbf{v} \in V$ we have $(\mathbf{w} \circ \phi)(\mathbf{v}) = \mathbf{w}(\phi(\mathbf{v})) = \mathbf{w}(d) = d\mathbf{w} = \mathbf{f}(\mathbf{v})$. Thus $\mathbf{f} = \mathbf{w} \circ \phi$.

It remains to show that the maps \mathbf{w} and ϕ are linear. Clearly, \mathbf{w} is linear. To show ϕ is linear, we show $\phi(\mathbf{v}_1 + \mathbf{v}_2) = \phi(\mathbf{v}_1) + \phi(\mathbf{v}_2)$; the proof that $\phi(c\mathbf{v}) = c\phi(\mathbf{v})$ is similar.

We have $\phi(\mathbf{v}_1 + \mathbf{v}_2) = d_{12}$, where $\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) = d_{12}\mathbf{w}$. Since \mathbf{f} is linear, $d_{12} = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2) = d_{12}\mathbf{w}$, i.e., $d_1\mathbf{w} + d_2\mathbf{w} = d_{12}\mathbf{w}$. We know $\mathbf{w} \neq \mathbf{0}$ (if it were, then \mathbf{f} would be rank-0), so $(d_1 + d_2)\mathbf{w} = d_{12}\mathbf{w}$, which means $d_1 + d_2 = d_{12}$. That is, $\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2)$. \square

Therefore, since any linear function $V \rightarrow W$, where V and W are finite-dimensional, is a finite sum of rank-1 linear functions, we have

$$\mathbf{f} = \sum_{i \in \{1, \dots, n\}} c_{ij}(\mathbf{w}_i \circ \phi_j),$$

where $\mathbf{w}_i \in \mathcal{L}(K \rightarrow W)$ is defined by $\mathbf{w}_i(c) = c\mathbf{w}_i$, $\phi_j \in \mathcal{L}(V \rightarrow K)$, and $c_{ij} \in K$.

Since we have seen that linear functions $V \rightarrow K$ are fundamental to this decomposition, we make the following definition.

Definition 3.12. (Dual space). Let V be a (not necessarily finite-dimensional) vector space over a field K . The *dual vector space* to V is the vector space over K , denoted V^* , consisting of the linear functions $V \rightarrow K$ under the operations of function addition and function scaling:

$$V^* := \mathcal{L}(V \rightarrow K).$$

One final “big leap” will complete our discovery. Recall, our original goal was to show $\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$. So, somehow, tensor product spaces will have to become involved.

We begin constructing the isomorphism by starting with $\mathbf{f} \in \mathcal{L}(V \rightarrow W)$ and decomposing it as described previously:

$$\mathbf{f} = \sum_{i \in \{1, \dots, n\}} c_{ij}(\mathbf{w}_i \circ \phi_j),$$

where $\mathbf{w}_i \in \mathcal{L}(K \rightarrow W)$ is defined by $\mathbf{w}_i(c) = c\mathbf{w}_i$, $\phi_j \in V^*$, and $c_{ij} \in K$.

The idea is to define a linear isomorphism $\mathbf{F} : \mathcal{L}(V \rightarrow W) \rightarrow W \otimes V^*$ that sends the rank-1 element $(\mathbf{w}_i \circ \phi_j) \in \mathcal{L}(V \rightarrow W)$ to the elementary tensor $\mathbf{w}_i \otimes \phi_j \in W \otimes V^*$:

$$\underbrace{\mathbf{w}_i \circ \phi_j}_{\in \mathcal{L}(V \rightarrow W)} \xrightarrow{\mathbf{F}} \underbrace{\mathbf{w}_i \otimes \phi_j}_{\in W \otimes V^*}.$$

We need to show that \mathbf{F} is a linear bijection. Ultimately, this is the case because \otimes is a bilinear map, and as \otimes correspondingly appears to be bilinear.

First, we show \mathbf{F} is linear on rank-1 compositions of the form $(\mathbf{w} \circ \phi) \in \mathcal{L}(V \rightarrow W)$. (Note, such rank-1 compositions are similar to elementary tensors in the sense that they do not need to be expressed as a linear combination of two or more other compositions). So, we need to show that

$$\begin{aligned} \mathbf{F}(\mathbf{f}_1 + \mathbf{f}_2) &= \mathbf{F}(\mathbf{f}_1) + \mathbf{F}(\mathbf{f}_2) \\ \mathbf{F}(c\mathbf{f}) &= c\mathbf{F}(\mathbf{f}), \end{aligned}$$

for all elementary compositions $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f} \in \mathcal{L}(V \rightarrow W)$ and scalars $c \in K$.

More explicitly, we need \mathbf{F} to satisfy

$$\begin{aligned} \mathbf{w}_i \circ \phi_k + \mathbf{w}_j \circ \phi_k &\xrightarrow{\mathbf{F}} \mathbf{w}_i \otimes \phi_k + \mathbf{w}_j \otimes \phi_k \\ \mathbf{w}_i \circ \phi_j + \mathbf{w}_i \circ \phi_k &\xrightarrow{\mathbf{F}} \mathbf{w}_i \otimes \phi_j + \mathbf{w}_i \otimes \phi_k \\ c(\mathbf{w}_i \circ \phi_j) &\xrightarrow{\mathbf{F}} c(\mathbf{w}_i \otimes \phi_j), \end{aligned}$$

where $\mathbf{w}_i \in \mathcal{L}(K \rightarrow W)$ is defined by $\mathbf{w}_i(c) = c\mathbf{w}_i$, $\phi_j \in V^*$, and $c \in K$.

As was alluded to before, the above is achieved due to the bilinearity of \circ and the seeming-bilinearity¹ of \otimes :

¹The fact that \circ is bilinear might seem rather abstract. It may be helpful to note that a familiar consequence of \circ being bilinear is the fact that matrix multiplication distributes over matrix addition. So, for example, $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

$$\begin{aligned}
\mathbf{w}_i \circ \phi_k + \mathbf{w}_j \circ \phi_k &= (\mathbf{w}_i + \mathbf{w}_j) \circ \phi_k \xrightarrow{\mathbf{F}} (\mathbf{w}_i + \mathbf{w}_j) \otimes \phi_k = \mathbf{w}_i \otimes \phi_k + \mathbf{w}_j \otimes \phi_k \\
\mathbf{w}_i \circ \phi_j + \mathbf{w}_i \circ \phi_k &= \mathbf{w}_i \circ (\phi_j + \phi_k) \xrightarrow{\mathbf{F}} \mathbf{w}_i \otimes \phi_j + \mathbf{w}_i \otimes \phi_k = \mathbf{w}_i \otimes (\phi_j + \phi_k) \\
c(\mathbf{w}_i \circ \phi_j) &= (c\mathbf{w}_i) \circ \phi_j \xrightarrow{\mathbf{F}} (c\mathbf{w}_i) \otimes \phi_j = c(\mathbf{w}_i \otimes \phi_j).
\end{aligned}$$

Because \mathbf{F} is linear on elementary compositions, we *impose* that \mathbf{F} is linear on nonelementary compositions to ensure its action on any defined $\mathbf{f} \in \mathcal{L}(V \rightarrow W)$ is defined, as such an \mathbf{f} is a linear combination of elementary compositions. This also “shows” that \mathbf{F} is linear for any $\mathbf{f} \in \mathcal{L}(V \rightarrow W)$.

The bijectivity of \mathbf{F} now follows easily. \mathbf{F} is surjective because any nonelementary tensor corresponds to a “nonelementary composition”, i.e., a linear combination of elementary compositions. \mathbf{F} is injective because it is injective when restricted to elementary compositions; the linearity of \mathbf{F} implies that this extends to “nonelementary compositions”. These are the main ideas of how to prove bijectivity; the explicit check is left to the reader.

So, we have proved the following theorem.

Theorem 3.13. ($\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$ naturally).

Let V and W be finite-dimensional vector spaces. Then there is a natural isomorphism

$$\mathcal{L}(V \rightarrow W) \cong W \otimes V^*.$$

This isomorphism is natural because it does not depend on a choice of basis. (See Definition 2.38).

Remark 3.14. (Rank-1 linear transformations correspond to elementary tensors).

In the derivation above, we saw that the natural isomorphism sends a rank-1 linear transformation $\mathbf{w} \circ \phi$, which we called an “elementary composition”, to an elementary tensor $\mathbf{w} \otimes \phi$.

Of course, not all linear transformations are rank-1, just as not all elements of $W \otimes V^*$ are elementary!

Remark 3.15. (Tensor product space as the structure behind composition).

In the derivation above, the bilinearity of \circ corresponded to the seeming-bilinearity of \otimes . These two notions of bilinearity are slightly different. The notion of bilinearity which \circ satisfies ultimately depends on how linear functions act on vectors, because the linearity condition $(\mathbf{f}_1 + \mathbf{f}_2) \circ \mathbf{g} = \mathbf{f}_1 \circ \mathbf{g} + \mathbf{f}_2 \circ \mathbf{g}$ ultimately depends on the definition of the function $\mathbf{f}_1 + \mathbf{f}_2$, which is $(\mathbf{f}_1 + \mathbf{f}_2)(\mathbf{v}) = \mathbf{f}_1(\mathbf{v}) + \mathbf{f}_2(\mathbf{v})$ (see, the vector \mathbf{v} is involved!). The notion of bilinearity which \otimes satisfies is simpler in the sense that it does not depend on previous notions in this way; \otimes expresses all the structure that matters without unnecessary excess.

Remark 3.16. (The two key ideas). Now that we have gone through the derivation, we can specifically see how the two key ideas of thinking of linear functions as vectors and “multilinear elements” have manifested.

We thought of the linear function $\mathbf{f} : V \rightarrow W$ as a vector when we decomposed it into a linear combination of “elementary compositions”. The notion of dual spaces allowed us to further abstract away the component $\phi \in V^* = \mathcal{L}(V \rightarrow K)$ in the “elementary composition” $\mathbf{w} \circ \phi$.

In order to distill “elementary compositions” $\mathbf{w} \circ \phi$ down into objects which express the key aspects of their bilinear structure, we used the seeming-bilinearity of \otimes .

3.3 Introduction to dual spaces

Recall that dual spaces are crucial to the concept of a $\binom{p}{q}$ tensor because they allow us to think of linear functions as vectors. As was previously mentioned, every linear function $V \rightarrow W$ is a linear combination of elements of V^* .

We now restate the definition of a dual space and make some additional remarks.

Definition 3.17. (Dual space). Let V be a (not necessarily finite-dimensional) vector space over a field K . The *dual vector space* to V is the vector space over K , denoted V^* , consisting of the linear functions $V \rightarrow K$ under the operations of function addition and function scaling:

$$V^* := \mathcal{L}(V \rightarrow K).$$

Elements of V^* have various names. They may be called *dual vectors*, *covectors*, *linear functionals*, or even *1-forms* (not to be mistaken with the notion of a *differential* 1-form).

Derivation 3.18. (Induced dual basis, dimension of finite-dimensional V^*).

Let V be a finite-dimensional vector space and let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V . Then we can discover an *dual basis* that is induced by E , as follows.

We want to use the fact that V^* is a vector space to decompose an arbitrary $\phi \in V^*$ into a basis sum. To achieve this decomposition, we utilize the correspondence between linear transformations and matrices.

Since $V^* = \mathcal{L}(V \rightarrow K)$, any element $\phi \in V^*$ is represented relative to the basis E by the (primitive) $1 \times n$ matrix $\phi(E)$ (see Derivation 2.51 and Remark 2.55).

Recall from Definition 2.49 that the matrix $\phi(E)$ is

$$\begin{pmatrix} \phi(\mathbf{e}_1) & \dots & \phi(\mathbf{e}_n) \end{pmatrix}.$$

Now we express $\phi(E)$ as a linear combination of “basis” row matrices:

$$\phi(E) = \sum_{i=1}^n \phi(\mathbf{e}_i) \hat{\mathbf{e}}_i^\top.$$

Thus, the action of any $\phi \in V^*$ on $\mathbf{v} \in V$ is expressed as

$$\phi(\mathbf{v}) = \phi(E)[\mathbf{v}]_E = \left(\sum_{i=1}^n \phi(\mathbf{e}_i) \hat{\mathbf{e}}_i^\top \right) \mathbf{v} = \sum_{i=1}^n \left(\phi(\mathbf{e}_i) \hat{\mathbf{e}}_i^\top \mathbf{v} \right).$$

This statement about matrices corresponds to a statement about linear functions. So, what linear functions do the (primitive) matrices $\hat{\mathbf{e}}_i^\top$ represent?

To address this, we define $\phi_{\mathbf{e}_i}$ to be the element of V^* that is represented relative to E by the (primitive) matrix $\hat{\mathbf{e}}_i^\top$:

$$\begin{aligned} \phi_{\mathbf{e}_i} &:= \text{the element of } V^* \text{ that is represented relative to } E \text{ by the (primitive) matrix } \hat{\mathbf{e}}_i^\top \\ \phi_{\mathbf{e}_i}(\mathbf{v}) &= \hat{\mathbf{e}}_i^\top [\mathbf{v}]_E. \end{aligned}$$

With our new definition of $\phi_{\mathbf{e}_i}$, the above becomes

$$\phi(\mathbf{v}) = \sum_{i=1}^n \left(\phi(\mathbf{e}_i) \phi_{\mathbf{e}_i}(\mathbf{v}) \right) = \left(\sum_{i=1}^n \phi(\mathbf{e}_i) \phi_{\mathbf{e}_i} \right) (\mathbf{v}).$$

So, in all, we have

$$\phi(\mathbf{v}) = \left(\sum_{i=1}^n \phi(\mathbf{e}_i) \phi_{\mathbf{e}_i} \right) (\mathbf{v}).$$

We see that any $\phi \in V^*$ is a linear combination of the $\phi_{\mathbf{e}_i}$; the $\phi_{\mathbf{e}_i}$ span V^* . They are also linearly independent because they are represented by the linearly independent (primitive) row-matrices $\hat{\mathbf{e}}_i^\top$.

The set $E^* = \{\phi_{\mathbf{e}_i}\}_{i=1}^n$ is therefore a basis for V^* . We call it the *dual basis* for V^* . We also say that E^* is *induced* by the choice of E because each basis vector $\phi_{\mathbf{e}_i}$ is defined as being represented relative to E by a (primitive) matrix, and therefore depends on how we choose E . *We cannot speak of the induced dual basis unless we have chosen a basis E for V .*

Since E^* contains n elements, we have seen that, when V is finite-dimensional, V^* is also finite-dimensional and is of the same dimension as V .

Theorem 3.19. (Equivalent definition of dual basis).

Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V . In the above, we defined $\phi_{\mathbf{e}_i}$ to be the dual vector that is represented relative to E by the (primitive) matrix $\hat{\mathbf{e}}_i^\top$.

An equivalent definition for the dual basis of V^* induced by E is to define the basis vectors $\phi_{\mathbf{e}_i}$ of V^* as acting on a basis vector \mathbf{e}_j of V by

$$\phi_{\mathbf{e}_i}(\mathbf{e}_j) = \hat{\mathbf{e}}_i^\top [\mathbf{e}_j]_E = \hat{\mathbf{e}}_i^\top \hat{\mathbf{e}}_j = \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij}.$$

Proof. We show that $(\phi_{\mathbf{e}_i}(\mathbf{v}) = \hat{\mathbf{e}}_i^\top [\mathbf{v}]_E) \iff (\phi_{\mathbf{e}_i}(\mathbf{e}_j) = \delta_{ij})$. For the forward direction, substitute $\mathbf{v} = \hat{\mathbf{e}}_j$. For the reverse direction, extend the statement on \mathbf{e}_j to a statement on arbitrary $\mathbf{v} \in V$ using the linearity of $\phi_{\mathbf{e}_i}$. \square

Remark 3.20. (An “unnatural” isomorphism $V \cong V^*$).

Suppose we’ve chosen a basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for V , so that we have the induced dual basis $E^* = \{\phi_{\mathbf{e}_1}, \dots, \phi_{\mathbf{e}_n}\}$ for V^* . We can define a linear isomorphism $V \rightarrow V^*$ that is defined on basis vectors by $\mathbf{e}_i \mapsto \phi_{\mathbf{e}_i}$.

This isomorphism is *not* natural (see Definition 2.38) because it depends on how we choose the basis E for V . Additionally, the map $\mathbf{e}_i \mapsto \phi_{\mathbf{e}_i}$ is only a surjection when V is finite-dimensional, because when V is infinite-dimensional the cardinality of V^* is strictly greater than the cardinality of V .

Remark 3.21. (Not every basis for the dual space is an induced basis).

We don’t have to pick a basis of V to pick a basis for V^* . Derivation 3.18 showed that when V is finite-dimensional, then V^* is finite-dimensional. Therefore, when V is finite-dimensional, we can pick an *arbitrary* basis for V^* .

Remark 3.22. (The misleading star notation for dual vectors).

Some authors use \mathbf{v}^* to denote an element of V^* , and use $\{\mathbf{e}_1^*, \dots, \mathbf{e}_n^*\}$ to denote an *arbitrary* basis of V^* . This notation is misleading because it suggests that there is a natural operation $*$: $V \rightarrow V^*$ that produces dual vectors from vectors. No such operation is natural, because every such operation requires a choice of basis for V . **category theoretical proof?**

(Credit goes to Mark Krusemeyer for his remarks on this in Advanced Linear Algebra).

Theorem 3.23. ($V \cong V^{**}$ naturally).

Let V be a finite-dimensional vector space. Once we have taken the dual V^* of V , we might ask “what happens if we take the dual again?”. The answer is that taking the double dual essentially returns the original space.

More formally, when V is finite-dimensional, then there is a natural linear isomorphism $\mathcal{L} : V \rightarrow V^{**}$ defined by $\mathcal{L}(\mathbf{v}) = \Phi_{\mathbf{v}}$, where $\Phi_{\mathbf{v}} : V^* \rightarrow K$ is the element of V^{**} defined by $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$.

Proof. We show that \mathcal{L} is linear, injective and surjective. Checking linearity is straightforward; \mathcal{L} is linear regardless of the dimensionality of V .

To show injectivity and surjectivity, it is useful to note that V^{**} is finite-dimensional and has the same dimension as V . (We know V^* is n -dimensional because V is n -dimensional. Replacing V in the last sentence with V^* shows that V^{**} is n -dimensional, as well).

Thus, since V and V^{**} have the same dimension, the map \mathcal{L} is injective iff it is surjective. We will show that it is injective.

\mathcal{L}_V is injective iff $(\phi(\mathbf{v}) = 0 \text{ for all } \phi \in V^* \implies \mathbf{v} = 0)$. Since V is finite-dimensional, we can show the contrapositive, $(\exists \phi \in V^* \phi(\mathbf{v}) = 0 \text{ and } \mathbf{v} \neq 0)$ after choosing a basis for V and obtaining a

dual basis for V^* . (In the infinite dimensional setting, we would need to assume the axiom of choice to get a basis for V). If $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for V , then each $\phi \in V^*$ is associated with the matrix $\phi(E)$, as shown in Derivation 3.18. We can choose ϕ to be such that $\phi(\mathbf{v}) \neq \mathbf{0}$ for $\mathbf{v} \neq \mathbf{0}$ by imposing that ϕ have a nontrivial kernel. \square

Definition 3.24. (Dual transformation).

Let V and W be finite-dimensional vector spaces, and let $\mathbf{f} : V \rightarrow W$ be a linear function. The *dual transformation of \mathbf{f}* , also called the *transpose of \mathbf{f}* , is the linear function $\mathbf{f}^* : W^* \rightarrow V^*$ defined by $\mathbf{f}^*(\chi) = \chi \circ \mathbf{f}$.

Theorem 3.25. (Dual transformation on finite-dimensional vector spaces is represented by transpose matrix).

Let V and W be finite-dimensional vector spaces, with bases E and F , respectively, and let E^* and F^* be the induced dual bases for V^* and W^* , respectively. Consider a linear function $\mathbf{f} : V \rightarrow W$. Recall that $[\mathbf{f}(E)]_F$ denotes the matrix of $\mathbf{f} : V \rightarrow W$ relative to E and F . The matrix $[\mathbf{f}^*(F^*)]_{E^*}$ of $\mathbf{f}^* : W^* \rightarrow V^*$ relative to F^* and E^* is $[\mathbf{f}(E)]_F^\top$.

Proof. Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$.

$[\mathbf{f}(E)]_F^\top$ and $[\mathbf{f}^*(F^*)]_{E^*}$ are both $n \times m$ matrices, so we can show that they are equal by showing that their ij entries are the same.

First, we compute the ij entry of $[\mathbf{f}(E)]_F^\top$. The j th column of $[\mathbf{f}(E)]_F$ is $[T(\mathbf{e}_j)]_F$, so the ij entry of $[\mathbf{f}(E)]_F^\top$ is

$$\mathbf{f}_i \cdot \mathbf{f}(\mathbf{e}_j) = \mathbf{f}_i^*(\mathbf{f}(\mathbf{e}_j)) = (\mathbf{f}_i^* \circ \mathbf{f})(\mathbf{e}_j) = \mathbf{f}^*(\mathbf{f}_i^*)(\mathbf{e}_j).$$

Thus the ij entry of $[\mathbf{f}(E)]_F^\top$ is $\mathbf{f}^*(\mathbf{f}_j^*)(\mathbf{e}_i)$.

Now we look at the ij entry of $[\mathbf{f}^*(F^*)]_{E^*}$. The j th column of $[\mathbf{f}^*(F^*)]_{E^*}$ is $[\mathbf{f}^*(\mathbf{f}_j^*)]_{E^*}$. By linearity,

$$\mathbf{f}^*(\mathbf{f}_j^*)(\mathbf{v}) = \mathbf{f}^*\left(\sum_{i=1}^n (\mathbf{e}_i^*(\mathbf{v}))\mathbf{e}_i\right) = \left(\sum_{i=1}^n \mathbf{f}^*(\mathbf{f}_j^*)(\mathbf{e}_i)\mathbf{e}_i^*\right)(\mathbf{v}).$$

In terms of functions, the above is expressed as

$$\mathbf{f}^*(\mathbf{f}_j^*) = \sum_{i=1}^n \mathbf{f}^*(\mathbf{f}_j^*)(\mathbf{e}_i)\mathbf{e}_i^*,$$

which means the j th column $[\mathbf{f}^*(\mathbf{f}_j^*)]_{E^*}$ of $[\mathbf{f}^*(F^*)]_{E^*}$, is $[\mathbf{f}^*(\mathbf{f}_j^*)]_{E^*} = \begin{pmatrix} \mathbf{f}^*(\mathbf{f}_j^*)(\mathbf{e}_1) \\ \vdots \\ \mathbf{f}^*(\mathbf{f}_j^*)(\mathbf{e}_n) \end{pmatrix}$. Thus the ij

entry of $[\mathbf{f}^*(F^*)]_{E^*}$ is $\mathbf{f}^*(\mathbf{f}_j^*)(\mathbf{e}_i)$.

Thus, the ij entry of $[\mathbf{f}(E)]_F^\top$ is the same as the ij entry of $[\mathbf{f}^*(F^*)]_{E^*}$. \square

Remark 3.26. (Motivations for defining the dual transformation).

The previous theorem reveals a new way to motivate the definition of the dual transformation. The first motivated definition, which we have already seen in Definition 3.24, is “given a linear function $\mathbf{f} : V \rightarrow W$, the dual transformation is the natural linear function $W^* \rightarrow V^*$ ”. The second motivated definition, which is informed by the previous theorem, is “if \mathbf{A} is the matrix of \mathbf{f} with respect to some bases, what linear transformation does \mathbf{A}^\top correspond to?”.

Theorem 3.27. (Dual of a composition).

Let U, V, W be finite-dimensional vector spaces, and let $\mathbf{f} : U \rightarrow V$, $\mathbf{g} : V \rightarrow W$ be linear functions. Then $(\mathbf{g} \circ \mathbf{f})^* = \mathbf{g}^* \circ \mathbf{f}^*$.

This fact is what underlies the fact $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$, which tells how to transpose a matrix-matrix product.

(The proof is left as an exercise).

Theorem 3.28. probably will remove this

When V, W are finite dimensional, $T = \mathcal{L}_W^{-1} \circ T^{**} \circ \mathcal{L}_V$. (This justifies the slight abuse of notation “ $T^{**} : V \rightarrow W$, $T^{**}(\mathbf{v}) = T(\mathbf{v})$ ”). That is, $\mathcal{L}_W \circ T = T^{**} \circ \mathcal{L}_V$, so the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ \mathcal{L}_V \downarrow & & \downarrow \mathcal{L}_W \\ V^{**} & \xrightarrow{T^{**}} & W^{**} \end{array}$$

3.4 $\binom{p}{q}$ tensors

Definition 3.29. ($\binom{p}{q}$ tensor).

Let V be a vector space. We define a $\binom{p}{q}$ tensor on V to be an element of the tensor product space $V^{\otimes p} \otimes (V^*)^{\otimes q}$. Here, we’ve used the notation $V^{\otimes k} := \underbrace{V \otimes \dots \otimes V}_{k \text{ times}}$.

We use $T_q^p(V)$ to denote the vector space of $\binom{p}{q}$ tensors on V .

Definition 3.30. (Coordinates of a $\binom{p}{q}$ tensor).

Let V be a finite-dimensional vector space over a field K with basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^* . The *coordinates of a $\binom{p}{q}$ tensor $T \in T_q^p(V)$ relative to E and E^** are the scalars $T_{j_1 \dots j_q}^{i_1 \dots i_p} \in K$ for which

$$T = \sum_{\substack{i_1, \dots, i_p \in \{1, \dots, n\} \\ j_1, \dots, j_q \in \{1, \dots, n\}}} T_{j_1 \dots j_q}^{i_1 \dots i_p} \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \phi_{\mathbf{e}_{j_1}} \otimes \dots \otimes \phi_{\mathbf{e}_{j_q}}.$$

Just as each coordinate of a $\binom{1}{1}$ tensor (a matrix) can be thought of as occupying a position in a $n \times n$ grid, each coordinate of a $\binom{p}{q}$ tensor can be thought of as occupying a position in a $n^{\times p} \times n^{\times q}$ grid. Thus, the coordinates of a $\binom{p}{q}$ tensor can be associated with a “multidimensional matrix”, or, more specifically, an element of $K^{n^p n^q}$.

Remark 3.31. (There are many definitions of “tensor”).

There are many ways to define the notion of a “tensor”. Here are three common ways to define what a tensor is that differ from our definition.

- (A physicist’s definition of a tensor). Physicists and engineers most commonly define tensors to be “multidimensional matrices” that follow the “the tensor transformation law” (which is really a change of basis formula; we will derive this in Theorem 4.39). This definition of tensor is clearly unmotivated, as it describes how tensors behave before explaining what they really are.
- (The more “concrete” but less insightful mathematical definition of a tensor). Mathematicians often define a $\binom{p}{q}$ tensor to be a multilinear map $(V^*)^{\times p} \times V^{\times q} \rightarrow K$. This definition is equivalent to the one we have used (we see why in Theorem 3.34), but it is less preferable because it obscures the concept of a “multilinear element” that tensor product spaces so nicely capture.
- (Another physicist’s definition of a tensor). Physicists also occasionally define an “ n th order tensor” to be² a linear map that sends a $(n-1)$ order tensor to a vector in V , where a tensor of order 2 is defined to be a linear map $V \rightarrow V$. This definition works because we have the natural isomorphism $T_1^1(V) = V \otimes V^* \cong \mathcal{L}(V \rightarrow V)$. Note also that, when a basis for V is fixed (which is often always done in physics, since often we have $V = \mathbb{R}^3$, so we can use the standard basis), there is no ambiguity when one says “second order tensor”, as $T_0^2(V) \cong T_1^2(V) \cong T_2^0(V)$ due to the (unnatural) isomorphism $V \cong V^*$ that is obtained by choosing a basis (see Remark 3.20).

²See p. 7 and p. 19 of Chapter 2 in [BW97] for a treatment of tensors in this way.

Definition 3.32. (Covariance and contravariance, coordinates of a $\binom{p}{q}$ tensor).

Let V be a vector space. For reasons that will be explained later, in [...], dual vectors (elements of V^*) are said to be *covariant vectors*, or *covectors*, and vectors (elements of V) are said to be *contravariant vectors*.

The coordinates of a covariant vector relative to a basis are indexed by lower subscripts; contrastingly, covariant vectors themselves are indexed by upper subscripts. So, for example, we would write a linear combination of covariant vectors as $c_1\phi^1 + \dots + c_n\phi^n$.

Contravariant vectors and their coordinates follow the opposite conventions. Coordinates of contravariant vectors are indexed by upper subscripts, and contravariant vectors themselves are indexed by lower subscripts. We would write a linear combination of contravariant vectors as $c^1\mathbf{v}_1 + \dots + c^n\mathbf{v}_n$.

Definition 3.33. (Valence and order of a tensor). The *valence* of a $\binom{p}{q}$ tensor is the tuple $\binom{p}{q}$. The *order* of a $\binom{p}{q}$ tensor is $p + q$.

Theorem 3.34. (Four fundamental natural isomorphisms for $\binom{p}{q}$ tensors).

Let V and W be finite-dimensional vector spaces over a field K . Then there exist natural isomorphisms

$$\begin{aligned} \mathcal{L}(V_1 \times \dots \times V_k \rightarrow W) &\cong \mathcal{L}(V_1 \otimes \dots \otimes V_k \rightarrow W) \\ \mathcal{L}(V \rightarrow W) &\cong W \otimes V^* \\ (V \otimes W)^* &\cong V^* \otimes W^* \end{aligned}$$

Most importantly, if $V = V_1 = \dots = V_k$ and $W = K$, then the above yields natural isomorphisms

$$\mathcal{L}(V^{\times k} \rightarrow K) \cong (V^{\otimes k})^* \cong (V^*)^{\otimes k},$$

so we have the natural isomorphism

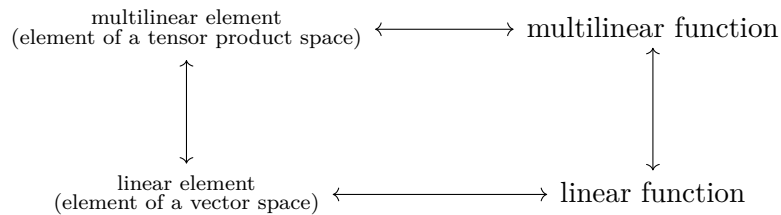
$$T_q^p(V) \cong T_p^q(V^*)$$

Proof. The first line in the first box is Theorem 3.10, and the second line in the first box is Theorem 3.13. We need to prove the third line in the first box; we need to prove that *taking the dual distributes over the tensor product*.

We do so by defining an isomorphism in the “reverse” direction. We define this isomorphism on elementary tensors and extend linearly. Given $\phi \otimes \psi \in V^* \otimes W^*$, we produce the linear map $\mathbf{f}_{\phi \otimes \psi} \in (V \otimes W)^*$, where $\mathbf{f}_{\phi \otimes \psi} : V \otimes W \rightarrow K$ is defined by $\mathbf{f}_{\phi \otimes \psi}(\mathbf{v} \otimes \mathbf{w}) = \phi(\mathbf{v})\psi(\mathbf{w})$. The explicit check that this is a linear isomorphism is left to the reader. \square

Remark 3.35. (The four-fold nature of $\binom{p}{q}$ tensors).

We have defined a $\binom{p}{q}$ tensor to be an element of a tensor product space; a $\binom{p}{q}$ tensor is a “multilinear element”. Due to the important natural isomorphisms of the previous theorem we can think of $\binom{p}{q}$ tensors in the four ways depicted by this diagram:



It’s instructive to apply these interpretations to vectors and to dual vectors. Vectors are 1-linear elements by definition, and they are less obviously “generalized linear transformations” because they are naturally identifiable with elements of V^{**} . Dual vectors are linear functions by definition, and they are less obviously 1-linear elements because they form a vector space.

Theorem 3.36. (Other useful natural isomorphisms for $\binom{p}{q}$ tensors).

Let V and W be (not necessarily finite-dimensional) vector spaces over K . Then we have natural isomorphisms

$$\begin{aligned} V \otimes K &\cong V \\ V \otimes W &\cong W \otimes V. \end{aligned}$$

The proof of this theorem is left as an exercise.

More about the natural isomorphism $\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$

We derived the natural isomorphism $\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$ for finite-dimensional V and W by defining an isomorphism $\mathcal{L}(V \rightarrow W) \rightarrow V^* \otimes W$ on rank-1 linear functions. We now present a theorem which details the explicit relationship between a matrix and its corresponding $\binom{1}{1}$ tensor, and an economical proof of the natural isomorphism $\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$.

Theorem 3.37. ($\binom{1}{1}$ tensor corresponding to a matrix).

Let V and W be finite-dimensional vector spaces, and let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ be bases for V and W , respectively. Let $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ be the basis for V^* induced by E . If (a_j^i) is the matrix of \mathbf{f} relative to E and F , then \mathbf{f} corresponds to the $\binom{1}{1}$ tensor $\sum_{ij} a_j^i \phi^{\mathbf{e}_i} \otimes \mathbf{f}_j$.

Proof. (Relationship between coordinates of a linear function and coordinates of a $\binom{1}{1}$ tensor).

We have $\sum_{ij} a_j^i \phi^{\mathbf{e}_i} \otimes \mathbf{f}_j = \sum_{ij} \phi^{\mathbf{e}_i} \otimes \mathbf{f}(\mathbf{e}_i)$. The explicit check that this is a linear isomorphism is left to the reader. It is enough to show linearity and surjectivity because $\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$ is a finite-dimensional vector space; injectivity automatically follows due to Theorem 2.37.

(You might first try sending \mathbf{f} to $\sum_{ij} a_j^i \phi^{\mathbf{e}_i} \otimes \mathbf{f}_j$. This won't work, but you'll get the idea to try $\sum_{ij} a_j^i \phi^{\mathbf{e}_i} \otimes \mathbf{f}_j = \sum_i (\phi^{\mathbf{e}_i} \otimes \sum_j a_j^i \mathbf{f}_j) = \sum_i \phi^{\mathbf{e}_i} \otimes \mathbf{f}(\mathbf{e}_i)$. \square)

We now present the traditional proof of the natural isomorphism $\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$. This proof is very economical, but, since it defines an isomorphism $W \otimes V^* \rightarrow \mathcal{L}(V \rightarrow W)$ going in the “reverse” direction, one is unlikely to discover this construction until they have proved $\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$ by more intuitive means.

Proof. We define an isomorphism $V^* \otimes W \rightarrow \mathcal{L}(V \rightarrow W)$ by $\phi \otimes \mathbf{w} \mapsto f_{\phi \otimes \mathbf{w}} : V \rightarrow W$, $f_{\phi \otimes \mathbf{w}}(\mathbf{v}_0) = \phi(\mathbf{v}_0)\mathbf{w}$. That is, $f_{\phi \otimes \mathbf{w}} = \mathbf{w}\phi$. Since $V^* \otimes W$ is finite-dimensional, it is enough to show that this map is linear and injective; surjectivity follows automatically from Theorem 2.37. \square

4

Bilinear forms, metric tensors, and coordinates of tensors

4.1 Bilinear forms and metric tensors

only really need coord thm for vectors and dual vectors from this chap; otherwise just gives more insights into tensors

Definition 4.1. (Linear k -form, bilinear form).

Let V_1, \dots, V_k be vector spaces over a field K . A *linear k -form on V_1, \dots, V_k* is a k -linear function $V_1 \times \dots \times V_k \rightarrow K$. (Unfortunately, the word “ k -form” is reserved to mean *differential k -form*. We have not defined differential k -forms yet).

Let V be a vector space over K . A *linear k -form on V* is a linear k -form on $\underbrace{V, \dots, V}_{k \text{ times}}$.

A *bilinear form on V_1 and V_2* is a linear 2-form on V_1 and V_2 , and a bilinear form on V is a linear 2-form on V and V , i.e., a bilinear form on V and V .

Remark 4.2. (Linear k -forms are naturally identified with $\binom{0}{k}$ tensors).

A linear k -form on V is an element of $\mathcal{L}(V^{\times k} \rightarrow K)$. Recalling Theorem ??, we have $\mathcal{L}(V^{\times k} \rightarrow K) \cong \mathcal{L}(V^{\otimes k} \rightarrow K) = (V^{\otimes k})^* \cong (V^*)^{\otimes k} = T_k^0(V)$. Therefore a linear k -form is naturally identified with a $\binom{0}{k}$ tensor.

Definition 4.3. (Nondegenerate bilinear form, the natural musical isomorphisms).

Let V and W be finite-dimensional vector spaces. If we have a bilinear form B on V and W , then there are natural linear maps $b_1 : V \rightarrow W^*$ and $b_2 : W \rightarrow V^*$ defined by $b_1(\mathbf{v})(\mathbf{w}) = B(\mathbf{v}, \mathbf{w})$ and $b_2(\mathbf{w})(\mathbf{v}) = B(\mathbf{v}, \mathbf{w})$. We denote $\mathbf{v}^{b_1} := b_1(\mathbf{v})$ and $\mathbf{w}^{b_2} := b_2(\mathbf{w})$.

What would it take for b_1 and b_2 to be linear isomorphisms? Well, if we knew that $b_1 : V \rightarrow W^*$ and $b_2 : W \rightarrow V^*$ were linear injections, then we would have $\dim(V) \leq \dim(W)$ and $\dim(W) \leq \dim(V)$, so we would have $\dim(V) = \dim(W)$, that is, $\dim(V) = \dim(W^*) = \dim(W) = \dim(V^*)$. Then, since b_1 and b_2 would be linear injections between finite-dimensional vector spaces of the same dimension, surjectivity would then follow automatically and b_1 and b_2 would be linear isomorphisms (see Theorem 2.37).

Therefore, if b_1 and b_2 are injections, then they are linear isomorphisms. When are b_1 and b_2 injections? This is the case if and only if their kernels are $\{\mathbf{0}\}$. In other words, b_1 and b_2 are isomorphisms iff the bilinear form B satisfies

$$\begin{aligned} B(\mathbf{v}_0, \mathbf{w}) &= \mathbf{0} \text{ for all } \mathbf{w} \in W \iff \mathbf{v}_0 = \mathbf{0} \\ B(\mathbf{v}, \mathbf{w}_0) &= \mathbf{0} \text{ for all } \mathbf{v} \in V \iff \mathbf{w}_0 = \mathbf{0}. \end{aligned}$$

(We have the reverse implications because $B(\mathbf{v}_0, \mathbf{0}) = B(\mathbf{v}_0, \mathbf{0} \cdot \mathbf{0}) = \mathbf{0} \cdot B(\mathbf{v}_0, \mathbf{0}) = \mathbf{0}$ and $B(\mathbf{0}, \mathbf{w}_0) = \mathbf{0}$ by the same argument).

A bilinear form B that satisfies the above conditions is called *nondegenerate*. We have contrived nondegenerate bilinear forms to be those for which $b_1 : V \rightarrow W^*$ and $b_2 : W \rightarrow V^*$ are natural linear isomorphisms. Note that when b_1 and b_2 are isomorphisms, they are indeed natural because they do not depend on a choice of basis (see Definition 2.38). When they are isomorphisms, b_1 and b_2 are called the *musical isomorphisms induced by B* . We denote the inverses of b_1 and b_2 by \sharp_1 and \sharp_2 , respectively: $\sharp_1 = b_1^{-1} : W^* \rightarrow V$ and $\sharp_2 = b_2^{-1} : V^* \rightarrow W$.

Definition 4.4. (The adjoint of a linear function).

Let V and W be finite-dimensional vector spaces, let B be a nondegenerate bilinear form on V and W , and consider the musical isomorphisms $b_1 : V \rightarrow W^*$ and $b_2 : W \rightarrow V^*$ induced by B .

There is an induced linear map $\mathbf{g} : V \rightarrow W$ obtained by using the musical isomorphisms on the domain and codomain of the dual transformation $\mathbf{f}^* : W^* \rightarrow V^*$. We have $\mathbf{g} = b_1^{-1} \circ \mathbf{f}^* \circ b_2$, or, equivalently, $\mathbf{g} \circ b_2 = \mathbf{f}^* \circ b_1$. So, the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{\mathbf{g}} & W \\ b_1 \downarrow & & \downarrow b_2 \\ W^* & \xrightarrow{\mathbf{f}^*} & V^* \end{array}$$

The map $\mathbf{g} : V \rightarrow W$, which can intuitively be thought of as “the dual transformation after identifying $V \cong W^*$ and $W \cong V^*$ ”, is called *the adjoint of \mathbf{f}* . Looking at the commutative diagram, we see that the adjoint satisfies $\mathbf{g}(\mathbf{v}_0)^{b_2} = \mathbf{f}^*(\mathbf{v}_0^{b_1})$. This condition further unravels after referring to the definitions of b_1 and b_2 :

$$B(\mathbf{v}_1, \mathbf{g}(\mathbf{v}_2)) = B(\mathbf{v}_2, \mathbf{f}(\mathbf{v}_1)) \text{ for all } \mathbf{v}_1, \mathbf{v}_2 \in V \text{ results from identifying } V \cong W^* \text{ and } W \cong V^*.$$

In a slight abuse of notation, we denote the adjoint $\mathbf{g} : V \rightarrow W$ of \mathbf{f} by $\mathbf{f}^* : V \rightarrow W$. (So, the adjoint of \mathbf{f} is distinguished from the dual of \mathbf{f} by the spaces that it maps between: the dual is written as $\mathbf{f} : W^* \rightarrow V^*$, and the adjoint is written as $\mathbf{f} : V \rightarrow W$). With this new notation, the above condition is restated as

$$B(\mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2)) = B(\mathbf{v}_2, \mathbf{f}(\mathbf{v}_1)) \text{ for all } \mathbf{v}_1, \mathbf{v}_2 \in V \text{ results from identifying } V \cong W^* \text{ and } W \cong V^*.$$

Theorem 4.5. (Induced bilinear form on the duals).

Let V and W be vector spaces. If B is a nondegenerate bilinear form on V and W , then there is an induced nondegenerate bilinear form \tilde{B} on W^* and V^* , $\tilde{B} = B \circ R^{-1}$, where $R : V \times W \rightarrow W^* \times V^*$ is defined by $R(\mathbf{v}, \mathbf{w}) = (\mathbf{v}^{b_1}, \mathbf{w}^{b_2})$. The bilinear form \tilde{B} supplies natural isomorphisms $\tilde{b}_1 : V^* \rightarrow W^{**}$, $\tilde{b}_2 : W^* \rightarrow V^{**}$ defined by $\phi^{b_1}(\psi) = \tilde{B}(\psi, \phi)$ and $\psi^{b_2}(\phi) = \tilde{B}(\psi, \phi)$. (Check that \tilde{B} is actually a nondegenerate bilinear form is an exercise).

The usefulness of this theorem is not apparent until we present Theorem 4.27.

Definition 4.6. (Metric tensor).

g is a *metric tensor on V and W* iff it is a nondegenerate bilinear form on V and W that is also *symmetric*, in the sense that $g(\mathbf{v}, \mathbf{w}) = g(\mathbf{w}, \mathbf{v})$ for all $\mathbf{v} \in V, \mathbf{w} \in W$. A *metric tensor on V* is a metric tensor on V and V .

(Technically, it does not make much sense to speak of a metric tensor on V and W ; we should have $V = W$ so that we can think of $\sqrt{g(\mathbf{v}, \mathbf{v})}$ as being the length of \mathbf{v} . We define the notion of a metric tensor on V and W for syntactical reasons; it is useful to be able to say “metric tensor on V and W ” rather than the more verbose “symmetric nondegenerate bilinear form on V and W ”).

Definition 4.7. (The notation \flat and \sharp).

When V is a finite-dimensional and there is a metric tensor g on V , then the musical isomorphisms $b_1 : V \rightarrow V^*$ and $b_2 : V \rightarrow V^*$ induced by g are the same because g is symmetric. This leads us to define $\flat := b_1 = b_2$ and $\sharp := \sharp_1 = \sharp_2 = b_1^{-1} = b_2^{-1}$.

Definition 4.8. (Inner product).

A metric tensor g on V is an *inner product* on V iff it is also *positive-definite*, that is, iff it is a metric tensor that satisfies $(g(\mathbf{v}, \mathbf{v}) \neq 0 \iff \mathbf{v} = \mathbf{0} \text{ for all } \mathbf{v} \in V)$. (We automatically have the reverse implication for any bilinear form g for the reasons discussed in Definition 4.3).

Iff g is an inner product, we denote it by $\langle \cdot, \cdot \rangle$ and use the notation $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle := g(\mathbf{v}_1, \mathbf{v}_2)$, for $\mathbf{v}_1, \mathbf{v}_2 \in V$.

Definition 4.9. (Vector space with inner product). Let V be a vector space over K . Iff there is an inner product $\langle \cdot, \cdot \rangle$ on V , then V is called a *vector space with inner product*, or an *inner product space*.

Remark 4.10. (Positive-definite \implies nondegenerate, but the converse does not hold).

The first part of the title of this remark is straightforwardly checked by looking at the definition of nondegenerate bilinear form. Therefore, all inner products are metric tensors, but not all metric tensors are inner products.

Example 4.11. The dot product on \mathbb{R}^n is an inner product on \mathbb{R}^n . (Proof left as exercise).

The dot product on K^n , defined analogously to the dot product on \mathbb{R}^n , is in general *not* an inner product because it is not positive-definite. For example, we have $\begin{pmatrix} 3 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 0$ when these vectors are elements of $\mathbb{Z}/9\mathbb{Z}$. (In $\mathbb{Z}/9\mathbb{Z}$, we have $3 \cdot 3 = 9 = 0$).

Length and orthogonality with respect to an inner product

Let V be a finite-dimensional inner product space with inner product $\langle \cdot, \cdot \rangle$.

Definition 4.12. (Length of a vector with respect to an inner product).

In analogy to the fact that the length of a vector in \mathbb{R}^n can be expressed using the dot product on \mathbb{R}^n (see Theorem 2.92), we define the *length of a vector* $\mathbf{v} \in V$ with respect to the inner product on V to be $\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

Definition 4.13. (Angle between vectors with respect to an inner product).

In analogy to the definition of the angle between vectors in \mathbb{R}^n , (see Definition 2.93), we define the *angle* θ between vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ with respect to the inner product on V to be $\cos^{-1} \left(\frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} \right)$.

Definition 4.14. (Adjoint of a linear function $V \rightarrow W$).

Let V and W be vector spaces with inner products $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_W$, respectively. Consider a linear function $\mathbf{f} : V \rightarrow W$, and let $\mathbf{f}^* : W^* \rightarrow V^*$ be its adjoint (see Definition 4.4). The condition on $\langle \cdot, \cdot \rangle$ induced by using the musical isomorphism $\flat : V \rightarrow V^*$ to identify $V \cong V^*$ (see the end of Definition 4.4) is

$$\langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle \text{ for all } \mathbf{v}_1, \mathbf{v}_2 \in V.$$

Remark 4.15. (Geometric inner product).

Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$. Then $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos(\theta)$, where θ is the angle between \mathbf{v}_1 and \mathbf{v}_2 with respect to the inner product on V . (This fact is the generalization of the geometric dot product on \mathbb{R}^n , which was discussed in Remark 2.94).

Theorem 4.16. (Cauchy-Schwarz inequality for vector spaces over \mathbb{R}).

Let V be a vector space over \mathbb{R} with inner product $\langle \cdot, \cdot \rangle$. Then the *Cauchy-Schwarz inequality* holds: $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle \leq \|\mathbf{v}_1\| \|\mathbf{v}_2\|$. Equivalently, the angle θ in V between \mathbf{v}_1 and \mathbf{v}_2 with respect to the inner product on V is in the set $[0, 2\pi)$.

Note that *if* we knew $\theta \in [0, 2\pi)$, the Cauchy-Schwarz inequality would immediately follow. We do not actually know $\theta \in [0, 2\pi)$ until we prove the Cauchy-Schwarz inequality!

Proof. Define $f : \mathbb{R} \rightarrow [0, \infty) \subseteq \mathbb{R}$ by $f(c) = \langle c\mathbf{v}_1 + \mathbf{v}_2, c\mathbf{v}_1 + \mathbf{v}_2 \rangle = c^2\langle \mathbf{v}_1, \mathbf{v}_1 \rangle + 2c\langle \mathbf{v}_1, \mathbf{v}_2 \rangle + \langle \mathbf{v}_2, \mathbf{v}_2 \rangle$. Since $\langle \cdot, \cdot \rangle$ is positive-definite, $f(c) \geq 0$, with $f(c) = 0$ only when $c\mathbf{v}_1 + \mathbf{v}_2 = 0$. Since f is nonnegative, then it must have either one or zero real roots, meaning $b^2 - 4ac = (2\langle \mathbf{v}_1, \mathbf{v}_2 \rangle)^2 - 4(\langle \mathbf{v}_1, \mathbf{v}_1 \rangle)(\langle \mathbf{v}_2, \mathbf{v}_2 \rangle) \leq 0$. Thus $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle^2 \leq \langle \mathbf{v}_1, \mathbf{v}_1 \rangle \langle \mathbf{v}_2, \mathbf{v}_2 \rangle = \|\mathbf{v}_1\|^2 \|\mathbf{v}_2\|^2$. Take the square root of each side to obtain the result. \square

Definition 4.17. (Orthogonality of vectors with respect to an inner product).

We say vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ are *orthogonal with respect to the inner product on V* iff the angle between \mathbf{v}_1 and \mathbf{v}_2 is $\frac{\pi}{2}$. That is, $\mathbf{v}_1, \mathbf{v}_2 \in V$ are orthogonal iff $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$.

Definition 4.18. (Orthonormal basis).

Let V be a finite-dimensional vector space with inner product $\langle \cdot, \cdot \rangle$. We say $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is *orthonormal* iff

- $\|\mathbf{e}_i\| = 1$ for all i
- \mathbf{e}_i and \mathbf{e}_j are orthogonal to each other when $i \neq j$

That is, E is an orthonormal basis iff $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$ for all i, j .

Theorem 4.19. (Gram-Schmidt algorithm).

Symmetric and orthogonal linear functions

This subsection is presented only completeness, so reading the entirety of this subsection is not necessary. The only results that are necessary to know are the conditions (3) and (4) of Definition 4.22 satisfied by an orthogonal linear function.

Let V be a vector space with inner products $\langle \cdot, \cdot \rangle$. Consider a linear function $\mathbf{f} : V \rightarrow V$, and let $\mathbf{f}^* : V \rightarrow V$ be its adjoint (see Definition 4.4). The defining condition of the adjoint (see the end of Definition 4.4) is

$$\langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle \text{ for all } \mathbf{v}_1, \mathbf{v}_2 \in V.$$

We will use this condition as a starting point for considering special classes of linear functions: we will investigate linear functions \mathbf{f} for which $\mathbf{f} = \mathbf{f}^*$ and for which $\mathbf{f}^* = \mathbf{f}^{-1}$.

Definition 4.20. (Transpose of a matrix).

The *transpose* of an $m \times n$ matrix (a_{ij}) is the $n \times m$ matrix (a_{ji}) .

Definition 4.21. (Symmetric linear function).

Let V be a vector space, consider a linear function $\mathbf{f} : V \rightarrow V$, and let \mathbf{f}^* be its adjoint. We define \mathbf{f} to be *symmetric* iff the following equivalent conditions hold:

1. $\mathbf{f} = \mathbf{f}^*$
2. $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{f}(\mathbf{v}_2) \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.
3. (a) If V is finite-dimensional and \hat{U} is an orthonormal basis of V , then the matrix of \mathbf{f} relative to an orthonormal basis is a *symmetric matrix*. That is, such a matrix is its own transpose: for any orthonormal basis \hat{U} , $[\mathbf{f}(\hat{U})]_{\hat{U}} = [\mathbf{f}(\hat{U})]_{\hat{U}}^\top$.
 (b) Moreover, the matrix of the adjoint $\mathbf{f}^* : V \rightarrow V$ relative to \hat{U} is $[\mathbf{f}(\hat{U})]_{\hat{U}}^\top$.

Proof.

(1 \iff 2).

(\implies). Use $\mathbf{f} = \mathbf{f}^*$ with $\langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle$.

(\impliedby). We have $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{f}(\mathbf{v}_2) \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $\langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle$.

Therefore $\langle \mathbf{v}_1, \mathbf{f}(\mathbf{v}_2) \rangle = \langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle$. Due to the cancelability of inner products (see Theorem ??), we have $\mathbf{f}(\mathbf{v}_2) = \mathbf{f}^*(\mathbf{v}_2)$ for all $\mathbf{v}_2 \in V$. So $\mathbf{f} = \mathbf{f}^*$.

(2 \iff 3a).

(\implies). Use $\mathbf{v}_1 = \mathbf{e}_i$ and $\mathbf{v}_2 = \mathbf{e}_j$ to see $a_{ij} = a_{ji}$.

(\impliedby). Let $\widehat{U} = \{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n\}$ be an orthonormal basis for V , and let the matrix \mathbf{A} of \mathbf{f} relative to \widehat{U} satisfy $a_{ij} = a_{ji}$. Since $a_{ij} = \langle \mathbf{f}(\hat{\mathbf{u}}_i), \hat{\mathbf{u}}_j \rangle$, then $\langle \mathbf{f}(\hat{\mathbf{u}}_i), \hat{\mathbf{u}}_j \rangle = \langle \hat{\mathbf{u}}_i, \mathbf{f}(\hat{\mathbf{u}}_j) \rangle$. Extend with multilinearity to obtain the conclusion.

(3b). The j th column of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$ is $[\mathbf{f}(\hat{\mathbf{u}}_j)]_{\widehat{U}}$, so the ij entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$ is $\langle \mathbf{f}(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle$. Similarly, the ij entry of the matrix of \mathbf{f}^* relative to \widehat{U} is $\langle \mathbf{f}^*(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle$. Due to the condition on $\langle \cdot, \cdot \rangle$ induced by the identification $V \cong V^*$, we have $\langle \mathbf{f}^*(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle = \langle \hat{\mathbf{u}}_j, \mathbf{f}(\hat{\mathbf{u}}_i) \rangle = \langle \mathbf{f}^*(\hat{\mathbf{u}}_i), \hat{\mathbf{u}}_j \rangle$. But this is the ji entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$, i.e., the ij entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$. Thus $[\mathbf{f}(\widehat{U})]_{\widehat{U}} = [\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$. \square

Definition 4.22. (Orthogonal linear function).

Let V be a vector space over K (where¹ we have $K \neq \mathbb{Z}/2\mathbb{Z}$), consider a linear function $\mathbf{f} : V \rightarrow V$, and let \mathbf{f}^* be its adjoint. We define \mathbf{f} to be *orthogonal* iff the following equivalent conditions hold:

1. $\mathbf{f}^* = \mathbf{f}^{-1}$.
2. $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{f}^{-1}(\mathbf{v}_2) \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.
3. \mathbf{f} preserves inner product: $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_2) \rangle$.
4. \mathbf{f} preserves length.
5. \mathbf{f} preserves length and angle.
6. If V is finite-dimensional and \widehat{U} is an orthonormal basis of V , then the matrix $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$ of \mathbf{f} relative to \widehat{U} has orthonormal columns.
7. If V is finite-dimensional and \widehat{U} is an orthonormal basis of V , then $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top [\mathbf{f}(\widehat{U})]_{\widehat{U}} = \mathbf{I}$ and $[\mathbf{f}(\widehat{U})]_{\widehat{U}} [\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top = \mathbf{I}$, so $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^{-1} = [\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$.

Proof. We prove (3) \iff (4) \iff (5) and then (1) \iff (2) \iff (3) \iff (6) \iff (7).

Here is the proof that (3) \iff (4) \iff (5):

(3 \implies 4). Length is a function of inner product, $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. Therefore, if inner product is preserved, then length is preserved.

(4 \iff 5). The reverse direction is obvious; we need to show the forward direction. The angle θ between \mathbf{v}_1 and \mathbf{v}_2 is $\theta = \cos^{-1} \left(\frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} \right)$. Since θ is a function of preserved quantities (dot product and length), it too is a preserved quantity.

(5 \implies 3). Replace the dot product \cdot on \mathbb{R}^n with the inner product $\langle \cdot, \cdot \rangle$ on V in the equation stated in the proof of Lemma 2.87 to show that $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \frac{1}{2}(\|\mathbf{v}_1 + \mathbf{v}_2\|^2 - (\|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2))$. That is, the inner product $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ is a function of $\|\mathbf{v}_1\|$ and $\|\mathbf{v}_2\|$. Applying the previous formula, the inner product $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_2) \rangle$ is a function of $\|\mathbf{f}(\mathbf{v}_1)\|$ and $\|\mathbf{f}(\mathbf{v}_2)\|$: $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_2) \rangle = \frac{1}{2}(\|\mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2)\|^2 - (\|\mathbf{f}(\mathbf{v}_1)\|^2 + \|\mathbf{f}(\mathbf{v}_2)\|^2))$. Since \mathbf{f} is linear, this becomes $\frac{1}{2}(\|\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2)\|^2 - (\|\mathbf{f}(\mathbf{v}_1)\|^2 + \|\mathbf{f}(\mathbf{v}_2)\|^2))$. If length is preserved, then this is the same as $\frac{1}{2}(\|\mathbf{v}_1 + \mathbf{v}_2\|^2 - (\|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2)) = \langle \mathbf{v}_1, \mathbf{v}_2 \rangle$.

Now we show (1) \iff (2) \iff (3) \iff (6) \iff (7).

(1 \iff 2).

(\implies). Use $\mathbf{f}^* = \mathbf{f}^{-1}$ with $\langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle$.

(\impliedby). $\langle \mathbf{v}_1, \mathbf{f}^{-1}(\mathbf{v}_2) \rangle = \langle \mathbf{f}(\mathbf{v}_1), \mathbf{v}_2 \rangle$ by hypothesis, and $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle$ by condition on $\langle \cdot, \cdot \rangle$ imposed by identifying $V \cong V^*$ for \mathbf{f}^* . Thus $\langle \mathbf{v}_1, \mathbf{f}^{-1}(\mathbf{v}_2) \rangle = \langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle$.

(2 \implies 3). Substitute $\mathbf{v}_3 = \mathbf{f}^{-1}(\mathbf{v}_2)$, so that we have $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_3) \rangle = \langle \mathbf{v}_1, \mathbf{v}_3 \rangle$ for all $\mathbf{v}_1, \mathbf{v}_3 \in V$.

(3 \iff 6).

¹This is a very technical condition, and not much attention should be paid to it. We require this so that $2 \neq 0$, which allows us to divide by 2.

(\implies). We have in particular that $\langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \langle \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j \rangle = \delta_j^i$. Therefore $\langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \delta_j^i$, so the columns $[\mathbf{f}(\hat{\mathbf{u}}_i)]_{\hat{U}}$ of the matrix of \mathbf{f} relative to \hat{U} are orthonormal.

(\impliedby). Since the columns $[\mathbf{f}(\hat{\mathbf{u}}_i)]_{\hat{U}}$ of the matrix of \mathbf{f} relative to \hat{U} are orthonormal, we have $\langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \delta_j^i$. Extend with multilinearity to obtain the conclusion.

(6 \iff 7).

(\implies). The ij entry of $[\mathbf{f}(\hat{U})]_{\hat{U}}[\mathbf{f}(\hat{U})]_{\hat{U}}^\top$ is $(i\text{th row of } [\mathbf{f}(\hat{U})]_{\hat{U}}) \cdot (j\text{th column of } [\mathbf{f}(\hat{U})]_{\hat{U}})^\top = (i\text{th row of } [\mathbf{f}(\hat{U})]_{\hat{U}}) \cdot (j\text{th row of } [\mathbf{f}(\hat{U})]_{\hat{U}}) = \langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \delta_j^i$.

(\impliedby). Reversing the logic of the forward direction, we know $\langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \delta_j^i$. Therefore (3) is satisfied. Then we use (3) \implies (6) \implies (7).

(5 \implies 1). Use the fact that $[\mathbf{f}(\hat{U})]_{\hat{U}}^\top$ is the matrix of the adjoint $\mathbf{f}^* : V \rightarrow V$. □

Self-duality and choice of basis

Definition 4.23. (Self-dual basis).

Let V be a finite-dimensional vector space, let B be a nondegenerate bilinear form on V , let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V , and let $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^* . We say E is *self-dual* iff the musical isomorphism $\flat : V \rightarrow V^*$ induced by B sends $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ for all i .

Theorem 4.24. (For vector spaces over \mathbb{R} , a musical isomorphism is the unique isomorphism enabling self-duality of a basis).

Let V be a finite-dimensional vector space over \mathbb{R} , let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V , let $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^* , and suppose there is an isomorphism $\mathbf{F} : V \rightarrow V^*$ sending $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$. Then $\mathbf{F} = \flat : V \rightarrow V^*$, where \flat is the musical isomorphism induced by the inner product $\langle \cdot, \cdot \rangle$ on \mathbb{R} defined by $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = [\mathbf{v}_1]_E \cdot [\mathbf{v}_2]_E$, and where \cdot is the dot product on \mathbb{R}^n . (It is left to the reader to check that this is indeed an inner product).

This theorem sheds light on the “unnatural” isomorphism $V \rightarrow V^*$ which sends $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ that was discussed in Remark 3.20). We said that the isomorphism which sends $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ is not natural because it depends on a choice of basis. (This is still true). Recently, we have seen that a nondegenerate bilinear form, such as an inner product, on V induces the natural musical isomorphism $\flat : V \rightarrow V^*$. This theorem shows that the unnatural isomorphism $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ can be viewed as arising from the usually natural but now, in this case, “unnatural” musical isomorphism $\flat : V \rightarrow V^*$ that arises from choosing a basis E for V .

Proof.

The induced dual basis vector $\phi^{\mathbf{e}_i} \in V^*$ acts on a vector $\mathbf{v} \in V$ by $\phi^{\mathbf{e}_i}(\mathbf{v}) = [\mathbf{e}_i]_E^\top [\mathbf{v}]_E = [\mathbf{e}_i]_E \cdot [\mathbf{v}]_E = \langle \mathbf{e}_i, \mathbf{v} \rangle = \mathbf{e}_i^\flat(\mathbf{v})$. Thus $\phi^{\mathbf{e}_i}(\mathbf{v}) = \mathbf{e}_i^\flat(\mathbf{v})$, so $\phi^{\mathbf{e}_i} = \mathbf{e}_i^\flat$. □

Theorem 4.25. (Self-dual \iff orthonormal).

Let V be a finite-dimensional vector space with inner product $\langle \cdot, \cdot \rangle$. Then a basis E for V is self-dual iff it is orthonormal with respect to the inner product on V .

Proof. Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, and let the induced dual basis be $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$. Let $\flat : V \rightarrow V^*$ be the musical isomorphism induced by the inner product (not by the self-duality of E) on V .

We have: self-dual $\iff \mathbf{e}_i^\flat = \phi^{\mathbf{e}_i} \iff \mathbf{e}_i^\flat(\mathbf{e}_j) = \phi^{\mathbf{e}_i}(\mathbf{e}_j) \iff \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \phi^{\mathbf{e}_i}(\mathbf{e}_j) \iff \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_j^i \iff$ orthonormal. □

4.2 Coordinates of $\binom{p}{q}$ tensors

Key ideas:

1. id V with V^* with musical isos
2. multiplication by the metric is matrix matrix mult (?) and therefore composition of linear transfs. leads to contraction

Theorem 4.26. (Coordinates of vectors and dual vectors).

Let V be a finite-dimensional vector space, let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V , and let $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ be the basis for V^* induced by E .

We have

$$\boxed{\begin{aligned} ([\mathbf{v}]_E)^i &= \phi^{\mathbf{e}_i}(\mathbf{v}) = \Phi_{\mathbf{v}}(\phi^{\mathbf{e}_i}) \\ ([\phi]_{E^*})_i &= \phi(\mathbf{e}_i) \end{aligned}}$$

Recall from Theorem 3.23 that \mathbf{v} is identified with $\Phi_{\mathbf{v}} \in V^{**}$, where $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$.

Proof. First, we prove the first equation in the first line:

$$\phi^{\mathbf{e}_i}(\mathbf{v}) = \phi^{\mathbf{e}_i}\left(\sum_{j=1}^n ([\mathbf{v}]_E)^j \mathbf{e}_j\right) = \sum_{j=1}^n \left(([\mathbf{v}]_E)^j \phi^{\mathbf{e}_i}(\mathbf{e}_j)\right) = \sum_{j=1}^n ([\mathbf{v}]_E)^j \delta_j^i = ([\mathbf{v}]_E)^i.$$

Now we prove the second line. By definition of $[\cdot]_{E^*}$, any $\phi \in V^*$ is of the form $\phi = \sum_{j=1}^n ([\phi]_{E^*})_j \phi^{\mathbf{e}_j}$, so we can compute $\phi(\mathbf{e}_i)$ as

$$\phi(\mathbf{e}_i) = \left(\sum_{j=1}^n ([\phi]_{E^*})_j \phi^{\mathbf{e}_j}\right)(\mathbf{e}_i) = \sum_{j=1}^n \left(([\phi]_{E^*})_j \phi^{\mathbf{e}_j}(\mathbf{e}_i)\right) = \sum_{j=1}^n ([\phi]_{E^*})_j \delta_i^j = ([\phi]_{E^*})_i.$$

Lastly, we prove the second equality of the first line. There are two ways to do so; each way presents a different insight. The first way is most economical, and the second way emphasizes the fact when a vector is identified with a vector of a different vector space, those vectors have the same coordinates relative to the bases that are identified with each other.

1. Recall from Theorem 3.23 that the definition of $\Phi_{\mathbf{v}}$ is $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$. We immediately get the result from this.
2. Since $\mathbf{v} \mapsto \Phi_{\mathbf{v}}$ is an isomorphism (see Theorem 3.23), then $([\mathbf{v}]_E)^i = ([\Phi_{\mathbf{v}}]_{E^{**}})^i$. By the second line, we have $([\Phi_{\mathbf{v}}]_{E^{**}})^i = \Phi_{\mathbf{v}}(\phi^{\mathbf{e}_i})$. Thus $([\mathbf{v}]_E)^i = \Phi_{\mathbf{v}}(\phi^{\mathbf{e}_i})$.

□

Theorem 4.27. (Relationship between coordinates of vectors and dual vectors for vector spaces).

“When we have a metric tensor g , then $V \cong V^*$ naturally, so we can convert between $\binom{k}{0}$ tensors and $\binom{0}{k}$ tensors. blah blah”

Let V and W be finite-dimensional vector spaces with bases E and F , and let $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, \dots, \psi^{\mathbf{f}_m}\}$ be corresponding induced dual bases of V^* and W^* . Let g be a metric tensor on V and W , and let \tilde{g} be the induced metric tensor on W^* and V^* (see Theorem 4.5). Let $b_1 : V \rightarrow W^*$ and $b_2 : W \rightarrow V^*$ be the musical isomorphisms induced by g .

Define the $n \times m$ matrices

$$\begin{aligned} \mathbf{g} &= (g_{ij}) := (g(\mathbf{e}_i, \mathbf{f}_j)) \\ \tilde{\mathbf{g}} &= (g^{ij}) := (\tilde{g}(\psi^{\mathbf{f}_i}, \phi^{\mathbf{e}_j})). \end{aligned}$$

Then

$$\begin{aligned} ([\mathbf{v}]_E)^i &= \sum_{j=1}^n g^{ij} ([\mathbf{v}^{b_1}]_{F^*})_j = (i\text{th row of } \tilde{\mathbf{g}}) \cdot [\mathbf{v}^{b_1}]_{F^*} \\ [\mathbf{v}]_E &= \tilde{\mathbf{g}} [\mathbf{v}^{b_1}]_{F^*} \\ ([\mathbf{v}^{b_1}]_{F^*})_i &= \sum_{j=1}^n g_{ij} ([\mathbf{v}]_E)^j = (i\text{th row of } \mathbf{g}) \cdot [\mathbf{v}]_E \\ [\mathbf{v}^{b_1}]_{F^*} &= \mathbf{g} [\mathbf{v}]_E. \end{aligned}$$

We will explain the significance of the lower indices in g_{ij} and the upper indices in g^{ij} in Remark 4.32.

Proof. We first prove the third equation, which implies the fourth equation, and then prove the first equation, which implies the second equation.

We prove the third equation by using the metric tensor g on V and the fact $([\phi]_{F^*})_i = \phi(\mathbf{f}_i)$, which is implied by the second line inside the box of the previous theorem:

$$([\mathbf{v}^{b_1}]_{F^*})_i = \mathbf{v}^{b_1}(\mathbf{f}_i) = g(\mathbf{v}, \mathbf{f}_i) = g\left(\sum_{j=1}^n ([\mathbf{v}]_E)^j \mathbf{e}_j, \mathbf{f}_i\right) = \sum_{j=1}^n ([\mathbf{v}]_E)^j g(\mathbf{e}_j, \mathbf{f}_i) = \sum_{j=1}^n ([\mathbf{v}]_E)^j g_{ji} = \sum_{j=1}^n g_{ij} ([\mathbf{v}]_E)^j.$$

We prove the first equation by using the induced metric tensor \tilde{g} on V^* . Since $\mathbf{v} \xrightarrow{b_1} \mathbf{v}^{b_1} \in W^*$ and $\mathbf{v}^{b_1} \xrightarrow{\tilde{b}_1} (\mathbf{v}^{b_1})^{\tilde{b}_1} \in V^{**}$, where these maps are isomorphisms, we have $([\mathbf{v}]_E)^i = ([(\mathbf{v}^{b_1})^{\tilde{b}_1}]_{E^{**}})^i$. Use the fact $([\mathbf{v}^{b_1})^{\tilde{b}_1}]_{E^{**}})^i = (\mathbf{v}^{b_1})^{\tilde{b}_1}(\phi^{\mathbf{e}_i})$, which is implied by the second equation on the first line inside the box of the previous theorem, to get the first equation:

$$\begin{aligned} ([\mathbf{v}]_E)^i &= ([(\mathbf{v}^{b_1})^{\tilde{b}_1}]_{E^{**}})^i = (\mathbf{v}^{b_1})^{\tilde{b}_1}(\phi^{\mathbf{e}_i}) = \tilde{b}_1(\mathbf{v}^{b_1})(\phi^{\mathbf{e}_i}) = \tilde{g}(\mathbf{v}^{b_1}, \phi^{\mathbf{e}_i}) = \tilde{g}\left(\sum_{j=1}^n ([\mathbf{v}^{b_1}]_{F^*})_j \psi^{\mathbf{f}_j}, \phi^{\mathbf{e}_i}\right) \\ &= \sum_{j=1}^n ([\mathbf{v}^{b_1}]_{F^*})_j \tilde{g}(\psi^{\mathbf{f}_j}, \phi^{\mathbf{e}_i}) = \sum_{j=1}^n ([\mathbf{v}^{b_1}]_{F^*})_j g^{ji} = \sum_{j=1}^n g^{ij} ([\mathbf{v}^{b_1}]_{F^*})_j. \end{aligned}$$

□

We now present two lemmas that enable a more “complete” statement of the previous theorem.

Lemma 4.28. (Primitive matrix of element of V^* as transposed coordinates).

Let V be a finite-dimensional vector space, let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V , and consider an element $\phi \in V^*$. Then the (primitive) matrix $\phi(E)$ of ϕ relative to E (see Remark 2.55) is $\phi(E) = [\phi]_{E^*}^\top$.

Proof.

$$\phi(E) = \begin{pmatrix} \phi(\mathbf{e}_1) & \dots & \phi(\mathbf{e}_n) \end{pmatrix} = \begin{pmatrix} ([\phi]_{E^*})_1 & \dots & ([\phi]_{E^*})_n \end{pmatrix} = [\phi]_{E^*}^\top$$

.

□

Lemma 4.29. Let V and W be finite-dimensional vector spaces with bases E and F , respectively, let g be a metric tensor on V , and let \tilde{g} be the induced metric tensor on W^* and V^* . Then

$$\begin{aligned} g(\mathbf{v}, \mathbf{w}) &= [\mathbf{v}]_E^\top \mathbf{g} [\mathbf{w}]_F = [\mathbf{w}]_F^\top \mathbf{g} [\mathbf{v}]_E \\ \tilde{g}(\psi, \phi) &= [\psi]_{F^*}^\top \mathbf{g}^{-1} [\phi]_{E^*} = [\phi]_{E^*}^\top \mathbf{g}^{-1} [\psi]_{F^*}. \end{aligned}$$

Most importantly, since we earlier defined $\tilde{\mathbf{g}} := (\tilde{g}(\psi^{\mathbf{f}_i}, \phi^{\mathbf{e}_j}))$, we have that $\tilde{\mathbf{g}} = \mathbf{g}^{-1}$ as a consequence. (Use $\psi = \psi^{\mathbf{f}_i}$ and $\phi = \phi^{\mathbf{e}_i}$ in the above to see this). That is,

$$\boxed{(g_{ij})^{-1} = (g^{ij})}$$

Proof. We show the first equation of each line. The second equation on each line follows by transposing the first equation on each line and using the fact that \mathbf{g} and $\tilde{\mathbf{g}}$ are symmetric matrices.

First, we show the first equation of the first line. We have $g(\mathbf{v}, \mathbf{w}) = \mathbf{v}^{b_1}(\mathbf{w})$. Notice that the previous lemma implies that the $1 \times m$ primitive matrix $\mathbf{v}^{b_1}(F)$ of $\mathbf{v}^{b_1} : W \rightarrow K$ relative to F is $\mathbf{v}^{b_1}(F) = [\mathbf{v}^{b_1}]_{F^*}^\top$. Applying the characterizing property of primitive matrices (see Derivation 2.51), we

have $g(\mathbf{v}, \mathbf{w}) = \mathbf{v}^{b_1}(\mathbf{w}) = \mathbf{v}^{b_1}(F)[\mathbf{w}]_F = [\mathbf{v}^{b_1}]_{F^*}^\top [\mathbf{w}]_F$. So $g(\mathbf{v}, \mathbf{w}) = [\mathbf{v}^{b_1}]_{F^*}^\top [\mathbf{w}]_F$. The fourth equation of Theorem 4.27 states that $[\mathbf{v}^{b_1}]_{F^*} = \mathbf{g}[\mathbf{v}]_E$. Thus $g(\mathbf{v}, \mathbf{w}) = (\mathbf{g}[\mathbf{v}]_E)^\top [\mathbf{w}]_F = [\mathbf{v}]_E^\top \mathbf{g}[\mathbf{w}]_F$, which is the first equation of the first line.

Now we show the first equation of the second line. We have $\tilde{g}(\psi, \phi) = g(b_1^{-1}(\psi), b_2^{-1}(\phi))$ by definition of \tilde{g} (see Theorem 4.5). We need to find the $\mathbf{v} \in V$ and $\mathbf{w} \in W$ for which $\mathbf{v} = b_1^{-1}(\psi)$ and $\mathbf{w} = b_2^{-1}(\phi)$, and then compute $\tilde{g}(\psi, \phi) = g(\mathbf{v}, \mathbf{w}) = [\mathbf{v}]_E^\top \mathbf{g}[\mathbf{w}]_F$.

As was noted in the first part of this proof, $[\psi]_{F^*} = \mathbf{g}[\mathbf{v}]_E$ due to Theorem 4.27. Therefore $b_1^{-1}(\psi) = \mathbf{v}$, where $[\mathbf{v}]_E = \mathbf{g}^{-1}[\psi]_{F^*}$. Similarly, $b_2^{-1}(\phi) = \mathbf{w}$, where $[\mathbf{w}]_F = \mathbf{g}^{-1}[\phi]_{E^*}$.

Therefore $\tilde{g}(\psi, \phi) = g(\mathbf{v}, \mathbf{w}) = [\mathbf{v}]_E^\top \mathbf{g}[\mathbf{w}]_F = (\mathbf{g}^{-1}[\psi]_{F^*})^\top \mathbf{g}(\mathbf{g}^{-1}[\phi]_{E^*}) = [\psi]_{F^*}^\top \mathbf{g}^{-1} \mathbf{g} \mathbf{g}^{-1}[\phi]_{E^*} = [\psi]_{F^*}^\top \mathbf{g}[\phi]_{E^*}$. In this last step, we used that \mathbf{g}^{-1} is symmetric. \square

Theorem 4.30. (Relationship between coordinates of vectors and dual vectors).

Consider the hypotheses of Theorem 4.27. The fact that $(g_{ij})^{-1} = (g^{ij})$, proved in the previous lemma, immediately implies that

$$\begin{aligned} ([\mathbf{v}]_E)^i &= \sum_{j=1}^n g^{ij}([\mathbf{v}^{b_1}]_{F^*})_j = (\text{ith row of } \mathbf{g}^{-1}) \cdot [\mathbf{v}^{b_1}]_{F^*} \\ [\mathbf{v}]_E &= \mathbf{g}^{-1}[\mathbf{v}^{b_1}]_{F^*} \\ ([\mathbf{v}^{b_1}]_{F^*})_i &= \sum_{j=1}^n g_{ij}([\mathbf{v}]_E)^j = (\text{ith row of } \mathbf{g}) \cdot [\mathbf{v}]_E \\ [\mathbf{v}^{b_1}]_{F^*} &= \mathbf{g}[\mathbf{v}]_E \end{aligned}$$

Remark 4.31. (The special case $g_{ij} = \delta_{ij}$).

When $V = W$, g is an inner product on V , and $E = \hat{U} = \{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n\}$ is an orthonormal basis (relative to the inner product) for V , then $g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j) = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$.

Then the converse is true as well (so long as $V = W$). If $g_{ij} = \delta_{ij}$, then g is an inner product on V , and, since $g(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) = g_{ij} = \delta_{ij}$ for some basis $\hat{U} = \{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n\}$ of V , then that basis is an orthonormal basis of V .

Remark 4.32. (Lower indices in g_{ij} and upper indices in g^{ij}).

We use lower indices in the notation $g_{ij} := g(\mathbf{e}_i, \mathbf{e}_j)$ because, if g is a metric tensor on V , then g is an element of $\mathcal{L}(V \times V \rightarrow K)$ and can thus be identified with a $\binom{0}{2}$ tensor on V (see Remark 4.2). We use lower indices for $g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j)$ to follow the index conventions of Definition 3.32.

What about the upper indices in g^{ij} ? Suppose again that g is a metric tensor on V . Just as g can be identified with a $\binom{0}{2}$ tensor on V , the induced bilinear form \tilde{g} on V^* can be identified with a $\binom{2}{0}$ tensor on V . (An element of $\mathcal{L}(V^* \times V^* \rightarrow K)$ can be identified with a $\binom{0}{2}$ tensor on V^* , which can be identified with a $\binom{2}{0}$ tensor on V). Therefore, since \mathbf{g}^{-1} stores the coordinates of a $\binom{2}{0}$ tensor, we use upper indices to denote the ij entry of \mathbf{g}^{-1} .

Remark 4.33. (δ^{ij} vs. δ_j^i vs. δ_i^j).

One may be confused when they consider that there are three ways to write the Kronecker delta: δ^{ij} , δ_j^i , and δ_i^j . The difference between these functions is straightforward: δ^{ij} is used for coordinates of a $\binom{2}{0}$ tensor, δ_j^i is used for a coordinates of a $\binom{1}{1}$ tensor, and δ_{ij} is used for coordinates of a $\binom{0}{2}$ tensor. All three functions are defined, as usual, to be 1 when $i = j$ and 0 otherwise.

Change of basis for $\binom{p}{q}$ tensors

Definition 4.34. (Coordinates of a $\binom{p}{q}$ tensor).

For convenience, we now restate Definition 3.30.

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, and let E^* be a basis for V^* . (Note that we have not assumed that E^* is induced by E). The *coordinates of a $\binom{p}{q}$ tensor $T \in T_q^p(V)$ relative to E and E^** are the scalars $T_{j_1 \dots j_q}^{i_1 \dots i_p}$ for which

$$T = \sum_{\substack{i_1, \dots, i_p \in \{1, \dots, n\} \\ j_1, \dots, j_q \in \{1, \dots, n\}}} T_{j_1 \dots j_q}^{i_1 \dots i_p} \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_q}.$$

Lemma 4.35. (Lemma for change of basis for vectors and dual vectors).

Let V be a finite-dimensional vector space with bases E and F , and let E^* and F^* be the corresponding induced dual bases for V^* . Then

$$([\phi]_{F^*})_i = [\phi]_{E^*}^\top [\mathbf{f}_i]_E.$$

Proof. Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$. Then, starting with the fact $([\phi]_{F^*})_i = \phi(\mathbf{f}_i)$ (see Theorem 4.26), we have

$$\begin{aligned} ([\phi]_{F^*})_i &= \phi(\mathbf{f}_i) = \phi\left(\sum_{j=1}^n ([\mathbf{f}_i]_E)_j \mathbf{e}_j\right) = \sum_{j=1}^n ([\mathbf{f}_i]_E)_j \phi(\mathbf{e}_j) = \sum_{j=1}^n ([\mathbf{f}_i]_E)_j ([\phi]_{E^*})_j \\ &= \left(([\phi]_{E^*})_1 \quad \dots \quad ([\phi]_{E^*})_n \right) [\mathbf{f}_i]_E = [\phi]_{E^*}^\top [\mathbf{f}_i]_E. \end{aligned}$$

□

Theorem 4.36. (Change of basis for vectors and dual vectors).

Let V be a finite-dimensional vector space with bases E and F , and let E^* and F^* be the corresponding induced dual bases for V^* . Then

$$\begin{aligned} [\mathbf{v}]_F &= [\mathbf{E}]_F [\mathbf{v}]_E = [\mathbf{F}]_E^{-1} [\mathbf{v}]_E \\ [\phi]_{F^*} &= [\mathbf{E}]_F^{-\top} [\phi]_{E^*} = [\mathbf{F}]_E^\top [\phi]_{E^*} \\ [\mathbf{E}]_F^{-1} &= [\mathbf{F}]_E \\ [\mathbf{E}^*]_{F^*} &= [\mathbf{E}]_F^{-\top} \end{aligned}$$

where $\mathbf{v} \in V$ and $\phi \in V^*$.

Proof. The first line of the boxed equation is Theorem 2.63, and the third line is Theorem 2.64. We will prove $[\mathbf{E}^*]_{F^*} = [\mathbf{E}]_F^{-\top}$, which is the fourth line. The second line then follows by applying the fourth line to the equation $[\phi]_{F^*} = [\mathbf{E}^*]_{F^*} [\mathbf{v}]_{E^*}$ (this last equation is implied by the first line).

As usual, set $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, $F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$, and $E^* = \{\phi^{e_1}, \dots, \phi^{e_n}\}$. We show $[\mathbf{E}^*]_{F^*} = [\mathbf{E}]_F^{-\top}$ by examining the ij entry of $[\mathbf{E}^*]_{F^*}$ and making use of the previous lemma. We have

$$([\mathbf{E}^*]_{F^*})_j^i = [\phi^{e_j}]_{F^*} = [\phi^{e_j}]_{E^*}^\top [\mathbf{f}_i]_E = \hat{\mathbf{e}}_j^\top [\mathbf{f}_i]_E = ([\mathbf{f}_i]_E)^j = ([\mathbf{F}]_E)_i^j = ([\mathbf{F}]_E^\top)_j^i \stackrel{\text{third line}}{=} ([\mathbf{E}]_F^{-\top})_j^i.$$

□

Remark 4.37. (What covariance and contravariance *really* mean).

The first two equations of the previous theorem can be restated as

$$\begin{aligned} [\mathbf{v}]_F &= [\mathbf{F}]_E^{-1} [\mathbf{v}]_E \\ [\phi]_{F^*}^\top &= [\phi]_{E^*}^\top [\mathbf{F}]_E. \end{aligned}$$

(We have simply copied the first equation from the previous theorem. The second equation has been obtained by transposing its counterpart from the previous theorem).

Paying close attention to the second above equation, we see that when we treat the coordinates of dual vectors taken relative to the E^* basis as row vectors (i.e. as transposed column vectors), then

these row vectors transform over to the F^* basis with use of $[\mathbf{F}]_E$. On the other hand, the first above equation states that the coordinates of vectors relative to E (when treated as column vectors, as usual) transform over to the F basis with use of $[\mathbf{F}]_E^{-1}$. For this reason, dual vectors are considered to be *covariant*, since they “co-vary” with $[\mathbf{F}]_E$, and vectors are considered to be *contravariant*, since they “contra-vary” against $[\mathbf{F}]_E$.

Theorem 4.38. (Change of basis for vectors and dual vectors in terms of basis vectors and basis dual vectors).

Let V be a finite-dimensional vector space with bases $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, \dots, \psi^{\mathbf{f}_n}\}$ be the corresponding induced dual bases for V^* . We have

$$\begin{aligned} \mathbf{f}_i &= \sum_{j=1}^n ([\mathbf{f}_i]_E)_j \mathbf{e}_j = \sum_{j=1}^n ([\mathbf{F}]_E)_i^j \mathbf{e}_j \\ \psi^{\mathbf{f}_i} &= \sum_{j=1}^n ([\psi^{\mathbf{f}_i}]_{E^*})_j \phi^{\mathbf{e}_j} = \sum_{j=1}^n ([\mathbf{F}]_E^{-\top})_i^j \phi^{\mathbf{e}_j} \end{aligned}$$

Proof. The first line in the boxed equation follows from the definition of $[\cdot]_F$ (and was stated earlier in Theorem 2.67). The second line in the boxed equation follows by applying the first line to the bases F^* and E^* for V^* . Specifically, the second equation in the second line follows because $\delta^i = \sum_{j=1}^n ([\mathbf{F}^*]_{E^*})_i^j \phi^{\mathbf{e}_j}$, where we have $[\mathbf{F}^*]_{E^*} = [\mathbf{F}]_E^{-\top}$ due to the previous theorem. \square

Theorem 4.39. (Change of basis for a $\binom{p}{q}$ tensor).

Let V be a finite-dimensional vector space with bases $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, \dots, \psi^{\mathbf{f}_n}\}$ be the corresponding induced dual bases for V^* .

We now derive how to change the coordinates of a $\binom{p}{q}$ tensor in $T_q^p(V)$. To do so, it is enough to relate the coordinates of the elementary $\binom{p}{q}$ tensor

$$T = \mathbf{f}_{i_1} \otimes \dots \otimes \mathbf{f}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes \dots \otimes \phi^{\mathbf{e}_{j_q}}$$

taken relative to F and F^* to the coordinates of T relative to E and E^* .

To obtain this relation, we apply the previous theorem to each basis vector in T :

$$\begin{aligned} T &= \left(\sum_{j_1=1}^n ([\mathbf{F}]_E)_{i_1}^{j_1} \mathbf{e}_{j_1} \right) \otimes \dots \otimes \left(\sum_{j_p=1}^n ([\mathbf{F}]_E)_{i_p}^{j_p} \mathbf{e}_{j_p} \right) \otimes \left(\sum_{i_1=1}^n ([\mathbf{F}_E]^{-1})_{i_1}^{j_1} \phi^{\mathbf{e}_{i_1}} \right) \otimes \dots \otimes \left(\sum_{i_q=1}^n ([\mathbf{F}_E]^{-1})_{i_q}^{j_q} \phi^{\mathbf{e}_{i_q}} \right) \\ &= \sum_{j_1=1}^n \dots \sum_{j_p=1}^n \sum_{i_1=1}^n \dots \sum_{i_q=1}^n \left(([\mathbf{F}]_E)_{i_1}^{j_1} \dots ([\mathbf{F}]_E)_{i_p}^{j_p} ([\mathbf{F}_E]^{-1})_{i_1}^{j_1} \dots ([\mathbf{F}_E]^{-1})_{i_q}^{j_q} \mathbf{e}_{j_1} \otimes \dots \otimes \mathbf{e}_{j_p} \otimes \phi^{\mathbf{e}_{i_1}} \otimes \dots \otimes \phi^{\mathbf{e}_{i_q}} \right). \end{aligned}$$

Though we have argued using an elementary tensor, this shows that an arbitrary $\binom{p}{q}$ tensor with an $\binom{i_1 \dots i_p}{j_1 \dots j_q}$ component of $T_{j_1 \dots j_q}^{i_1 \dots i_p}$ relative to F and F^* has an $\binom{i_1 \dots i_p}{j_1 \dots j_q}$ component relative to E and E^* of

$$\sum_{k_1=1}^n \dots \sum_{k_p=1}^n \sum_{\ell_1=1}^n \dots \sum_{\ell_q=1}^n \left(([\mathbf{F}]_E)_{\ell_1}^{k_1} \dots ([\mathbf{F}]_E)_{\ell_p}^{k_p} ([\mathbf{F}_E]^{-1})_{\ell_1}^{k_1} \dots ([\mathbf{F}_E]^{-1})_{\ell_q}^{k_q} T_{\ell_1 \dots \ell_q}^{k_1 \dots k_p} \right).$$

(It is possible to this expression using the fact that $([\mathbf{F}]_E)_j^i ([\mathbf{F}_E]^{-1})_j^i = \delta_j^i$. Let's not do that, because that would require introducing the max function to account for whether $p \geq q$ or $q < p$).

This change of basis formula is sometimes called the *Ricci transformation law*, or the *tensor transformation law*.

At this stage, it would be remiss not to mention what is called *Einstein summation notation*. In Einstein summation notation, we assume that there is an “implied summation” over any index that

appears in both a lower and upper index. We can use Einstein notation to write the $\binom{i_1 \dots i_p}{j_1 \dots j_q}$ component of the previously mentioned $\binom{p}{q}$ tensor relative to E and E^* as

$$([\mathbf{F}]_E)_{\ell_1}^{k_1} \dots ([\mathbf{F}]_E)_{\ell_p}^{k_p} ([\mathbf{F}_E]^{-1})_{\ell_1}^{k_1} \dots ([\mathbf{F}_E]^{-1})_{\ell_q}^{k_q} T_{\ell_1 \dots \ell_q}^{k_1 \dots k_p} \quad (\text{Einstein notation}).$$

Remark 4.40. (Tensors as “multidimensional matrices” that “transform like tensors”).

As was mentioned in Remark 3.31, physicists often define tensors to be “multidimensional matrices” that follow the change of basis formula of the previous theorem.

Tensor contraction

Derivation 4.41. (Composition of linear functions with contraction).

Let V, W and Z be vector spaces over a field K . Notice that the map \circ which composes linear functions $V \rightarrow W$ and $W \rightarrow Z$ is itself a bilinear map $\mathcal{L}(V \rightarrow W) \times \mathcal{L}(W \rightarrow Z) \xrightarrow{\circ} \mathcal{L}(V, Z)$. (Check this as an exercise!). Also recall from Section 3.2 that every element of $\mathcal{L}(V \rightarrow W)$ and $\mathcal{L}(W \rightarrow Z)$ are linear combinations of rank-1 compositions of linear functions, i.e., of “elementary compositions”. We can understand the composition map \circ more deeply by looking at how it acts on such elementary compositions.

Lastly, recall the convention of Section 3.2 which, for $\mathbf{w} \in W$, uses the same symbol \mathbf{w} to denote the linear map $\mathbf{w} \in \mathcal{L}(K \rightarrow W)$ defined by $\mathbf{w}(c) = c\mathbf{w}$. Then, under the composition map, $(\mathbf{z} \circ \phi, \mathbf{w} \circ \phi) \in \mathcal{L}(V \rightarrow W) \times \mathcal{L}(W \rightarrow Z)$ is sent to

$$(\mathbf{w} \circ \phi, \mathbf{z} \circ \psi) \xrightarrow{\circ} (\mathbf{z} \circ \psi) \circ (\mathbf{w} \circ \phi) = \mathbf{z} \circ (\psi \circ \mathbf{w}) \circ \phi.$$

Notice that $\phi \circ \mathbf{w}$ is the linear map $K \rightarrow K$ sending $c \mapsto c\psi(\mathbf{w})$. If we extend the convention of Section 3.2 mentioned above to elements of K , and denote the linear map $K \rightarrow K$ sending $c \mapsto c\psi(\mathbf{w})$ by $\psi(\mathbf{w})$, then we have

$$(\mathbf{w} \circ \phi, \mathbf{z} \circ \psi) \xrightarrow{\circ} \mathbf{z} \circ \psi(\mathbf{w}) \circ \phi = \psi(\mathbf{w}) \circ \mathbf{z} \circ \phi = \mathbf{z} \circ \phi \circ \psi(\mathbf{w})$$

(In the last two equalities, we were able to commute $\psi(\mathbf{w})$ because it is a linear map $K \rightarrow K$).

The action of $\phi \in W^*$ on $\mathbf{w} \in W$ is said to be the result of evaluating the *natural pairing map on W and W^** , or, equivalently, the result of *contracting W against W^** . Therefore, we see that the composition of linear maps, when we restrict the linear maps to be elementary compositions, involves *contraction*. These notions are formalized in the next definition.

Definition 4.42. (Tensor contraction).

Let V be a vector space, and consider also its dual space V^* . There is a natural bilinear form C on V and V^* , often called the *natural pairing (of V and V^*)*, that is defined by $C(\mathbf{v}, \phi) = \phi(\mathbf{v})$.

Now suppose that V is finite-dimensional. We define the $\binom{k}{\ell}$ *contraction* on elementary $\binom{p}{q}$ tensors, and extend with multilinearity. The $\binom{k}{\ell}$ contraction of an elementary tensor is defined as follows:

$$\begin{aligned} & \mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_p \otimes \phi^1 \otimes \dots \otimes \phi^q \\ & \quad \binom{k}{\ell} \text{ contraction} \\ & \quad \longmapsto \\ & C(\mathbf{v}_k, \phi^\ell)(\mathbf{v}_1 \otimes \dots \otimes \cancel{\mathbf{v}_k} \otimes \dots \otimes \mathbf{v}_p \otimes \phi^1 \otimes \dots \otimes \cancel{\phi^\ell} \otimes \dots \otimes \phi^q) \\ & \quad = \\ & \phi^\ell(\mathbf{v}_k)(\mathbf{v}_1 \otimes \dots \otimes \cancel{\mathbf{v}_k} \otimes \dots \otimes \mathbf{v}_p \otimes \phi^1 \otimes \dots \otimes \cancel{\phi^\ell} \otimes \dots \otimes \phi^q). \end{aligned}$$

Remark 4.43. (Contraction with upper and lower indices).

The convention we laid out in 3.32 requires that lower indices (e.g. those which appear in \mathbf{v}_k) are always contracted against upper indices (e.g. those which appear in ϕ^ℓ), and vice versa. Lower indices are never contracted against other lower indices, and upper indices are never contracted against other upper indices.

Remark 4.44. (Composition of linear functions with tensor contraction, revisited).

The map \circ which composes linear functions is itself a bilinear map $\mathcal{L}(V \rightarrow W) \times \mathcal{L}(W \rightarrow Z) \xrightarrow{\circ} \mathcal{L}(V, Z)$. Due to Theorem 3.34, we have the natural isomorphism $\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$, so \circ can be identified with a linear map $\tilde{\circ} : (W \otimes V^*) \otimes (Z^* \otimes W) \rightarrow Z \otimes V^*$. Following a similar argument as was presented in Derivation 4.41, we see that $\tilde{\circ}$ acts on elementary tensors by $(\mathbf{w} \otimes \phi) \otimes (\mathbf{z} \otimes \psi) \xrightarrow{\tilde{\circ}} C(\mathbf{w}, \psi)(\mathbf{z} \otimes \phi) = \psi(\mathbf{w})(\mathbf{z} \otimes \phi)$.

Theorem 4.45. (Coordinates of a contracted tensor).

Let V be an n -dimensional vector space, let E be a basis for V , and let E^* be a basis for V^* . Consider a $\binom{p}{q}$ tensor $\mathbf{T} \in T_q^p(V)$. If the $\binom{i_1 \dots i_p}{j_1 \dots j_q}$ coordinate of \mathbf{T} relative to E and E^* is $T_{j_1 \dots j_q}^{i_1 \dots i_p}$, then the $\binom{i_1 \dots i_{p-1}}{j_1 \dots j_{q-1}}$ component of the $\binom{k}{\ell}$ contraction of \mathbf{T} relative to E and E^* is $\sum_{r=1}^n T_{j_1 \dots j_{\ell-1} \ r \ j_{\ell+1} \dots j_{q-1}}^{i_1 \dots i_{k-1} \ r \ i_{k+1} \dots i_{p-1}}$.

Proof. Assume \mathbf{T} has $\binom{i_1 \dots i_p}{j_1 \dots j_q}$ component $T_{j_1 \dots j_q}^{i_1 \dots i_p}$, so $\mathbf{T} = \sum_{I_p, J_q} T_{J_q}^{I_p} \mathbf{e}^{I_p} \otimes \phi_{J_q}$, where $I_p = \{i_1, \dots, i_p\}$, and $J_q = \{j_1, \dots, j_q\}$. Then, using linearity of the contraction map, the $\binom{k}{\ell}$ contraction of \mathbf{T} is

$$\sum_{I_p, J_q} \epsilon_{j_\ell}(\mathbf{e}^{i_k}) T_{J_q}^{I_p} \mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}},$$

where $I_p - \{i_k\} = \{i_1, \dots, \cancel{j_k}, \dots, i_p\}$, $J_q - \{j_\ell\} = \{j_1, \dots, \cancel{j_\ell}, \dots, j_p\}$. Because $\epsilon_{j_\ell}(\mathbf{e}^{i_k}) = \delta_{i_k j_\ell}$, the above sum reduces to

$$\sum_{I_p - \{i_k\}, J_q - \{j_\ell\}} \sum_r \left(T_{J_q}^{I_p} \mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}} \right),$$

where $I_p' = \{i_1, \dots, i_{k-1}, r, i_{k+1}, \dots, i_p\}$, $J_q' = \{j_1, \dots, j_{\ell-1}, r, j_{\ell+1}, \dots, j_p\}$. Since $\mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}}$ does not vary with r , this is the same as

$$\begin{aligned} & \sum_{I_p - \{i_k\}, J_q - \{j_\ell\}} \left(\mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}} \sum_r T_{J_q}^{I_p'} \right) \\ &= \sum_{I_p - \{i_k\}, J_q - \{j_\ell\}} \left(\sum_r T_{J_q}^{I_p'} \right) \mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}} \end{aligned}$$

So, we can see that the $\binom{k}{\ell}$ contraction of T is a $\binom{p-1}{q-1}$ tensor, with a $\binom{I_p'}{J_q'}$ component of $\sum_r T_{J_q}^{I_p'}$. In other words, the $\binom{i_1 \dots i_{k-1} i_{k+1} \dots i_p}{j_1 \dots j_{\ell-1} j_{\ell+1} \dots j_p}$ component of the $\binom{k}{\ell}$ contraction of T is

$$\sum_r T_{j_1 \dots j_{\ell-1} \ r \ j_{\ell+1} \dots j_p}^{i_1 \dots i_{k-1} \ r \ i_{k+1} \dots i_p}.$$

Equivalently, after shifting the subindices of i_{k+1}, \dots, i_p down by one (which is valid because the subindices k, ℓ of i, j are no longer “occupied”) we see the $\binom{i_1 \dots i_{p-1}}{j_1 \dots j_{q-1}}$ component of the $\binom{k}{\ell}$ contraction of T is

$$\sum_r T_{j_1 \dots j_{\ell-1} \ r \ j_{\ell+1} \dots j_{q-1}}^{i_1 \dots i_{k-1} \ r \ i_{k+1} \dots i_{p-1}}.$$

□

Theorem 4.46. Taking any $\binom{k}{\ell}$ contraction is basis-independent.

Proof. This follows from the fact that the natural bilinear form C on V and V^* defined by $C(\mathbf{v}, \phi) = \phi(\mathbf{v})$ is basis-independent, as no bases are required to define C . \square

Theorem 4.47. (The trace is the $\binom{1}{1}$ contraction of a $\binom{1}{1}$ tensor).

Let V be a finite-dimensional vector space over a field K .

The *trace* of a square matrix (a_j^i) with entries in K is defined to be the sum of the matrix's diagonal entries: $\text{tr}(a_j^i) := \sum_{i=1}^n a_i^i$. We have that $\text{tr}(a_j^i)$ is the $\binom{1}{1}$ contraction of the $\binom{1}{1}$ tensor corresponding to (a_j^i) .

Thus, we see the trace is a special case of tensor contraction.

Proof. Recall from Theorem 3.37 that when (a_j^i) is interpreted to be the matrix relative to bases of a linear function $V \rightarrow V$, then (a_j^i) is identified with the $\binom{1}{1}$ tensor $\sum_{ij} a_i^j \phi^{\mathbf{e}_j} \otimes \mathbf{e}^i$, where $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ is the basis for V^* induced by the basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for V . The $\binom{1}{1}$ contraction of this $\binom{1}{1}$ tensor is $\sum_{ij} a_i^j \phi^{\mathbf{e}_j}(\mathbf{e}^i) = \sum_{ij} a_i^j \delta_i^j = \sum_i a_i^i = \text{tr}(a_j^i)$. \square

Slanted indices

Theorem 4.48. Correspondence between multidimensional matrices and $\binom{p}{q}$ tensors

Theorem 4.49. (Converting vectors to dual vectors and dual vectors to vectors in a $\binom{p}{q}$ tensor).

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, \dots, \epsilon^n\}$ be a basis for V^* .

We can “convert” a $\binom{p}{q}$ tensor $\mathbf{T} \in T_q^p(V)$ to a $\binom{p-1}{q+1}$ tensor as follows:

$$\begin{aligned} \mathbf{T} &= \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_k} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_q} \\ &\longmapsto \\ &\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_k}^{b_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_q}. \end{aligned}$$

Using Theorem 4.30, we compute $\mathbf{e}_{i_k}^{b_1}$ to be

$$\mathbf{e}_{i_k}^{b_1} = \sum_{r=1}^n ([\mathbf{e}_{i_k}^{b_1}]_{E^*})_r \epsilon^r = \sum_{r=1}^n \left(\sum_{j=1}^n g_{rj} ([\mathbf{e}_{i_k}]_E)^j \right) \epsilon^r = \sum_{r=1}^n \left(\sum_{j=1}^n g_{rj} \delta_{i_k}^j \right) \epsilon^r = \sum_{r=1}^n g_{i_k r} \epsilon^r,$$

so \mathbf{T} is ultimately sent to

$$\begin{aligned} &\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_{k-1}} \otimes \sum_{r=1}^n \left(g_{i_k r} \epsilon^r \right) \otimes \mathbf{e}_{i_{k+1}} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_q} \\ &= \sum_{r=1}^n g_{i_k r} \left(\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_{k-1}} \otimes \epsilon^r \otimes \mathbf{e}_{i_{k+1}} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_q} \right). \end{aligned}$$

Rearranging the position of ϵ^r within the sum, we see that this $\binom{p}{q}$ tensor is identifiable with the tensor

$$\sum_{r=1}^n g_{i_k r} \left(\mathbf{e}_{i_1} \otimes \dots \otimes \cancel{\mathbf{e}_{i_k}} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^r \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_q} \right).$$

Overall, we have converted the $\binom{p}{q}$ tensor \mathbf{T} to a $\binom{p-1}{q+1}$ tensor as follows:

$$\begin{aligned}
\mathbf{T} &= \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_k} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_q} \\
&\longmapsto \\
&\sum_{r=1}^n g_{i_k r} \left(\mathbf{e}_{i_1} \otimes \dots \otimes \cancel{\mathbf{e}_{i_k}} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^r \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_q} \right).
\end{aligned}$$

If the coordinates of \mathbf{T} relative to E and E^* were originally $T_{j_1 \dots j_q}^{i_1 \dots i_p}$, then they get sent to $\sum_{r=1}^n g_{i_k r} T_{r j_1 \dots \cancel{j_k} \dots j_q}^{i_1 \dots i_p}$. Following a similar process to above, we can convert a $\binom{p}{q}$ tensor to a $\binom{p+1}{q-1}$ tensor like this:

$$\begin{aligned}
\mathbf{T} &= \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \dots \otimes \epsilon^{j_\ell} \otimes \dots \otimes \epsilon^{j_q} \\
&\longmapsto \\
&\sum_{r=1}^n g^{j_\ell r} \left(\epsilon^r \otimes \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \dots \otimes \cancel{\epsilon^{j_\ell}} \otimes \dots \otimes \epsilon^{j_q} \right).
\end{aligned}$$

(Note that we have used the contravariant metric tensor, with upper indices, here). Here, if the coordinates of \mathbf{T} relative to E and E^* were originally $T_{j_1 \dots j_q}^{i_1 \dots i_p}$, then they get sent to $\sum_{r=1}^n g^{i_k r} T_{j_1 \dots \cancel{j_k} \dots j_q}^{r i_1 \dots i_p}$.

To make sure the above two mappings are invertible (i.e. reversable), we adopt the convention that if we have two identifications $V \rightarrow V^*$, then the result of the second identification (which will be $\mathbf{e}_{i_u}^b$ for some positive integer u) goes immediately to the left of the previous identification (which will be \mathbf{e}_{i_v} for some positive integer $v < u$), where the first identification is immediately to the left of the leftmost “original” basis dual vector (which is ϵ^{j_1} in both of the above mappings). Similarly, if we have two identifications $V^* \rightarrow V$, then the result of the second identification goes immediately to the left of the previous identification, where the first identification is immediately to the right of the leftmost “original” basis vector (which is \mathbf{e}_{i_1} in both of the above mappings).

5

Exterior powers, the determinant, and orientation

5.1 Exterior powers

Antisymmetric tensors

Definition 5.1. (Permutations on $\{1, \dots, n\}$).

Let X be any set. A *permutation on X* is a bijection $X \rightarrow X$. Intuitively, a permutation on X is thought of as redistributing the names of the elements of X .

We use S_n to denote the *set of permutations on $\{1, \dots, n\}$* . Formally, $S_n := \{\{1, \dots, n\} \rightarrow \{1, \dots, n\}\}$.

Remark 5.2. (Sign of a permutation).

Without proof, we will use the fact that every permutation in S_n can be decomposed into a composition of “two-element swaps”. (Formally, “two-element swaps” are called *transpositions*). The *sign* function on permutations in S_n is the function $\text{sgn} : S_n \rightarrow \{-1, 1\}$ defined by $\text{sgn}(\sigma) := (-1)^n$, where n is the number of “two-element swaps” that occur in *any* of the decompositions of σ into only two-element swaps. An equivalent definition of sgn is

$$\text{sgn}(\sigma) := \begin{cases} 1 & \text{there are an even number of “two-element” swaps in any decomposition of } \sigma \\ -1 & \text{there are an odd number of “two-element” swaps in any decomposition of } \sigma \end{cases}.$$

It may seem surprising that sgn is a well-defined function. That is, it may seem surprising that the parity (the “evenness” or “oddness”) of the number of two-element swaps in a permutation’s decomposition into only two-element swaps is always the same. This might seem relatively surprising, but it’s true! (We do not prove this statement about permutations, either).

Definition 5.3. (Permuting an element of a tensor product space).

Let V_1, \dots, V_k be vector spaces over the same field. Given a permutation $\sigma \in S_k$ and a tensor $\mathbf{T} \in V_1 \otimes \dots \otimes V_k$, we define the map $(\cdot)^\sigma : V_1 \otimes \dots \otimes V_k \rightarrow V_1 \otimes \dots \otimes V_k$ sending $\mathbf{T} \mapsto \mathbf{T}^\sigma$ by specifying its action on elementary tensors and extending linearly. We define

$$(\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k)^\sigma := \mathbf{v}_{\sigma(1)} \otimes \dots \otimes \mathbf{v}_{\sigma(k)}.$$

Definition 5.4. (Antisymmetric tensor).

Let V_1, \dots, V_k be vector spaces over the same field. We say a tensor in $V_1 \otimes \dots \otimes V_k$ is *antisymmetric* iff $\mathbf{T}^\sigma = \text{sgn}(\sigma)\mathbf{T}$.

Remark 5.5. (We want to construct an algebra of antisymmetric tensors).

Consider a tuple $(A, K, +, \cdot, \star)$, where A is a set, K is a field, and $\cdot : K \times A \rightarrow A$, and $\star : A \times A \rightarrow A$ are binary functions. We say that A is an *algebra over K* iff

- $(A, K, +, \cdot)$ is a vector space
- $\star : A \times A \rightarrow A$ is bilinear

Let V_1, \dots, V_k be vector spaces over the same field. Observe that the tuple

$$(\{\text{antisymmetric tensors in } V_1 \otimes \dots \otimes V_k\}, K, \cdot, \otimes)$$

is *not* an algebra, since the tensor product \otimes of antisymmetric two tensors in $V_1 \otimes \dots \otimes V_k$ is *not* necessarily another antisymmetric tensor. Consider, for example, the¹ antisymmetric $\binom{1}{0}$ tensors (i.e. vectors) \mathbf{v}_1 and \mathbf{v}_2 ; their tensor product $\mathbf{v}_1 \otimes \mathbf{v}_2$ does not satisfy $\mathbf{v}_1 \otimes \mathbf{v}_2 = -\mathbf{v}_2 \otimes \mathbf{v}_1$.

We will construct a function $\wedge : (V_1 \otimes \dots \otimes V_k) \times (V_1 \otimes \dots \otimes V_k) \rightarrow V_1 \otimes \dots \otimes V_k$, called the *wedge product*, such that

$$(\{\text{antisymmetric tensors in } V_1 \otimes \dots \otimes V_k\}, K, \cdot, \wedge)$$

is an algebra.

Constructing the wedge product

Definition 5.6. (Antisymmetrization of elements of tensor product spaces).

Let V_1, \dots, V_k be vector spaces over the same field. We define an *antisymmetrizing function* $\text{alt} : V_1 \otimes \dots \otimes V_k \rightarrow V_1 \otimes \dots \otimes V_k$ that converts any tensor $\mathbf{T} \in V_1 \otimes \dots \otimes V_k$ into an antisymmetric tensor in $V_1 \otimes \dots \otimes V_k$. We define alt on elementary tensors and extend linearly: for an elementary tensor $\mathbf{T} \in V_1 \otimes \dots \otimes V_k$, we define

$$\text{alt}(\mathbf{T}) := \frac{1}{k!} \sum_{\sigma \in S_k} \text{sgn}(\sigma) \mathbf{T}^\sigma.$$

The division by $k!$ is used so that $\text{alt}(\mathbf{T}) = \mathbf{T}$ when \mathbf{T} is antisymmetric.

How might you come up with this formula? Well, you first might start by noticing the case of $k = 2$, without the division by $2!$. That is, notice that we can send $\mathbf{v} \otimes \mathbf{w}$ to the antisymmetric tensor $\mathbf{v} \otimes \mathbf{w} - \mathbf{w} \otimes \mathbf{v}$.

Proof. We need to show that $\text{alt}(\mathbf{T})$ is antisymmetric, i.e., that $\text{alt}(\mathbf{T})^\pi = \text{sgn}(\pi) \text{alt}(\mathbf{T})$. We have

$$\text{alt}(\mathbf{T})^\pi = \left(\sum_{\sigma \in S_k} \text{sgn}(\sigma) (\mathbf{T}^\sigma) \right)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma) (\mathbf{T}^\sigma)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma) \mathbf{T}^{\pi \circ \sigma}.$$

Since S_k is closed under taking the inverse of a permutation (recall, a permutation in S_k is a bijection on $\{1, \dots, k\}$), then for every $\tau \in S_k$ there is a $\sigma \in S_k$ such that $\tau = \pi \circ \sigma$; namely, $\sigma = \pi^{-1} \circ \tau$. So the sum becomes

$$\sum_{\tau \in S_k} \text{sgn}(\pi^{-1} \circ \tau) \mathbf{T}^\tau = \text{sgn}(\pi^{-1}) \sum_{\tau \in S_k} \text{sgn}(\tau) \mathbf{T}^\tau = -\text{sgn}(\pi) \text{alt}(\mathbf{T}).$$

We have shown $\text{alt}(\mathbf{T})^\pi = \text{sgn}(\pi) \text{alt}(\mathbf{T})$. □

Definition 5.7. (Wedge product).

Let V_1, \dots, V_k be vector spaces over the same field. We define the *wedge product* $\wedge : (V_1 \otimes \dots \otimes V_k) \otimes (V_1 \otimes \dots \otimes V_k) \rightarrow \text{alt}((V_1 \otimes \dots \otimes V_k) \otimes (V_1 \otimes \dots \otimes V_k))$ by $\mathbf{T} \wedge \mathbf{S} := \text{alt}(\mathbf{T} \otimes \mathbf{S})$.

¹ \mathbf{v}_1 is antisymmetric because its only permutation \mathbf{v}_1^σ is equal to $\mathbf{v}_1^\sigma = \text{sgn}(\sigma) \mathbf{v}_1$, where $\sigma = i$ is the identity. For the same reason, \mathbf{v}_2 is antisymmetric.

Lemma 5.8. (Lemma for associativity of wedge product).

Let V_1, \dots, V_k be vector spaces over the same field, and consider $\mathbf{T}, \mathbf{S} \in V_1 \otimes \dots \otimes V_k$. If $\text{alt}(\mathbf{T}) = \mathbf{0}$, then $\mathbf{T} \wedge \mathbf{S} = \mathbf{0} = \mathbf{S} \wedge \mathbf{T}$.

Proof. (This proof requires some abstract algebra. Understanding this proof is not really necessary to understand exterior powers; just take this theorem as an axiom).

Assume $\text{alt}(\mathbf{T}) = \mathbf{0}$. Let $T = \mathbf{v}_{i_1} \otimes \dots \otimes \mathbf{v}_{i_k}$ and $\mathbf{S} = \mathbf{v}_{j_1} \otimes \dots \otimes \mathbf{v}_{j_\ell}$. We must show $\text{alt}(\mathbf{T} \otimes \mathbf{S}) = \mathbf{0}$.

To do so, let H be the subgroup of S_{2k} whose elements fix all of j_1, \dots, j_ℓ , and consider the right cosets $\{H\sigma \mid \sigma \in S_{2k}\}$ of H in S_{2k} . Since these right cosets partition S_{2k} , then

$$\begin{aligned} \text{alt}(\mathbf{T} \otimes \mathbf{S}) &= \sum_{[\pi]_\tau \sigma \in \{\text{right cosets}\}} (-1)^{[\pi]_\tau \sigma} (\mathbf{T} \otimes \mathbf{S})^{\pi_\tau \sigma} = \sum_{\sigma \in S_{k+m}} \sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau \sigma} (\mathbf{T} \otimes \mathbf{S})^{\pi_\tau \sigma} \\ &= \sum_{\sigma \in S_{k+m}} \left(\sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} (\mathbf{T} \otimes \mathbf{S})^{\pi_\tau} \right)^\sigma. \end{aligned}$$

Since $[\pi]_\tau \in H$, where H is the subgroup of S_{2k} whose elements fix all of j_1, \dots, j_ℓ , then $(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} = \mathbf{T}^{[\pi]_\tau} \otimes \mathbf{S}$. With this, the innermost sum becomes

$$\sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} (\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} = \sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} \mathbf{T}^{[\pi]_\tau} \otimes \mathbf{S} = \left(\sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} \mathbf{T}^{[\pi]_\tau} \right) \otimes \mathbf{S}.$$

Now define $\pi \in S_k$ by $\pi = \tau^{-1}[\pi]_\tau \tau$, where $\tau = (i_1, \dots, i_k)$, using one-line notation (so $\tau(i) = j_i$). Then the above is

$$\left(\sum_{\pi \in S_k} (-1)^\pi \mathbf{T}^\pi \right) \otimes \mathbf{S} = \text{alt}(\mathbf{T}) \otimes \mathbf{S} = \mathbf{0} \otimes \mathbf{S} = \mathbf{0}.$$

The last equality follows by the seeming-multilinearity of \otimes . □

Theorem 5.9. (Wedge product is associative).

Let V_1, \dots, V_k be vector spaces over the same field. Then for all $\mathbf{T}, \mathbf{S}, \mathbf{R} \in V_1 \otimes \dots \otimes V_k$, we have $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R})$, and are therefore justified in denoting both as $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R}) := \mathbf{T} \wedge \mathbf{S} \wedge \mathbf{R}$.

Proof. We will show $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$; a similar argument shows $\mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R}) = \text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$.

First, we have by definition of \wedge that

$$(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \text{alt}((\mathbf{T} \wedge \mathbf{S}) \otimes \mathbf{R})$$

Subtracting $\text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$ from both sides and using linearity of alt , we get that

$$(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} - \text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R}) = \text{alt}((\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \otimes \mathbf{R}) = (\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R}.$$

If we show $(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R} = \mathbf{0}$, then our claim is true. The previous lemma says that if $\text{alt}(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) = \mathbf{0}$, then $(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R} = \mathbf{0}$. And this is true, since $\text{alt}(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) = \text{alt}(\mathbf{T} \wedge \mathbf{S}) - \text{alt}(\mathbf{T} \otimes \mathbf{S}) = \mathbf{T} \wedge \mathbf{S} - \mathbf{T} \wedge \mathbf{S} = \mathbf{0}$, by linearity of alt and with use of the fact that $\mathbf{T} \wedge \mathbf{S}$ is antisymmetric. □

Wedge product spaces and exterior powers

Theorem 5.10. (Properties of the wedge product).

Let V_1, \dots, V_k be vector space over a field $K \neq \mathbb{Z}/2\mathbb{Z}$, and consider the tensor product space $V_1 \otimes \dots \otimes V_k$. The wedge product \wedge satisfies the following properties...

1. \wedge looks as if it is bilinear, just as was the case with \otimes . That is, ...
 - 1.1. $(\mathbf{T} + \mathbf{S}) \wedge \mathbf{R} = \mathbf{T} \wedge \mathbf{R} + \mathbf{S} \wedge \mathbf{R}$ for all $\mathbf{T}, \mathbf{S}, \mathbf{R} \in V_1 \otimes \dots \otimes V_k$.
 - 1.2. $\mathbf{T} \wedge (\mathbf{S} + \mathbf{R}) = \mathbf{T} \wedge \mathbf{S} + \mathbf{T} \wedge \mathbf{R}$ for all $\mathbf{T}, \mathbf{S}, \mathbf{R} \in V_1 \otimes \dots \otimes V_k$.
 - 1.3. $(c\mathbf{T}) \wedge \mathbf{S} = c(\mathbf{T} \wedge \mathbf{S}) = \mathbf{T} \wedge (c\mathbf{S})$ for all $\mathbf{T}, \mathbf{S} \in V_1 \otimes \dots \otimes V_k$ and $c \in K$.
2. \wedge is associative, just as was the case with \otimes : $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R})$ for all $\mathbf{T}, \mathbf{S} \in V_1 \otimes \dots \otimes V_k$.
3. \wedge looks as if it is an antisymmetric map: $\mathbf{T} \wedge \mathbf{S} = -\mathbf{S} \wedge \mathbf{T}$ for all $\mathbf{T}, \mathbf{S} \in V_1 \otimes \dots \otimes V_k$.
4. \wedge is *skew-commutative*: if $\mathbf{T} \in V_1 \otimes \dots \otimes V_k$ and $\mathbf{S} \in V_1 \otimes \dots \otimes V_\ell$, then $\mathbf{S} \wedge \mathbf{T} = (-1)^{k+\ell}(\mathbf{T} \wedge \mathbf{S})$.
5. $\mathbf{T} \wedge \mathbf{T} = \mathbf{0}$ for all $\mathbf{T} \in V_1 \otimes \dots \otimes V_k$.

Proof. Property (1) follows by checking the definition $\mathbf{T} \wedge \mathbf{S} := \text{alt}(\mathbf{T} \otimes \mathbf{S})$. Property (2) was proved in Theorem 5.9. Property (3) follows from the definition of alt . Conditions (3) and (4) are logically equivalent, and conditions (3) and (5) are logically equivalent when $K \neq \mathbb{Z}/2\mathbb{Z}$. (Again, we need to require $K \neq \mathbb{Z}/2\mathbb{Z}$ here so that $2 \neq 0$, which allows division by 2). \square

In Remark 5.5, we said that we wanted to find a function $\wedge : (V_1 \otimes \dots \otimes V_k) \times (V_1 \otimes \dots \otimes V_k) \rightarrow V_1 \otimes \dots \otimes V_k$ for which the tuple $(\{\text{antisymmetric tensors in } V_1 \otimes \dots \otimes V_k\}, K, +, \cdot, \wedge)$ is an algebra. We have done so by constructing the wedge product \wedge , and now formalize this result with the next definitions.

Definition 5.11. (Wedge product space).

Let V_1, \dots, V_k be vector spaces over a field K . We define the *wedge product space* $V_1 \wedge \dots \wedge V_k$ to be the algebra

$$V_1 \wedge \dots \wedge V_k := (\text{alt}(V_1 \otimes \dots \otimes V_k), K, +, \cdot, \wedge).$$

Theorem 5.12. (Basis and dimension for wedge product spaces).

don't prove until have proven thm about basis for tensor product spaces

Theorem 5.13. (Exterior powers).

Let V be a vector space. We define the k th exterior power $\Lambda^k(V)$ of V to be the wedge product space $\Lambda^k(V) := V^{\wedge k}$.

Remark 5.14. Intuition of k -blades and volume

Definition 5.15. (Alternating function).

Let V_1, \dots, V_k, W be vector spaces over the same field. We say a function $\mathbf{f} : V_1 \times \dots \times V_k \rightarrow W$ is a k -alternating function iff for all $\sigma \in S_n$, we have $\mathbf{f}(\mathbf{v}_{\sigma(1)}, \dots, \mathbf{v}_{\sigma(k)}) = \text{sgn}(\sigma)\mathbf{f}(\mathbf{v}_1, \dots, \mathbf{v}_k)$. Equivalently, \mathbf{f} is k -alternating iff $\mathbf{f}(\mathbf{v}_1, \dots, \mathbf{v}_i, \mathbf{v}_{i+1}, \mathbf{v}_k) = -\mathbf{f}(\mathbf{v}_1, \dots, \mathbf{v}_{i+1}, \mathbf{v}_i, \dots, \mathbf{v}_k)$ for all $i \in \{1, \dots, k\}$.

When k is clear from the context, k -linear functions are called *alternating functions*.

Remark 5.16. ("Antisymmetric" vs. "alternating").

Many authors refer to the "antisymmetric tensors" of Definition [...] as "alternating tensors". Doing so is technically incorrect, because an "alternating tensor" is an element of a certain "quotient algebra". This quotient algebra is isomorphic to an algebra of antisymmetric tensors only when the "characteristic" of the field K is infinity. In other words, "antisymmetric tensors" are only the same as "alternating tensors" in certain special cases.

Given the previous paragraph, one might think that “alternating functions” should really be called “antisymmetric functions”. This is actually not the case. We are correct to define the functions of the above to be “alternating functions” because the alternating functions of the above definition correspond to alternating tensors (which, remember, are elements of a certain quotient algebra) via *the universal property of the exterior algebra*. So, alternating functions share the same level of generality as alternating tensors (which we have not defined). Antisymmetric tensors, which we have defined, are what are “more specific”.

We have not defined what the exterior algebra is (and won’t), but it is good to know this context.

Definition 5.17. (Vector space of alternating functions).

If V_1, \dots, V_k, W are vector spaces over a field K , then we use $(\text{alt}\mathcal{L})(V_1 \times \dots \times V_k \rightarrow W)$ to denote the vector space over K formed by the set of k -alternating functions $V_1 \times \dots \times V_k \rightarrow W$ under the operations of function addition and function scaling.

Theorem 5.18. (Universal property for exterior powers).

Let V_1, V_2, W be vector spaces over the same field, and let $\mathbf{f} : V_1 \times V_2 \rightarrow W$ be an alternating bilinear function. Then there exists a linear function $\mathbf{h} : V_1 \wedge V_2 \rightarrow W$ with $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ where \mathbf{h} uniquely depends on \mathbf{f} , and where $\mathbf{g} : V_1 \times V_2 \rightarrow V_1 \otimes V_2$.

In Remark 5.16, we mentioned that there exists a “universal property of the exterior algebra”. Note that this theorem is *not* the universal property of the exterior algebra, but a special case of it. (As mentioned before, we have not stated the universal property of the exterior algebra, and will not state it in this book).

Proof. The proof is similar to the proof of the universal property of tensor product spaces (Theorem 3.9). The only difference is that the maps we define in this proof are extended using antisymmetry and bilinearity, rather than just bilinearity. \square

Theorem 5.19. (Fundamental natural isomorphisms for exterior powers).

Theorem 3.34 stated that there are natural isomorphisms

$$\begin{aligned}\mathcal{L}(V_1 \times \dots \times V_k \rightarrow W) &\cong \mathcal{L}(V_1 \otimes \dots \otimes V_k \rightarrow W) \\ \mathcal{L}(V \rightarrow W) &\cong W \otimes V^* \\ (V \otimes W)^* &\cong V^* \otimes W^* \\ T_q^p(V) &\cong T_p^q(V^*)\end{aligned}$$

Analogously, there are natural isomorphisms

$$\begin{aligned}(\text{alt}\mathcal{L})(V_1 \times \dots \times V_k \rightarrow W) &\cong (\text{alt}\mathcal{L})(V_1 \wedge \dots \wedge V_k \rightarrow W) \\ (\text{alt}\mathcal{L})(V \rightarrow W) &\cong W \wedge V^* \\ (V \wedge W)^* &\cong V^* \wedge W^* \\ \Lambda^k(V)^* &\cong \Lambda^k(V^*)\end{aligned}$$

Proof. To show the first equation in the box, show that the map sending an alternating bilinear function to its unique linear counterpart defined on wedge product spaces (which is guaranteed to exist by the universal property for exterior powers) is a linear isomorphism. (This is what we did when we proved the corresponding fact for tensor product spaces; those steps can essentially be repeated for this proof. See Theorem 3.10). This proves the first equation for the case $k = 2$. The general result follows by induction.

To prove the second line, use a similar isomorphism as was presented at the end of Section 3.2, when we derived the natural isomorphism $\mathcal{L}(V \rightarrow W) \cong W \otimes V^*$. That is, take an element $\mathbf{f} \in (\text{alt}\mathcal{L})(V \rightarrow W)$, decompose it into a linear combination of “alternating elementary compositions”, and then send each alternating elementary composition $\mathbf{w} \circ \phi \mapsto \mathbf{w} \wedge \phi$. The formal check that this map is a linear isomorphism is essentially the same as the check described at the end of Section 3.2.

The third line in the box is proved similarly as was in Theorem 3.34; the only difference is that it is necessary to extend with antisymmetry and bilinearity rather than just bilinearity. The fourth line follows from the third line.

read <https://math.stackexchange.com/questions/18595/exterior-power-of-dual-space/18628#18628>. even better, ask Owen □

Pushforward and pullback

We now momentarily take a step backwards and introduce the *pushforward* and *pullback* operations for $\binom{k}{0}$ and $\binom{0}{k}$ tensors. This prepares us to understand pushforward and pullback for exterior powers.

Definition 5.20. (Pushforward of $\binom{k}{0}$ tensors).

Let V and W be finite-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \rightarrow W$. Using \mathbf{f} , we can *push forward* $\binom{k}{0}$ tensors in $T_0^k(V) = V^{\otimes k}$ to $\binom{k}{0}$ tensors in $T_0^k(W) = W^{\otimes k}$ as follows:

$$\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k \xrightarrow{\otimes_0^k \mathbf{f}} \mathbf{f}(\mathbf{v}_1) \otimes \dots \otimes \mathbf{f}(\mathbf{v}_k).$$

(As usual, we extend this map, which is stated in terms of elementary $\binom{k}{0}$ tensors, to act on arbitrary $\binom{k}{0}$ tensors with the seeming-multilinearity of \otimes). We have denoted this map by $\otimes_0^k \mathbf{f}$, so we can also state the above as

$$(\otimes_0^k \mathbf{f})(\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k) = \mathbf{f}(\mathbf{v}_1) \otimes \dots \otimes \mathbf{f}(\mathbf{v}_k).$$

The map $\otimes_0^k \mathbf{f} : T_0^k(V) \rightarrow T_0^k(W)$ is called the *pushforward*. Since we have extended it with the seeming-multilinearity of \otimes , the pushforward is a linear map.

Definition 5.21. (Pullback of $\binom{0}{k}$ tensors).

As in the previous definition, let V and W be finite-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \rightarrow W$. We can use the dual $\mathbf{f}^* : W^* \rightarrow V^*$ of \mathbf{f} to construct a map that *pulls back* elements of $T_k^0(W) = T_0^k(W^*)$ to elements of $T_k^0(V) = T_0^k(V^*)$, as follows:

$$\mathbf{w}_1 \otimes \dots \otimes \mathbf{w}_k \xrightarrow{\otimes_k^0 \mathbf{f}^*} \mathbf{f}^*(\mathbf{w}_1) \otimes \dots \otimes \mathbf{f}^*(\mathbf{w}_k).$$

(As in the previous definition, this map is extended to act on arbitrary $\binom{0}{k}$ tensors with the seeming-multilinearity of \otimes). We have denoted this map by $\otimes_k^0 \mathbf{f}^*$, so we can also state the above as

$$(\otimes_k^0 \mathbf{f}^*)(\mathbf{w}_1 \otimes \dots \otimes \mathbf{w}_k) = \mathbf{f}^*(\mathbf{w}_1) \otimes \dots \otimes \mathbf{f}^*(\mathbf{w}_k).$$

The map $\otimes_k^0 \mathbf{f}^* : T_k^0(W) \rightarrow T_k^0(V)$ is called the *pullback*. As was the case with the pushforward, the pullback is a linear map because its definition was extended with the seeming-multilinearity of \otimes .

Theorem 5.22. [Lee, p.320] (Properties of tensor pullbacks).

Having introduced pushforward and pullback for $\binom{k}{0}$ and $\binom{0}{k}$ tensors, we now are ready to look at pushforward and pullback for exterior powers.

Definition 5.23. (Pushforward on $\Lambda^k(V)$).

Let V and W be finite-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \rightarrow W$. Using \mathbf{f} , we can *push forward* elements of $\Lambda^k(V)$, $k \leq \max(\dim(V), \dim(W))$, to elements of $\Lambda^k(W)$ by using the *pushforward* $\Lambda^k \mathbf{f}$ on $\Lambda^k(V)$ defined by

$$\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_k \xrightarrow{\Lambda^k \mathbf{f}} \mathbf{f}(\mathbf{v}_1) \wedge \dots \wedge \mathbf{f}(\mathbf{v}_k).$$

We extend this map to act on arbitrary elements of $\Lambda^k(V)$ with the seeming-multilinearity and antisymmetry \wedge .

Definition 5.24. (Pullback on $\Lambda^k(W)$).

Let V and W be finite-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \rightarrow W$. Using \mathbf{f} , we can *pull back* elements of $\Lambda^k(W)$ to elements of $\Lambda^k(V)$, $k \leq \max(\dim(V), \dim(W))$, by using the *pullback* $\Lambda^k \mathbf{f}^*$ on $\Lambda^k(W)$ defined by

$$\mathbf{w}_1 \wedge \dots \wedge \mathbf{w}_k \xrightarrow{\Lambda^k \mathbf{f}^*} \mathbf{f}^*(\mathbf{w}_1) \wedge \dots \wedge \mathbf{f}^*(\mathbf{w}_k).$$

We extend this map to act on arbitrary elements of $\Lambda^k(W)$ with the seeming-multilinearity and antisymmetry \wedge .

Remark 5.25. (Star notation for pushforward and pullback).

The above notation of using $\otimes_k^0 \mathbf{f} : V \rightarrow W$ or $\Lambda^k \mathbf{f}$ for the pushforward and using $\otimes_0^k \mathbf{f}^* : W^* \rightarrow V^*$ or $\Lambda^k \mathbf{f}^*$ for the pullback is nonstandard. It is more common to denote a pullback by $\mathbf{f}^* : T_k^0(W) \rightarrow T_k^0(V)$ or $\mathbf{f}^* : \Lambda^k(W) \rightarrow \Lambda^k(V)$. Somewhat less commonly, a pushforward is denoted by $\mathbf{f}_* : T_0^k(V) \rightarrow T_0^k(W)$ or $\mathbf{f}_* : \Lambda^k(V) \rightarrow \Lambda^k(W)$.

Once the above definitions of pushforward and pullback are understood, this “star” notation can be useful. But it can be difficult to understand what pushforward and pullback are if this notation is used from the start, due to the potential for confusing the dual $\mathbf{f}^* : W^* \rightarrow V^*$ with either of the pullbacks $\mathbf{f}^* : T_k^0(W) \rightarrow T_k^0(V)$, $\mathbf{f}^* : \Lambda^k(W) \rightarrow \Lambda^k(V)$.

We will use the star notation after we define a pullback of a differential form in Chapter 9. Until then, we do not use this notation.

5.2 The determinant

Definition 5.26. (The determinant).

Let K be a field, and let $\hat{\mathbf{e}} = \{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n\}$ be the standard basis of K^n . We want to define a function $(K^n)^{\times n} \rightarrow K$ which, given $\mathbf{c}_1, \dots, \mathbf{c}_n \in K^n$, returns the n -dimensional volume of the parallelapiped spanned by $\mathbf{c}_1, \dots, \mathbf{c}_n$. We will denote this function by $\det : (K^n)^{\times n} \rightarrow K$. We require that \det satisfy the following axioms:

1. $\det(\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n) = 1$, since we want the unit n -cube to have an n -dimensional volume of 1.
2. \det is multilinear, because...
 - The volume of a parallelapiped that is the disjoint union of two smaller parallelapipeds should be the sum of the volumes of the smaller parallelapipeds.
 - Scaling one of the sides of a parallelapiped by $c \in K$ should increase that paralellapiped’s volume by a factor of c .
3. $\det(\mathbf{c}_1, \dots, \mathbf{c}_i, \dots, \mathbf{c}_j, \dots, \mathbf{c}_n) = 0$ when $\mathbf{c}_i = \mathbf{c}_j$ for all $\mathbf{c}_k \in K$, $k \in \{1, \dots, n\}$. This should hold because when two sides of a parallelapiped coincide, its n -dimensional volume is zero.

When $K \neq \mathbb{Z}/2\mathbb{Z}$ so that $2 \neq 0$, which enables division by 2, then, due to the multilinearity of \det , the third axiom is logically equivalent to \det being an alternating function. (Proof left as exercise). This is the case when $K = \mathbb{R}$, for example. The fact that \det is (almost always) alternating means that our intuitive assumptions about volume require that volume be *signed*, or *oriented*; the volume of the parallelapiped spanned by $\mathbf{c}_1, \dots, \mathbf{c}_j, \dots, \mathbf{c}_i, \dots, \mathbf{c}_n$ is the negation of the volume of the parallelapiped spanned by $\mathbf{c}_1, \dots, \mathbf{c}_i, \dots, \mathbf{c}_j, \dots, \mathbf{c}_n$.

The fact that the third axiom is logically equivalent to alternatingness also gives us a concise characterization of the determinant: $\det : (K^n)^{\times n} \rightarrow K$, when $K \neq \mathbb{Z}/2\mathbb{Z}$, is the unique multilinear alternating function satisfying $\det(\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n) = 1$.

Definition 5.27. (Determinant of a square matrix).

We define the *determinant of a square matrix* to be the result of applying \det to the column vectors of that matrix.

Theorem 5.28. (Consequent properties of the determinant).

5. \det is invariant under linearly combining input vectors into a different input vector. That is, $\det(\mathbf{c}_1, \dots, \mathbf{c}_i, \dots, \mathbf{c}_n) = \det(\mathbf{c}_1, \dots, \mathbf{c}_i + \sum_{j=1, j \neq i}^n d_j \mathbf{c}_j, \dots, \mathbf{c}_n)$ for all $i \in \{1, \dots, n\}$.
6. $\det(\mathbf{c}_1, \dots, \mathbf{c}_n) = 0$ iff $\{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ is a linearly dependent set.

Proof.

5. Using the axiom $\det(\mathbf{c}_1, \dots, \mathbf{c}_i, \dots, \mathbf{c}_j, \dots, \mathbf{c}_n) = 0$ when $\mathbf{c}_i = \mathbf{c}_j$ together with the multilinearity of the determinant, we have

$$\begin{aligned} \det(\mathbf{c}_1, \dots, \mathbf{c}_i, \dots, \mathbf{c}_n) &= \det(\mathbf{c}_1, \dots, \mathbf{c}_i, \dots, \mathbf{c}_n) + \sum_{j=1, j \neq i}^n \left(d_j \det(\mathbf{c}_1, \dots, \mathbf{c}_j, \dots, \mathbf{c}_j, \dots, \mathbf{c}_n) \right) \\ &= \det \left(\mathbf{c}_1, \dots, \mathbf{c}_i + \sum_{j=1, j \neq i}^n (d_j \mathbf{c}_j), \dots, \mathbf{c}_n \right). \end{aligned}$$

6.

($\det(\mathbf{c}_1, \dots, \mathbf{c}_n) = 0 \implies \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ is a linearly dependent set). If the input vectors are linearly dependent, we can use the invariance of \det under linearly combining some columns into others (which we just proved) to produce an equal determinant in which two columns are the same. By the third axiom, this determinant is zero.

($\det(\mathbf{c}_1, \dots, \mathbf{c}_n) = 0 \iff \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ is a linearly dependent set). Suppose for contradiction that the determinant of a set of n linearly independent vectors is zero. These n linearly independent vectors form a basis for K^n , so we have shown that the determinant of a basis set is zero. But then, using multilinearity together with the invariance of \det under linearly combining some vectors into a different vector, we can show that $\det(\mathbf{c}_1, \dots, \mathbf{c}_n) = 0$ for *all* $\mathbf{c}_1, \dots, \mathbf{c}_n \in K^n$. This contradicts the first axiom; we must have $\det(\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n) = 1$.

□

Derivation 5.29. (Permutation formula for the determinant).

We now derive *permutation formula* for the \det . The formula we obtain shows that the function \det specified in Definition [...] exists; since this formula is derived from the axioms of the determinant, it also shows that \det is unique.

Consider vectors $\mathbf{c}_1, \dots, \mathbf{c}_n \in K^n$, and² set $(a_j^i) := ([\mathbf{c}_i]_{\hat{\mathbf{e}}})^j$. Then...

²There is no hidden meaning behind the upper and lower indices on a_j^i here; we only want to consider an arbitrary $n \times n$ matrix of scalars in K , and prefer to think of this matrix as storing the coordinates of a $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ tensor rather those of a $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$ or $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$ tensor.

$$\begin{aligned}
\det((a_j^i)) &= \det(\mathbf{c}_1, \dots, \mathbf{c}_n) = \det\left(\sum_{i_1=1}^n a_{i_1}^1 \hat{\mathbf{e}}_{i_1}, \dots, \sum_{i_n=1}^n a_{i_n}^n \hat{\mathbf{e}}_{i_n}\right) \\
&= \sum_{i_1=1}^n \det\left(a_{i_1}^1 \hat{\mathbf{e}}_{i_1}, \dots, \sum_{i_n=1}^n a_{i_n}^n \hat{\mathbf{e}}_{i_n}\right) \\
&= \sum_{i_1=1}^n \det(a_{i_1}^1 \hat{\mathbf{e}}_{i_1}, \dots, \sum_{i_n=1}^n a_{i_n}^n \hat{\mathbf{e}}_{i_n}) \\
&\vdots \\
&= \sum_{i_1=1}^n \dots \sum_{i_n=1}^n \det(a_{i_1}^1 \hat{\mathbf{e}}_{i_1}, \dots, a_{i_n}^n \hat{\mathbf{e}}_{i_n}) \\
&= \sum_{i_1=1}^n \dots \sum_{i_n=1}^n \det(a_{i_1}^1 \hat{\mathbf{e}}_{i_1}, \dots, a_{i_n}^n \hat{\mathbf{e}}_{i_n}), \text{ where } i_1, \dots, i_n \text{ are distinct from each other} \\
&= \sum_{\sigma \in S_n} \det(a_{\sigma(1)}^1 \hat{\mathbf{e}}_{\sigma(1)}, \dots, a_{\sigma(n)}^n \hat{\mathbf{e}}_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} a_{\sigma(1)}^1 \dots a_{\sigma(n)}^n \det(\hat{\mathbf{e}}_{\sigma(1)}, \dots, \hat{\mathbf{e}}_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} a_{\sigma(1)}^1 \dots a_{\sigma(n)}^n \operatorname{sgn}(\sigma) \det(\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n) \\
&= \sum_{\sigma \in S_n} a_{\sigma(1)}^1 \dots a_{\sigma(n)}^n \operatorname{sgn}(\sigma)
\end{aligned}$$

Therefore, we have

$$\boxed{\det(\mathbf{c}_1, \dots, \mathbf{c}_n) = \sum_{\sigma \in S_n} a_{\sigma(1)}^1 \dots a_{\sigma(n)}^n \operatorname{sgn}(\sigma)}$$

In this derivation, we have mostly used the multilinearity of the determinant. Though, the expression labeled with “where i_1, \dots, i_n are distinct from each other” results from the previous line due to the third axiom of the determinant, $\det(\mathbf{c}_1, \dots, \mathbf{c}_i, \dots, \mathbf{c}_j, \dots, \mathbf{c}_n) = 0$ when $\mathbf{c}_i = \mathbf{c}_j$.

There are four major steps in the derivation. The first step is to use the multilinearity of the determinant to turn the determinant of (a_j^i) into a sum of the determinants of matrices that only have one nonzero entry in each column (see the line directly above the line labeled with “where i_1, \dots, i_n are distinct from each other”). The second step is to disregard all determinants in this previous sum whose matrix arguments have two or more columns that have their nonzero entries in the same row (i.e. whose matrix arguments are matrices of linearly dependent columns). This leaves us with a sum of determinants of diagonal matrices whose columns have been shuffled (this corresponds to the line labeled with “where i_1, \dots, i_n are distinct from each other” and the line directly below it). The third step, which corresponds to the third to last line, is to use multilinearity to pull out all the constants. The fourth step is to use the alternatingness of the determinant so that every determinant argument in the sum is the identity matrix; this results in multiplying each term in the sum by $\operatorname{sgn}(\sigma)$.

Theorem 5.30. (Determinant of a matrix is transpose-invariant).

Let \mathbf{A} be a square matrix with entries in K . Recall from the discussion after the statement of the permutation formula for the determinant that the determinant of a matrix is a sum of determinants of diagonal matrices whose columns have been shuffled. Each shuffled diagonal matrix in this sum can be momentarily converted to a diagonal matrix, transposed, and then re-shuffled (so that the columns of the shuffled-transposed-reshuffled matrix are in the order of the columns of the original shuffled diagonal matrix). Reversing the expansion that was accomplished with the multilinearity of \det in the derivation of the permutation formula for the determinant, we see that the sum of the determinants of these shuffled-transposed-reshuffled matrices is equal to the determinant of \mathbf{A}^\top . Therefore

$$\det(\mathbf{A}) = \det(\mathbf{A}^\top).$$

Theorem 5.31. (Laplace expansion for the determinant).

Consider an $n \times n$ matrix $\mathbf{A} = (a_j^i)$, and let \mathbf{A}_j^i denote the so-called *ij minor matrix* obtained by erasing the i th row and j th column of \mathbf{A} . We have

$$\begin{aligned}\det(\mathbf{A}) &= \sum_{i=1}^n a_j^i \det(\mathbf{A}_j^i) \text{ for all } i \in \{1, \dots, n\} \\ \det(\mathbf{A}) &= \sum_{j=1}^n a_j^i \det(\mathbf{A}_j^i) \text{ for all } j \in \{1, \dots, n\}\end{aligned}$$

The first equation is called the *Laplace expansion for the determinant along the i th row*, and the second equation is called the *Laplace expansion for the determinant along the j th column*. Note that each equation implies the other because $\det(\mathbf{A}) = \det(\mathbf{A}^\top)$.

Proof. We prove the second equation of the theorem.

Consider all terms in the permutation formula's sum for $\det(\mathbf{A})$ that have the factor a_j^i . Let \mathbf{B} denote the shuffled diagonal matrix that corresponds to one of these terms. We can view $\det(\mathbf{B})$ as $\det(\mathbf{B}) = \pm a_j^i \det(\mathbf{B}_j^i)$, where \mathbf{B}_j^i is the determinant of the matrix obtained by removing the i th column and j th row from \mathbf{B} . The \pm sign is a result of the fact that the matrices \mathbf{B} and \mathbf{B}_j^i may have different inversion counts.

The main effort of this proof is to determine the \pm sign and specify how the inversion counts of \mathbf{B} and \mathbf{B}_j^i differ.

As a first step, note that the difference in the inversion count between \mathbf{B} and \mathbf{B}_j^i is the number of inversions that involve a_j^i . Thus, our problem reduces to determining an expression for the number of inversions that involve a_j^i . So, divide the matrix \mathbf{B} into quadrants that are centered on a_j^i . Let k_1, k_2, k_3, k_4 be the number of inversions in the upper left, upper right, lower left, and bottom right corners of \mathbf{A} , respectively. The number of inversions involving a_j^i is $k_2 + k_3$. Since we know $k_1 + k_2 + 1 = i$ and $k_1 + k_3 + 1 = j$, we have $k_2 + k_3 = i + j - 2 - 2k_1 = i + j - 2(k_1 + 1)$. (We also know $k_1 + k_2 + k_3 + k_4 = n$, but this is not that helpful). Thus, if σ is the permutation corresponding to \mathbf{B} and π is the permutation corresponding to \mathbf{B}_j^i , then $\text{sgn}(\sigma) = \text{sgn}(\pi)(-1)^{i+j-2(k_1+1)} = \text{sgn}(\pi)(-1)^{i+j}$. Thus $\text{sgn}(\sigma) = (-1)^{i+j} \text{sgn}(\pi) \iff \text{sgn}(\pi) = (-1)^{i+j} \text{sgn}(\sigma)$.

So,

$$\begin{aligned}a_j^i \det(\mathbf{B}_j^i) &= a_j^i \sum_{\pi \in S_n} a_1^{\pi(1)} \dots \cancel{a_j^{\pi(i)}} \dots a_n^{\pi(n)} \text{sgn}(\pi) \\ &= a_j^i \sum_{\sigma \in S_n} a_1^{\sigma(1)} \dots \cancel{a_j^{\sigma(i)}} \dots a_n^{\sigma(n)} (-1)^{i+j} \text{sgn}(\sigma) \\ &= (-1)^{i+j} a_j^i \det(\mathbf{B})\end{aligned}$$

Thus $a_j^i \det(\mathbf{B}_j^i) = (-1)^{i+j} a_j^i \det(\mathbf{B}) \iff \det(\mathbf{B}) = (-1)^{i+j} a_j^i \det(\mathbf{B}_j^i)$. Now sum all of the \mathbf{B} 's (the diagonal shuffled matrices) to get $\det(\mathbf{A}) = \sum_{j=1}^n a_j^i \det(\mathbf{A}_j^i)$. \square

Theorem 5.32. (Determinant of an upper triangular matrix).

Theorem 5.33. (Adjoint and Cramer's rule).

Definition 5.34. (Determinant of a linear function).

Let V and W be finite-dimensional vector spaces of the same dimension, and let E and F be bases for V and W . We define the *determinant of a linear function* $\mathbf{f} : V \rightarrow W$ to be the determinant of the matrix of \mathbf{f} relative to E and F , $\det(\mathbf{f}) := \det([\mathbf{f}(E)]_F)$.

Remark 5.35. We have not yet shown that the determinant of a linear function $V \rightarrow V$ is well defined; we have not shown that it doesn't on the basis chosen for V . We will see that this is the case soon.

Theorem 5.36. (Determinant of a matrix is dual-invariant).

Let V and W be finite-dimensional vector spaces of the same dimension, and consider a linear function $\mathbf{f} : V \rightarrow W$. Consider also the dual $\mathbf{f} : W^* \rightarrow V^*$ (recall Definition 4.4). Then $\det(\mathbf{f}^*) = \det(\mathbf{f})$.

Proof. Recall from³ condition (3) of Definition 4.21 that if \mathbf{A} is the matrix of \mathbf{f} relative to orthonormal bases \widehat{U}_1 and \widehat{U}_2 , then the matrix of \mathbf{f}^* relative to the induced dual bases \widehat{U}_2^* and \widehat{U}_1^* is \mathbf{A}^\top . Since the determinant of a matrix is transpose-invariant (recall Theorem 5.30,) we have $\det(\mathbf{f}) = \det(\mathbf{A}) = \det(\mathbf{A}^\top) = \det(\mathbf{f}^*)$. \square

Theorem 5.37. (The pushforward on the top exterior power is multiplication by the determinant).

Let V and W be n -dimensional vector spaces over a field K , let $\mathbf{f} : V \rightarrow W$ be a linear function, and consider $\Lambda^n(V)$. We call $\Lambda^n(V)$ the *top exterior power of V* because n is the largest positive integer for which $\Lambda^k(V)$ is not a zero-dimensional vector space. Consequently, it is the exterior power of smallest dimension; $\dim(\Lambda^n(V)) = \binom{n}{n} = 1$.

Now consider the pushforward $\Lambda^n \mathbf{f} : \Lambda^n(V) \rightarrow \Lambda^n(W)$ on the top exterior power, which (recall [...]) is defined on elementary tensors by $\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n \mapsto \mathbf{f}(\mathbf{v}_1) \wedge \dots \wedge \mathbf{f}(\mathbf{v}_n)$ and is extended with multilinearity. Defined this way, the pushforward $\Lambda^n \mathbf{f}$ is a multilinear alternating map. This pushforward is also a map of 1-dimensional vector spaces, so it must be multiplication by a constant. We will determine what this constant is.

Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis of V , and consider the action of the pushforward on the basis vectors of E ,

$$\mathbf{f}(\mathbf{e}_1) \wedge \dots \wedge \mathbf{f}(\mathbf{e}_n).$$

Because $\Lambda^n(\mathbf{f})$ is multilinear and alternating, the wedge product on the left hand side closely resembles the determinant $\det(\mathbf{f}(\mathbf{e}_1), \dots, \mathbf{f}(\mathbf{e}_n)) = \det([\mathbf{f}(E)]_F) = \det(\mathbf{f})$. So, set $(a_j^i) = [\mathbf{f}(E)]_F$, and then use essentially the same argument as was made to derive the permutation formula on the left hand side of the above. We obtain

$$\begin{aligned} \mathbf{f}(\mathbf{e}_1) \wedge \dots \wedge \mathbf{f}(\mathbf{e}_n) &= \sum_{\sigma \in S_n} (a_{\sigma(1)}^1 \dots a_{\sigma(n)}^n \operatorname{sgn}(\sigma) \mathbf{e}_1 \wedge \dots \wedge \mathbf{e}_n) = \left(\sum_{\sigma \in S_n} a_{\sigma(1)}^1 \dots a_{\sigma(n)}^n \operatorname{sgn}(\sigma) \right) \mathbf{e}_1 \wedge \dots \wedge \mathbf{e}_n \\ &= \det([\mathbf{f}(E)]_F) (\mathbf{e}_1 \wedge \dots \wedge \mathbf{e}_n) = \det(\mathbf{f}) (\mathbf{e}_1 \wedge \dots \wedge \mathbf{e}_n). \end{aligned}$$

So, we have the statement on the basis E

$$\mathbf{f}(\mathbf{e}_1) \wedge \dots \wedge \mathbf{f}(\mathbf{e}_n) = \det(\mathbf{f}) (\mathbf{e}_1 \wedge \dots \wedge \mathbf{e}_n).$$

Using the seeming-multilinearity of \wedge , we can extend this fact to apply to any list of vectors in V :

$$\boxed{\mathbf{f}(\mathbf{v}_1) \wedge \dots \wedge \mathbf{f}(\mathbf{v}_n) = \det(\mathbf{f}) (\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n) \text{ for all } \mathbf{v}_1, \dots, \mathbf{v}_n \in V}$$

We can explicitly involve the pullback $\Lambda^n \mathbf{f}$ and write the above as

$$(\Lambda^n \mathbf{f})(\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n) = \det(\mathbf{f}) (\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n) \text{ for all } \mathbf{v}_1, \dots, \mathbf{v}_n \in V.$$

Thus, $\Lambda^n(\mathbf{f})$ is multiplication by $\det(\mathbf{f})$.

³Technically, the equivalent conditions of the definition we reference only apply to linear functions $V \rightarrow V$. This is not an issue because V and W have the same dimension; if we want to be very formal, we can use the linear function $\tilde{\mathbf{f}} : V \rightarrow V$ that is obtained from \mathbf{f} by identifying $W \cong V$ with the identification that sends basis vectors of W to basis vectors of V .

Theorem 5.38. (The pullback on the top exterior power is multiplication by the determinant).

Let V and W be n -dimensional vector spaces, and consider a linear function $\mathbf{f} : V \rightarrow W$. Consider additionally the dual $\mathbf{f}^* : W^* \rightarrow V^*$ (recall Definition 4.4). Then, applying the previous theorem and using that $\det(\mathbf{f}) = \det(\mathbf{f}^*)$, we see that the pullback $\Lambda^n \mathbf{f}^*$ on the top exterior power satisfies

$$(\Lambda^n \mathbf{f}^*)(\mathbf{w}_1 \wedge \dots \wedge \mathbf{w}_n) = \det(\mathbf{f})(\mathbf{w}_1 \wedge \dots \wedge \mathbf{w}_n) \text{ for all } \mathbf{w}_1, \dots, \mathbf{w}_n \in V.$$

That is,

$$\boxed{\mathbf{f}^*(\mathbf{w}_1) \wedge \dots \wedge \mathbf{f}^*(\mathbf{w}_n) = \det(\mathbf{f})(\mathbf{w}_1 \wedge \dots \wedge \mathbf{w}_n)}$$

Theorem 5.39. (Product rule for determinants). Let V, W and Z be finite-dimensional vector spaces of the same dimension, and consider linear functions $\mathbf{f} : V \rightarrow W$ and $\mathbf{g} : W \rightarrow Z$. Then $\det(\mathbf{g} \circ \mathbf{f}) = \det(\mathbf{g}) \det(\mathbf{f})$. Thus, if \mathbf{A} is an $m \times n$ matrix and \mathbf{B} is an $n \times p$ matrix, then $\det(\mathbf{BA}) = \det(\mathbf{B}) \det(\mathbf{A})$.

Proof. Set $n := \dim(V) = \dim(W) = \dim(Z)$. By the previous theorem, $\det(\mathbf{g} \circ \mathbf{f})$ satisfies

$$(\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1) \wedge \dots \wedge (\mathbf{g} \circ \mathbf{f})(\mathbf{v}_n) = \det(\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n)$$

for all $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$.

Notice that the left hand side is

$$\mathbf{g}(\mathbf{f}(\mathbf{v}_1)) \wedge \dots \wedge \mathbf{g}(\mathbf{f}(\mathbf{v}_n)) = \det(\mathbf{g})(\mathbf{f}(\mathbf{v}_1) \wedge \dots \wedge \mathbf{f}(\mathbf{v}_n)) = \det(\mathbf{g}) \det(\mathbf{f})(\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n).$$

Thus

$$\det(\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n) = \det(\mathbf{g}) \det(\mathbf{f})(\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n).$$

This is a statement on the basis vector $\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n$ of $\Lambda^n(V)$. Extending this statement to a statement on any vector $\mathbf{T} \in \Lambda^n(V)$, we have $\det(\mathbf{g} \circ \mathbf{f})\mathbf{T} = \det(\mathbf{g}) \det(\mathbf{f})\mathbf{T}$. Thus $(\det(\mathbf{g} \circ \mathbf{f}) - \det(\mathbf{g}) \det(\mathbf{f}))\mathbf{T} = \mathbf{0}$ for all $\mathbf{T} \in \Lambda^n(V)$. Since we can choose $\mathbf{T} \neq \mathbf{0}$, this forces $\det(\mathbf{g} \circ \mathbf{f}) - \det(\mathbf{g}) \det(\mathbf{f}) = 0$. \square

Theorem 5.40. (Determinant of an inverse function).

Let V and W be finite-dimensional vector spaces of the same dimension, and consider a linear function $\mathbf{f} : V \rightarrow W$. Then $\det(\mathbf{f}^{-1}) = \frac{1}{\det(\mathbf{f})}$.

Proof. We have $\det(\mathbf{f} \circ \mathbf{f}^{-1}) = \det(\mathbf{I}) = 1$, and $\det(\mathbf{f} \circ \mathbf{f}^{-1}) = \det(\mathbf{f}) \det(\mathbf{f}^{-1})$ by the previous theorem, so $\det(\mathbf{f}) \det(\mathbf{f}^{-1}) = 1$. \square

Theorem 5.41. (Determinant of a linear function is well-defined).

Let V and W be finite-dimensional vector spaces of the same dimension, and consider a linear function $\mathbf{f} : V \rightarrow W$. For simplicity, consider the linear function $\tilde{\mathbf{f}} : V \rightarrow V$ induced by \mathbf{f} , where the “second” V is thought of as coming from the identification $W \cong V$ that identifies basis vectors of W with basis vectors of V . Let E and F be bases for V . Then, the matrix $[\tilde{\mathbf{f}}(F)]_F$ of $\tilde{\mathbf{f}}$ relative to F and F is related to the matrix $[\tilde{\mathbf{f}}(E)]_E$ of $\tilde{\mathbf{f}}$ relative to E and E by $[\tilde{\mathbf{f}}(F)]_F = [\mathbf{F}]_F [\tilde{\mathbf{f}}(E)]_E [\mathbf{F}]_F^{-1}$ (see Theorem 2.66). Therefore, by the previous theorem, $\det(\tilde{\mathbf{f}})$ is well-defined, because $\det([\tilde{\mathbf{f}}(F)]_F) = \det([\mathbf{F}]_F [\tilde{\mathbf{f}}(E)]_E [\mathbf{F}]_F^{-1}) = \det([\mathbf{F}]_F) \det([\tilde{\mathbf{f}}(E)]_E) \det([\mathbf{F}]_F^{-1}) = \det([\tilde{\mathbf{f}}(E)]_E)$, so $\det([\tilde{\mathbf{f}}(F)]_F) = \det([\tilde{\mathbf{f}}(E)]_E)$.

5.3 Orientation of finite-dimensional vector spaces

Orientation is the mathematical formalization of the notions of “clockwise” and “counterclockwise”; it is the notion which distinguishes different “rotational configurations” from each other.

Our discussion of orientation will be as follows. First, we treat inner product spaces, so that we can speak of orthonormality; we will define an orientation on an inner product space to be a choice of an *ordered* orthonormal basis. Then, in order to complete the definition of orientation for inner product spaces, we introduce rotations in n -dimensions. After we finish the definition of orientation for inner product spaces, we end the subsection on oriented inner product spaces by presenting the fact that the determinant “tracks” orientation. This fact allows us to generalize the notion of orientation to finite dimensional vector spaces that may or may not have an inner product. Lastly, we show how the top exterior power of a finite-dimensional vector space can be used for the purposes of orientation.

First notions of orientation for inner product spaces

Definition 5.42. (Ordered basis).

We will formalize the notion of orientation relying on the concept of an ordered basis. An *ordered basis*, which is a basis for a finite-dimensional vector space in which the order that vectors are specified matters.

For example, if V is a 2-dimensional vector space, then the ordered bases $E_1 = \{\mathbf{e}_1, \mathbf{e}_2\}$ and $E_2 = \{\mathbf{e}_2, \mathbf{e}_1\}$ for V are not equal, $E_1 \neq E_2$.

Definition 5.43. (Permutation acting on an ordered basis).

Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be an ordered basis for some finite-dimensional vector space. Given a permutation $\sigma \in S_n$, we define $E^\sigma := \{\mathbf{e}_{\sigma(1)}, \dots, \mathbf{e}_{\sigma(n)}\}$.

We now discover a consequence of imposing that the bases under consideration be ordered.

Derivation 5.44. (Intuition for the antisymmetry of ordered bases).

Consider the plane \mathbb{R}^2 , and consider also two permutations of the standard ordered basis $\hat{\mathbf{e}} = \{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ for \mathbb{R}^2 : $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ and $\{\hat{\mathbf{e}}_1, -\hat{\mathbf{e}}_2\}$. (Draw these ordered bases out on paper). Notice that no matter how you rotate the entire second ordered basis (rotate each vector in the second ordered basis by the same amount), it is impossible to make all vectors from the second ordered basis simultaneously align with their counterparts from the first ordered basis. This is also impossible for the ordered bases $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ and $\{\hat{\mathbf{e}}_2, \hat{\mathbf{e}}_1\}$. Finally, consider the ordered bases $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ and $\{-\hat{\mathbf{e}}_2, \hat{\mathbf{e}}_1\}$ of \mathbb{R}^2 . It *is* possible to make each vector from the first ordered basis with its counterpart from the second ordered basis by rotating either the entire first ordered basis or the entire second ordered basis.

What we have discovered is that *swapping adjacent vectors in an ordered basis of two vectors produces an ordered basis that is equivalent under rotation to the ordered basis obtained from the original by negating one of the vectors that have been swapped*. We refer to this fact as the *antisymmetry of ordered bases*. The notion of “equivalent under rotation” is formalized in the following definition.

We now work to formalize the antisymmetry of ordered bases by stating it in the context of a 2-dimensional inner product space. The notion of *rotational equivalence* is what facilitate this formalization. Before we define rotational equivalence, however, we must define what a 2-rotation is.

Definition 5.45. (2-rotation).

Let V be a 2-dimensional inner product space, and let \hat{U} be an orthonormal ordered basis for V . A *2-rotation on the 2-dimensional inner product space V* is a linear function $\mathbf{R}_\theta : V \rightarrow V$ whose matrix relative to \hat{U} and \hat{U} is

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}$$

for some $\theta \in [0, 2\pi)$.

Now consider the case when V is an n -dimensional inner product space, $n > 2$. Suppose $\widehat{U} = \{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n\}$ is an orthonormal ordered basis for V . An *extension of a 2-rotation on the oriented subspace*⁴ $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ to V is a linear function $V \rightarrow V$ for which

- The map $(\mathbf{v}_i, \mathbf{v}_j) \mapsto \mathbf{R}_\theta(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n)$ is a 2-rotation on $\text{span}(\mathbf{v}_i, \mathbf{v}_j)$ for some i, j
- The map $(\mathbf{v}_1, \dots, \cancel{\mathbf{v}}_i, \dots, \cancel{\mathbf{v}}_j, \dots, \mathbf{v}_n) \mapsto \mathbf{R}_\theta(\mathbf{v}_1, \dots, \cancel{\mathbf{v}}_i, \dots, \cancel{\mathbf{v}}_j, \dots, \mathbf{v}_n)$ is the identity on $\text{span}(\mathbf{v}_1, \dots, \cancel{\mathbf{v}}_i, \dots, \cancel{\mathbf{v}}_j, \dots, \mathbf{v}_n)$.

Note that the extension of a 2-rotation restricts to a 2-rotation on the subspace $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$. It's also worth noting that the matrix of such an extension relative to \widehat{U} and \widehat{U} is, for some $i, j \in \{1, \dots, n\}$,

$$\begin{pmatrix} 1 & \dots & \text{\scriptsize ith column} \cos(\theta) & 0 & \text{\scriptsize jth column} -\sin(\theta) & 0 \\ 0 & & 0 & \vdots & 0 & \vdots \\ 0 & & \vdots & 1 & \vdots & \vdots \\ \vdots & & 0 & \vdots & 0 & \vdots \\ 0 & \dots & \sin(\theta) & 0 & \cos(\theta) & 1 \end{pmatrix}.$$

(The columns other than the i th and j th columns are the columns of the $n \times n$ identity matrix).

If \mathbf{R}_θ is a 2-rotation on a 2-dimensional inner product space or is an extension of a 2-dimensional rotation on an n -dimensional inner product space, it is simply called a *2-dimensional rotation*. In this looser terminology, the phrase “a 2-rotation defined on the oriented subspace $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ ” really means “an extension of a 2-rotation, defined on the oriented subspace $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ ”.

Definition 5.46. (Equivalence under rotation for 2-dimensional inner product spaces).

We define orthonormal ordered bases $\widehat{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ and $\widetilde{U} = \{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2\}$ of a 2-dimensional inner product space V to be *equivalent under rotation* iff there exists a 2-dimensional rotation \mathbf{R}_θ for which $\widetilde{U} = \mathbf{R}_\theta(\widehat{U})$. (Recall Definition 2.49 for the meaning of $\mathbf{R}_\theta(\widehat{U})$).

Theorem 5.47. (Antisymmetry of ordered bases for a 2-dimensional inner product space).

Now we see how the notion of rotational equivalence for 2-dimensional inner product spaces formalizes the antisymmetry of ordered bases. Let $\widehat{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ be an orthonormal ordered basis of a 2-dimensional inner product space V . When $\theta = -\frac{\pi}{2}$ or $\theta = \frac{\pi}{2}$, the matrix of \mathbf{R}_θ relative to \widehat{U} and \widehat{U} is

$$\pm \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Thus, as we noticed in the informal discussion (Derivation [...]), the following ordered bases are rotationally equivalent:

$$\begin{aligned} \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\} &\sim \{-\hat{\mathbf{u}}_2, \hat{\mathbf{u}}_1\} \sim \{\hat{\mathbf{u}}_2, -\hat{\mathbf{u}}_1\} \\ \{\hat{\mathbf{u}}_2, \hat{\mathbf{u}}_1\} &\sim \{-\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\} \sim \{\hat{\mathbf{u}}_1, -\hat{\mathbf{u}}_2\}. \end{aligned}$$

We now generalize the equivalence under rotation and the “swap negate principle” to n -dimensional inner product spaces.

Definition 5.48. (Equivalence under rotation for n -dimensional inner product spaces).

We define orthonormal ordered bases $\widehat{U} = \{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n\}$ and $\widetilde{U} = \{\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_n\}$ of an n -dimensional inner product space V to be *equivalent under rotation*, and thus write $\widehat{U} \sim \widetilde{U}$, iff there is a composition of 2-dimensional rotations $\mathbf{R} = \mathbf{R}_k \circ \dots \circ \mathbf{R}_1$ defined on some oriented subspaces $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ for which $\mathbf{R}(\widehat{U}) = \widetilde{U}$. This notion of “equivalence under rotation” is indeed an equivalence relation on orthonormal ordered bases of V .

⁴When we say “oriented subspace”, we mean that the orientation of $\text{span}(\mathbf{u}_i, \mathbf{u}_j)$ is given by the ordered basis $\{\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j\}$.

Theorem 5.49. (Antisymmetry of ordered bases for an n -dimensional inner product space).

Similarly to what was done in the previous derivation, we use $\theta = -\frac{\pi}{2}$ and $\theta = \frac{\pi}{2}$ in the matrix relative to bases of a 2-dimensional rotation defined on the oriented subspace $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ to obtain the formal statement of the antisymmetry of ordered bases: for any orthonormal ordered basis $\hat{U} = \{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n\}$ of V , we have

$$\{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}, \dots, \hat{\mathbf{u}}_n\} \sim -\{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_{i+1}, \hat{\mathbf{u}}_i, \dots, \hat{\mathbf{u}}_n\}.$$

Equivalently, for any orthonormal ordered basis $\hat{U} = \{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n\}$ of V and any permutation $\sigma \in S_n$,

$$\{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}, \dots, \hat{\mathbf{u}}_n\} \sim \text{sgn}(\sigma) \{\hat{\mathbf{u}}_{\sigma(1)}, \dots, \hat{\mathbf{u}}_{\sigma(n)}\}.$$

Definition 5.50. (Orientation of permuted ordered bases).

Let V be an n -dimensional inner product space, and fix an orthonormal ordered basis $\hat{U} = \{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n\}$ for V . Suppose $\hat{U}^\sigma = \{\hat{\mathbf{u}}_{\sigma(1)}, \dots, \hat{\mathbf{u}}_{\sigma(n)}\}$, where $\sigma \in S_n$, is a permutation of \hat{U} that is not rotationally equivalent to \hat{U} (so choose any σ with $\text{sgn}(\sigma) = -1$). By the antisymmetry of ordered bases, any other permutation $\hat{U}^\pi = \{\hat{\mathbf{u}}_{\pi(1)}, \dots, \hat{\mathbf{u}}_{\pi(n)}\}$, $\pi \in S_n$, of \hat{U} is rotationally equivalent either to \hat{U} or to \hat{U}^σ . In other words, there are only two equivalence classes⁵ of “equivalence under rotation”.

We can now begin to set up the notion of orientation. An *orientation for the n -dimensional inner product space V* is a choice of an orthonormal ordered basis \hat{U} for V . Iff V is given the orientation \hat{U} , then the *orientation of a permutation of \hat{U} (relative to \hat{U})* is then defined to be *positive* iff that permutation of \hat{U} is rotationally equivalent to \hat{U} , and is defined to be *negative* otherwise. Per the previous paragraph, every permutation of \hat{U} is either positively oriented or negatively oriented relative to \hat{U} .

Remark 5.51. (The formalization of “counterclockwise” and “clockwise”).

At the beginning of this section, we said that orientation would formalize the notions of “clockwise” and “counterclockwise”. This formalization has been achieved by the previous definition.

A *counterclockwise rotational configuration* is another name for the orientation given to \mathbb{R}^3 by the standard basis, *when we use the normal human convention* of drawing the ordered basis $\hat{\mathcal{E}} = \{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ such that $\hat{\mathcal{E}}$ can be rotated so that $\hat{\mathbf{e}}_1$ points out of the page, $\hat{\mathbf{e}}_2$ points to the right, and $\hat{\mathbf{e}}_3$ points upwards on the page. In this visual convention, the direction of each basis vector corresponds to its position in $\hat{\mathcal{E}}$.

A *clockwise rotational configuration* then corresponds to the ordered bases which are not rotationally equivalent to $\hat{\mathcal{E}}$. One such ordered basis, $\{-\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ can be depicted using the visual convention just established by drawing $\hat{\mathbf{e}}_1$ as pointing into the page (i.e. $-\hat{\mathbf{e}}_1$ points out of the page), $\hat{\mathbf{e}}_2$ as pointing to the right, and $\hat{\mathbf{e}}_3$ as pointing upwards on the page.

We could have easily picked a different visual convention (i.e. a different permutation of in/out, left/right, up/down) to represent the ordering of the basis that is considered to orient the space.

At this point, we need some definitions and facts about n -rotations before we complete our development of orientation.

Rotations in n -dimensions

Definition 5.52. (n -rotation).

Let V be an n -dimensional inner product space, and consider $k \leq n$. An *extension of a k -rotation on the oriented subspace $\text{span}(\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_k})$ to V* is a linear function $\mathbf{R} : V \rightarrow V$ such that

- The map $(\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_k}) \mapsto \mathbf{R}(\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_k}, \dots, \mathbf{v}_n)$ is an n -rotation on $\text{span}(\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_k})$ for some i_1, \dots, i_k .

⁵I find this relatively surprising. My intuition is that there would be something like 2^n or $n!$ equivalence classes of “equivalence under rotation” in n dimensions, but nope! There are 2 equivalence classes of “equivalence under rotation” for every n .

- The map $(\mathbf{v}_1, \dots, \mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_k}, \dots, \mathbf{v}_n) \mapsto \mathbf{R}(\mathbf{v}_1, \dots, \mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_k}, \dots, \mathbf{v}_n)$ is the identity on $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_k}, \dots, \mathbf{v}_n)$.

We define a n -rotation to be a composition of extensions of $(n-1)$ -rotations.

Theorem 5.53. Every n -rotation has determinant 1. [...]. Therefore, every n -rotation is an orthogonal linear function.

Proof. Every n -rotation is a composition of 2-rotations. Since 2-rotations have determinant 1, it follows that any n -rotation has determinant 1. \square

Lemma 5.54. (3-rotations acting on 2-dimensional orthonormal oriented subspaces).

Let V be an inner product space with $\dim(V) \geq 2$, and consider 3-dimensional subspaces \widetilde{W} and W of V . If $\widetilde{U} = \{\widetilde{\mathbf{u}}_1, \widetilde{\mathbf{u}}_2\}$ and $\widehat{U} = \{\widehat{\mathbf{u}}_1, \widehat{\mathbf{u}}_2\}$ are orthonormal ordered bases of \widetilde{W} and W , respectively, then there exists a 3-rotation which takes \widetilde{U} to \widehat{U}^σ for some $\sigma \in S_2$.

Proof. By definition, a 3-rotation is of the form $\mathbf{R} = \mathbf{R}_\gamma \circ \mathbf{R}_\beta \circ \mathbf{R}_\alpha$, where $\mathbf{R}_\alpha, \mathbf{R}_\beta$, and \mathbf{R}_γ are 2-rotations. (Sidenote: $\alpha, \beta, \gamma \in [0, 2\pi)$ are called *Euler angles*).

We must show that there exist α, β, γ for which $\mathbf{R}(\widetilde{U}) = \widehat{U}$. We choose α, β, γ such that $\mathbf{R}(\widetilde{\mathbf{u}}_1) = \widehat{\mathbf{u}}_1$. **proof that such α, β, γ exist?** 2-rotations are orthogonal linear functions, so they preserve the orthonormality of bases. Thus, $\mathbf{R}(\widetilde{\mathbf{u}}_2)$ must have length 1 and be orthogonal to $\mathbf{R}(\widetilde{\mathbf{u}}_1) = \widehat{\mathbf{u}}_1$, so $\mathbf{R}(\widetilde{\mathbf{u}}_2)$ is either $\widehat{\mathbf{u}}_2$ or $-\widehat{\mathbf{u}}_2$. Therefore, $\mathbf{R}(\widetilde{U}) = \{\widehat{\mathbf{u}}_1, \pm \widehat{\mathbf{u}}_2\}$. More formally, $\mathbf{R}(\widetilde{U}) = \widehat{U}^\sigma$ for some $\sigma \in S_2$. \square

Remark 5.55. The contribution of the previous lemma to the theorem we prove next is that the previous lemma captures the notion of rotating a lesser dimensional subspace within a higher dimensional ambient vector space. This machinery is required in the next theorem in the case when n is odd.

Theorem 5.56. (n -rotations acting on orthonormal ordered bases).

Let V be an n -dimensional vector space, let $k \leq n$, and consider k -dimensional subspaces \widetilde{W} and W of V . If $\widetilde{U} = \{\widetilde{\mathbf{u}}_1, \dots, \widetilde{\mathbf{u}}_n\}$ and $\widehat{U} = \{\widehat{\mathbf{u}}_1, \dots, \widehat{\mathbf{u}}_n\}$ are orthonormal ordered bases of \widetilde{W} and W , respectively, then there is an n -rotation taking \widetilde{U} to some permutation \widehat{U}^σ of \widehat{U} .

Proof. When $n \in \{2, 3\}$, the previous lemma yields the desired n -rotation as a composition of 3-rotations (for $n = 2$, just take the restriction of the 3-rotations). We need to show that the theorem holds when $n > 3$. We consider the cases when n is even and n is odd.

If n is even, then, for $i \in \{1, \dots, n\}$, let \mathbf{R}_i be the 3-rotation taking $\{\widehat{\mathbf{u}}_i, \widehat{\mathbf{u}}_{i+1}\}$ to $\{\widetilde{\mathbf{u}}_{\sigma_i(i)}, \widetilde{\mathbf{u}}_{\sigma_i(i+1)}\}$, where $\sigma_i \in S(\{i, i+1\})$ is some permutation. Then $\mathbf{R}_{\frac{n}{2}} \circ \dots \circ \mathbf{R}_1$ is a composition of 3-rotations taking \widehat{U} to \widetilde{U}^σ , where $\sigma \in S_n$ is the permutation defined by $\sigma(i) = \sigma_{j(i)}(i)$, where $j(i) = i$ when i is odd and $j(i) = i-1$ when i is even.

If n is odd, then let \mathbf{R} be the 3-rotation taking $\{\widehat{\mathbf{u}}_{n-2}, \widehat{\mathbf{u}}_{n-1}, \widehat{\mathbf{u}}_n\}$ to $\{\widetilde{\mathbf{u}}_{\sigma_{n-2}(n-2)}, \widetilde{\mathbf{u}}_{\sigma_{n-2}(n-1)}, \widetilde{\mathbf{u}}_{\sigma_{n-2}(n)}\}$, where $\sigma_{n-2} \in S(\{n-2, n-1, n\})$ is some permutation.

By the previous case (when n was even), there is a composition $\mathbf{R}_{\frac{n-3}{2}} \circ \dots \circ \mathbf{R}_1$ of 3-rotations taking $\widetilde{U} - \{\widetilde{\mathbf{u}}_{n-2}, \widetilde{\mathbf{u}}_{n-1}, \widetilde{\mathbf{u}}_n\}$ to $\widehat{U} - \{\widehat{\mathbf{u}}_{n-2}, \widehat{\mathbf{u}}_{n-1}, \widehat{\mathbf{u}}_n\}^\sigma$, for some permutation $\sigma \in S_n$. Thus $\mathbf{R} \circ \mathbf{R}_{\frac{n-3}{2}} \circ \dots \circ \mathbf{R}_1$ is a composition of 3-rotations taking \widetilde{U} to \widehat{U}^π , where $\pi(i) = \sigma(i)$ when $i \in \{1, \dots, n-3\}$ and $\pi(i) = \sigma_{n-2}(i)$ when $i \in \{n-2, n-1, n\}$. \square

Completing the definition of orientation for inner product spaces

Definition 5.57. (Orientation of arbitrary ordered bases).

Let V be an n -dimensional inner product space, and fix an orthonormal basis $\widehat{U} = \{\widehat{\mathbf{u}}_1, \dots, \widehat{\mathbf{u}}_n\}$ for V . We know how to “orient” ordered bases for V that happen to be permutations of \widehat{U} . Now, we generalize the notion of orientation so that it applies to any ordered orthonormal basis of V .

We define the *orientation of an orthonormal ordered basis E of V* that is not necessarily a permutation of \widehat{U} to be the orientation of the unique permutation \widehat{U}^σ , $\sigma \in S^n$, of \widehat{U} for which there exists an n -rotation taking E to \widehat{U}^σ .

Then, we define the *orientation of an arbitrary orthonormal ordered basis E of V* to be the orientation of the unique orthonormal basis \widehat{U}_E obtained from performing the Gram-Schmidt process on E (see Theorem [...]).

Theorem 5.58. (The determinant tracks orientation).

Let V be an n -dimensional inner product space with an orientation given by an orthonormal ordered basis \widehat{U} . Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be any ordered basis (not necessarily orthonormal) of V . We have $\det([\mathbf{E}]_{\widehat{U}}) > 0$ iff E is positively oriented relative to \widehat{U} , and $\det([\mathbf{E}]_{\widehat{U}}) < 0$ iff E is negatively oriented relative to \widehat{U} .

Proof. This proof has two overarching steps. First, we pass the definition of orientation for arbitrary ordered bases of V to the definition of orthonormal ordered bases of V by obtaining an orthonormal ordered basis \widehat{U}_E from E . Then we pass the definition of orientation for orthonormal ordered bases of V that are not permutations of \widehat{U} to the definition of orientation for orthonormal ordered bases of V that are permutations of \widehat{U} .

To begin the first step, consider $\det([\mathbf{E}]_{\widehat{U}}) = \det([\mathbf{e}_1]_{\widehat{U}}, \dots, [\mathbf{e}_n]_{\widehat{U}})$, and perform Gram-Schmidt on $\{[\mathbf{e}_1]_{\widehat{U}}, \dots, [\mathbf{e}_n]_{\widehat{U}}\}$. In the i th step of Gram-Schmidt, a linear combination of the vectors $[\mathbf{e}_1]_{\widehat{U}}, \dots, [\mathbf{e}_i]_{\widehat{U}}, \dots, [\mathbf{e}_n]_{\widehat{U}}$ is added to $[\mathbf{e}_i]_{\widehat{U}}$. Recall from Theorem 5.28 that the determinant is invariant under linearly combining input vectors into a different input vector. Therefore, performing Gram-Schmidt does not change the determinant. That is, if $\widehat{U}_E = \{\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_n\}$ is the orthonormal basis obtained by performing Gram-Schmidt on E , then

$$\det([\mathbf{E}]_{\widehat{U}}) = \det([\mathbf{e}_1]_{\widehat{U}}, \dots, [\mathbf{e}_n]_{\widehat{U}}) = \det([\tilde{\mathbf{u}}_1]_{\widehat{U}}, \dots, [\tilde{\mathbf{u}}_n]_{\widehat{U}}) = \det([\widehat{U}_E]_{\widehat{U}}).$$

In performing this first step of the proof, the determinant has stayed the same as we've passed from E to \widehat{U}_E . We now show that the determinant continues to stay the same as we pass from \widehat{U}_E to some permutation \widehat{U}^σ of \widehat{U} .

Theorem 5.56 says that there is a n -rotation \mathbf{R} taking \widehat{U}_E to \widehat{U}^σ , for some $\sigma \in S_n$, and Theorem 5.53 guarantees that $\det(\mathbf{R}) = 1$. Thus, since $\widehat{U}^\sigma = \mathbf{R}(\widehat{U}_E)$, we have

$$\det([\widehat{U}_E]_{\widehat{U}}) = \det([\mathbf{R}(\widehat{U}_E)]_{\widehat{U}}) \det([\widehat{U}_E]_{\widehat{U}}) = \det([\mathbf{R} \circ \mathbf{I}](\widehat{U}_E))_{\widehat{U}} = \det([\mathbf{R}(\widehat{U}_E)]_{\widehat{U}}) = \det([\widehat{U}^\sigma]_{\widehat{U}}).$$

To conclude the proof, we will show that $\det([\widehat{U}^\sigma]_{\widehat{U}}) = \text{sgn}(\sigma) \det([\widehat{U}]_{\widehat{U}}) = \text{sgn}(\sigma)$. Since any permutation is a composition of “swaps” (a “swap” is a permutation defined on a two-element set), then \widehat{U}^σ can be obtained from \widehat{U} by repeatedly swapping vectors in \widehat{U} . Whenever vectors are swapped in the determinant, the sign of the determinant is multiplied by -1 . This accounts for the $\text{sgn}(\sigma)$ factor in the equation $\det([\widehat{U}^\sigma]_{\widehat{U}}) = \text{sgn}(\sigma) \det([\widehat{U}]_{\widehat{U}}) = \text{sgn}(\sigma)$. \square

We state the next theorem only for completeness. (We will not use the forward implication of the next theorem, and our last use of the reverse implication, which was already shown, was in the proof of Theorem 5.58).

Orientation of finite-dimensional vector spaces

The fact that the determinant tracks orientation is the main result of our discussion of orientation. Because determinants do not rely on the existence of an inner product, the determinant can be used to generalize the notion of orientation to any finite-dimensional vector space.

Definition 5.59. (Orientation of a finite-dimensional vector space).

Let V be a finite-dimensional vector space (not necessarily an inner product space). An *orientation on V* is a choice of ordered basis E for V . (Notice here that E is not necessarily orthonormal, because V might not have an inner product!). If we have given V the orientation E , then we say that an ordered basis F of V is *positively oriented (relative to E)* iff $\det([\mathbf{F}]_E) > 0$, and that F is *negatively oriented (relative to E)* iff $\det([\mathbf{F}]_E) < 0$.

A finite-dimensional vector space that has an orientation is called an *oriented (finite-dimensional) vector space*.

Remark 5.60. (Antisymmetry of ordered bases).

Notice that we still have the previous antisymmetry of ordered bases due to the antisymmetry of the determinant.

We now show how the top exterior power of a vector space can be used to describe orientation.

Theorem 5.61. (Orientation with n -wedges).

Let V be a finite-dimensional vector space with an orientation $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. All positively oriented bases of V are scalar multiples of E , and all negatively oriented bases of V are scalar multiples of $-E$, where $-E = E^\sigma$ for some σ with $\text{sgn}(\sigma) < 0$.

Notice, we can identify $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ with $\mathbf{e}_1 \wedge \dots \wedge \mathbf{e}_n \in \Lambda^n(V)$, because the antisymmetry of ordered bases is manifested in elements of $\Lambda^n(V)$ due to the antisymmetry \wedge . Once one has noticed this, it is a natural next step to check that the union of the sets of positively oriented and negatively oriented bases, when considered under the operations of “basis addition” and multiplication by a scalar, is a vector space that is isomorphic to $\Lambda^n(V)$.

Therefore, another way to give an orientation to a finite-dimensional vector space is to choose an element of $\Lambda^n(V)$. Notice also that the fact that the pushforward on the top exterior power is multiplication by the determinant (recall Theorem 5.37) plays nicely into this interpretation: once an orientation $\mathbf{e}_1 \wedge \dots \wedge \mathbf{e}_n$ of V has been chosen, then we have $\mathbf{f}_1 \wedge \dots \wedge \mathbf{f}_n = \det(\mathbf{f})\mathbf{e}_1 \wedge \dots \wedge \mathbf{e}_n$, where \mathbf{f} is the linear function $V \rightarrow V$ sending $\mathbf{e}_i \mapsto \mathbf{f}_i$.

5.4 The cross product

This section gives an intuitive and formal understanding of the cross product.

Lemma 5.62. (The algebraic dot product on \mathbb{R}^n is positive definite, and therefore cancelable).

The algebraic dot product on \mathbb{R}^n is *positive definite*; that is, it satisfies the property $(\mathbf{v} \cdot \mathbf{v} = 0 \iff \mathbf{v} = \mathbf{0})$.

As a consequence, we have the fact that, when $\mathbf{v} \in \mathbb{R}^n$ is nonzero and $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, then $((\mathbf{v}_1 \cdot \mathbf{v} = \mathbf{v}_2 \cdot \mathbf{v}) \implies \mathbf{v}_1 = \mathbf{v}_2)$.

Proof. We first show that the dot product on \mathbb{R}^n satisfies the property $(\mathbf{v} \cdot \mathbf{v} = 0 \iff \mathbf{v} = \mathbf{0})$. The reverse implication follows immediately. For the forward implication, let $\mathbf{v} \in \mathbb{R}^n$, and suppose $\mathbf{v} \cdot \mathbf{v} = 0$; we must show $\mathbf{v} = \mathbf{0}$. We have $\mathbf{v} \cdot \mathbf{v} = \sum_{i=1}^n ([\mathbf{v}]_{\hat{\mathbf{e}}})_i^2$. Each term in the sum is a nonnegative number. Therefore, the sum is only zero if all terms in the sum are zero, so $([\mathbf{v}]_{\hat{\mathbf{e}}})_i = 0$ for each i , which means $\mathbf{v} = \mathbf{0}$.

Now we show that when $\mathbf{v} \in \mathbb{R}^n$ is nonzero and $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, then $((\mathbf{v}_1 \cdot \mathbf{v} = \mathbf{v}_2 \cdot \mathbf{v}) \implies \mathbf{v}_1 = \mathbf{v}_2)$. If $\mathbf{v}_1 \cdot \mathbf{v} = \mathbf{v}_2 \cdot \mathbf{v}$, then $(\mathbf{v}_1 - \mathbf{v}_2) \cdot \mathbf{v} = \mathbf{0}$ by the bilinearity of \cdot . Since $\mathbf{v} \neq \mathbf{0}$, then we must have $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$ due to the positive definiteness of \cdot . That is, $\mathbf{v}_1 = \mathbf{v}_2$. \square

Derivation 5.63. Prove that $(\mathbf{v}_1 \bar{\wedge} \dots \bar{\wedge} \mathbf{v}_{n-1}) \cdot \mathbf{x} = \det(\mathbf{v}_1, \dots, \mathbf{v}_{n-1}, \mathbf{x})$.

Then derive induced operation \times on $V \times V$ corresponding to $\bar{\wedge}$ when $\dim(V) = 3$

Remark 5.64. (Interpretation of k -wedges as oriented volumes).

Definition 5.65. (Cross product). The *cross product* of vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ is the unique vector $(\mathbf{v}_1 \times \mathbf{v}_2) \in \mathbb{R}^3$ for which

$$(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{x} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^3.$$

Proof. Let $\mathbf{x} \in \mathbb{R}^3$. We need to show that the vector $\mathbf{c} \in \mathbb{R}^3$ for which $\mathbf{c} \cdot \mathbf{x} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{x})$ is unique. So, let $\mathbf{d} \in \mathbb{R}^3$ satisfy this relation as well; suppose $\mathbf{d} \cdot \mathbf{x} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{x})$. Then $\mathbf{c} \cdot \mathbf{x} = \mathbf{d} \cdot \mathbf{x}$. Since \cdot is cancelable (see the previous lemma), then $\mathbf{c} = \mathbf{d}$. \square

Theorem 5.66. (Magnitude, direction of cross product).

right hand rule

$\sin(\theta)$ formula

Part II

Calculus and basic topology

6

Review of calculus

Notation for covariance and contravariance is not used in this chapter. The use of both upper and lower indices to distinguish between “covariant” and “contravariant” will not be used in the following chapter of multivariable calculus review, even though these concepts have already been introduced. Only lower indices will be used.

6.1 Single-variable calculus

Definition 6.1. (The derivative).

$$f'(t) := \lim_{\Delta t \rightarrow 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}$$

Definition 6.2. (Common notation for derivatives).

Suppose $U \subseteq \mathbb{R}$ is an open set and $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is a differentiable function. We define the *Leibniz* and *operator* notations for differentiation.

Leibniz notation

$$\frac{df}{dt} := f'$$

$$\left. \frac{df}{dt} \right|_{t=t_0} = f'(t_0)$$

Operator notation

$$\frac{d}{dt}f := f'$$

$$\frac{d}{dt}f(t) := \left(\frac{d}{ds}f \right) \Big|_{s=t} = f'(t)$$

$$\frac{df(t)}{dt} := \frac{d}{dt}f(t)$$

Definition 6.3. (Derivative with respect to a function).

This definition formalizes a convention that is often used but rarely explained.

Suppose $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ and $x : V \subseteq \mathbb{R} \rightarrow \mathbb{R}$ satisfy the differentiability conditions of the chain rule (see Theorem [...]), so that $f \circ x$ is differentiable. We define $\frac{df}{dx} : ? \rightarrow ?$ to be the function defined by

$$\left. \frac{df}{dx} \right|_{t=t_0} := \left. \frac{df}{dt} \right|_{t=x(t_0)}.$$

That is, $\frac{df}{dx} := f' \circ x$.

With this notation, the chain rule is

$$\frac{d(f \circ x)}{dt} = \frac{df}{dx} \frac{dx}{dt}.$$

This is more elegant than the following statement of the chain rule employing a substitution, which is often presented in standard calculus textbooks:

$$\left. \frac{d(f \circ x)}{dt} \right|_{t=t_0} = \frac{df(u)}{du} \frac{du(t)}{dt}, \text{ where } u = x(t_0).$$

check the above standard chain rule statement

Remark 6.4. (Letters in the denominator).

The definition $\frac{df}{dt} := f'$ from Definition 6.2 technically implies that $\frac{df}{da} = \frac{df}{db} = \frac{df}{dc} = \dots = \frac{df}{dz} = f'$; it does not matter which letter is used in the “denominator”.

On the other hand, when the letter in the “denominator” represents a function $\mathbb{R} \rightarrow \mathbb{R}$, the letter used *does* matter.

In calculus, we often intentionally conflate real numbers with real-valued functions so that we can start with theorems of the form “if $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is a differentiable function and $x \in \mathbb{R}$, and ..., then ...” and then think of the real number x as a real-valued function, apply the notion of derivative with respect to a function, and leverage the chain rule to obtain theorems of the form “if $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ and $x : V \subseteq \mathbb{R} \rightarrow \mathbb{R}$ are differentiable functions, and ..., then ...”. Since there is always the potential for real numbers to become real-valued functions, it’s best to think of the letters in the “denominator” as mattering in all cases.

Of course, the choice of letter in the “denominator” inherently matters for partial derivatives.

Definition 6.5. (Antiderivative). Let $f : I \rightarrow \mathbb{R}$ be a function defined on an interval I . An *antiderivative of f (on I)* is a function $F : I \rightarrow \mathbb{R}$ whose derivative is f , i.e., $F' = f$.

We denote a particular antiderivative of f by $\int f$. If f has an antiderivative $\int f$, then all other antiderivatives of f are of the form $\int f + c$, where $c \in \mathbb{R}$ is a constant.

Note, we have not shown yet that any function actually has an antiderivative. A corollary of the first part of the fundamental theorem of calculus is that every continuous function has an antiderivative.

Definition 6.6. (Difference notation). If $f : A \rightarrow \mathbb{R} \rightarrow \mathbb{R}$ is function defined on any nonempty set $A \subseteq \mathbb{R}$, then we define, for $a, b \in A$, the notation $f|_a^b := f(b) - f(a)$.

Theorem 6.7. (Fundamental theorem of calculus).

1. If $f : [a, b] \rightarrow \mathbb{R}$ is continuous, then the “accumulation function” $\int_a^x f$ is differentiable on $[a, b]$ and

$$\frac{d}{dx} \int_a^x f = f.$$

2. If $f : [a, b] \rightarrow \mathbb{R}$ is continuous, then

$$\int_a^b f = \left(\int f \right) \Big|_a^b.$$

Proof. See <http://www2.clarku.edu/~djoyce/ma121/FTCproof.pdf>. □

Remark 6.8. (Intuition for the fundamental theorem of calculus).

The first part of the theorem says that the rate of the accumulation f is f itself.

The second part of the fundamental theorem of calculus follows from the first part. It states that accumulating f over the region $[a, b]$ is “the same” as finding the difference in the antiderivative between b and a . More geometrically, if we define $F = \int f$, then it relates the definite integral of F' over the region $[a, b]$ to how F behaves on the boundary $\{\{a\}, \{b\}\}$. Stokes’ theorem (Theorem [...]) generalizes this result.

(The second part of the fundamental theorem as stated above also holds if f is assumed to be Riemann-integrable. The focus of this book is geometry, not analysis, so we are not concerned with this).

Theorem 6.9. All continuous functions $[a, b] \rightarrow \mathbb{R}$ have antiderivatives.

This is a corollary of the first part of the fundamental theorem of calculus: an antiderivative of a continuous function f is its accumulation function.

6.2 Multivariable calculus

Lemma 6.10. (Multivariable chain rule for differentiable functions $\mathbb{R}^n \rightarrow \mathbb{R}$).

Let $\mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}^n$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Set $\mathbf{x}(t_0) = \mathbf{x}_0$.

It can be proved that

$$\left. \frac{d(f \circ \mathbf{x})}{dt} \right|_{\mathbf{x}=\mathbf{x}_0} = \left. \frac{\partial f(\mathbf{x})}{\partial x_1} \right|_{\mathbf{x}=\mathbf{x}_0} \left. \frac{dx_1}{dt} \right|_{t=t_0} + \dots + \left. \frac{\partial f(\mathbf{x})}{\partial x_n} \right|_{\mathbf{x}=\mathbf{x}_0} \left. \frac{dx_n}{dt} \right|_{t=t_0}.$$

In other words,

$$\left. \frac{d(f \circ \mathbf{x})}{dt} \right|_{\mathbf{x}=\mathbf{x}_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \left. \frac{d\mathbf{x}(t)}{dt} \right|_{t=t_0},$$

where we have defined the *gradient of f with respect to the function \mathbf{x}* to be

$$\nabla_{\mathbf{x}} f := \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Note that since $\mathbf{x}(\mathbf{p}) = (x_1(\mathbf{p}), \dots, x_n(\mathbf{p}))^\top$, each x_i is a function. Thus, the derivative $\frac{\partial}{\partial x_i}$ is a derivative with respect to a function, in the sense of Definition 6.3. This is why we say the gradient $\nabla_{\mathbf{x}} f$ is “with respect to \mathbf{x} ”.

We can interpret the dot product to act on vector-valued functions (the dot product of vector-valued functions is equal to the dot product of the evaluated vector-valued functions at each point) so that the above is expressed as

$$\left. \frac{d(f \circ \mathbf{x})}{dt} \right|_{\mathbf{x}=\mathbf{x}_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \left. \frac{d\mathbf{x}(t)}{dt} \right|_{t=t_0}$$

Definition 6.11. (Directional derivative of a differentiable function $\mathbb{R}^n \rightarrow \mathbb{R}$).

Consider $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Let $\mathbf{r} : \mathbb{R} \rightarrow \mathbb{R}^n$ be the curve with $\mathbf{r}(t_0) = \mathbf{x}_0$ and $\left. \frac{d\mathbf{r}}{dt} \right|_{t=t_0} = \mathbf{v}$. We define the *directional derivative $\frac{\partial f}{\partial \mathbf{v}}$ of f in the direction of \mathbf{v}* to be

$$\left. \frac{\partial f}{\partial \mathbf{v}} \right|_{\mathbf{x}=\mathbf{x}_0} := \left. \frac{d(f \circ \mathbf{r})}{dt} \right|_{\mathbf{x}=\mathbf{x}_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \left. \frac{d\mathbf{r}(t)}{dt} \right|_{t=t_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \mathbf{v}$$

Therefore the directional derivative is expressed as

$$\boxed{\begin{aligned}\frac{\partial f}{\partial \mathbf{v}} \Big|_{\mathbf{x}=\mathbf{x}_0} &= (\nabla_{\mathbf{x}} f)_{\mathbf{x}=\mathbf{x}_0} \cdot \mathbf{v} \\ \frac{\partial f}{\partial \mathbf{v}} &= \nabla f \cdot \mathbf{v}\end{aligned}}$$

In the second line, we interpret ∇ as the function sending $\mathbf{x} \mapsto \nabla_{\mathbf{x}}$.

Most authors denote $\frac{\partial f}{\partial \mathbf{v}}$ as $D_{\mathbf{x}}f(\mathbf{v})$ or as $Df[\mathbf{v}](\mathbf{x})$.

Remark 6.12. [Lee, p.282, 283] (Different interpretations and notation for the directional derivative).

We have defined the directional derivative as the ratio

$$\frac{\text{change in } f \text{ that results from moving away from } \mathbf{x} = \mathbf{x}_0 \text{ with instantaneous velocity } \mathbf{v}}{\text{“infinitesimal” time interval over which this moving away occurs}}.$$

This interpretation of the directional derivative motivates why we write the directional derivative as $\frac{\partial f}{\partial \mathbf{v}}$. We will continue to favor this notation because it plays most nicely with the multivariable chain rule.

It should be noted, however, that the directional derivative can also be thought of as the “infinitesimal” change in f resulting from an “infinitesimally small velocity” \mathbf{v} . (So, to recap, the first interpretation relies on the concept of an “infinitesimal” time interval, while the second interpretation relies on the concept of an “infinitesimal” velocity). Thinking of the directional derivative in this way leads to notating it by $D_{\mathbf{p}}f(\mathbf{v})$, since this notation is reminiscent of the expression “ df ”. Taylor’s theorem justifies this interpretation as follows: if $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth function, then $(\Delta_{\mathbf{v}}f)|_{\mathbf{p}} = f(\mathbf{p} + \mathbf{v}) - f(\mathbf{p})$ is well approximated as $\Delta_{\mathbf{v}}f \approx \sum_{i=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} ([\mathbf{v}]_{\hat{\mathbf{e}}})_i = (\nabla_{\mathbf{x}} f) \cdot \mathbf{v} = (D_{\mathbf{p}}f)(\mathbf{v})$ when $\|\mathbf{v}\|$ is small.

Theorem 6.13. (Directional derivative as a limit).

It is sometimes useful to express the directional derivative as a limit.

To do so, first consider the curve \mathbf{r} from the definition of the directional derivative. Since $\mathbf{r}(t_0) = \mathbf{x}_0$ and $\frac{d\mathbf{r}}{dt} \Big|_{t=t_0} = \mathbf{v}$, then $\mathbf{r}(t) = \mathbf{x}_0 + \mathbf{v}t$. Thus

$$\begin{aligned}\frac{\partial f}{\partial \mathbf{v}} \Big|_{\mathbf{x}=\mathbf{x}_0} &= \frac{d(\mathbf{f} \circ \mathbf{r})(t)}{dt} \Big|_{t=t_0} = \frac{d\mathbf{f}(\mathbf{x}_0 + \mathbf{v}t)}{dt} \Big|_{t=t_0} = \lim_{h \rightarrow 0} \left(\frac{\mathbf{f}(\mathbf{x}_0 + \mathbf{v}(t+h)) - \mathbf{f}(\mathbf{x}_0 + \mathbf{v}t)}{h} \right) \Big|_{t=t_0} \\ &= \lim_{h \rightarrow 0} \left(\left(\frac{\mathbf{f}(\mathbf{x}_0 + \mathbf{v}(t+h)) - \mathbf{f}(\mathbf{x}_0 + \mathbf{v}t)}{h} \right) \Big|_{t=t_0} \right) = \lim_{h \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + h\mathbf{v}) - \mathbf{f}(\mathbf{v})}{h}.\end{aligned}$$

Therefore

$$\frac{\partial f}{\partial \mathbf{v}} \Big|_{\mathbf{x}=\mathbf{x}_0} = \lim_{h \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + h\mathbf{v}) - \mathbf{f}(\mathbf{v})}{h}.$$

Many authors define the directional derivative using this formula.

Theorem 6.14. (Gradient is direction of greatest increase).

Remark 6.15. (Directional derivative simplifies to partial derivative). We have $\frac{\partial}{\partial \hat{\mathbf{e}}_i} = \frac{\partial}{\partial x_i}$.

Lemma 6.16. (Multivariable chain rule for differentiable functions $\mathbb{R}^n \rightarrow \mathbb{R}^m$).

Let $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}^n$ be [...sufficiently differentiable...], and set $\mathbf{x}_0 = \mathbf{x}(t_0)$.

$$\frac{d(\mathbf{f} \circ \mathbf{x})(t)}{dt} \Big|_{\mathbf{x}=\mathbf{x}_0} = \begin{pmatrix} \frac{d}{dt} f_1(\mathbf{x}(t)) \\ \vdots \\ \frac{d}{dt} f_m(\mathbf{x}(t)) \end{pmatrix} \Big|_{\mathbf{x}=\mathbf{x}_0} = \begin{pmatrix} (\nabla_{\mathbf{x}} f_1)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}}{dt} \Big|_{t=t_0} \\ \vdots \\ (\nabla_{\mathbf{x}} f_m)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}}{dt} \Big|_{t=t_0} \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \end{pmatrix} \Big|_{\mathbf{x}=\mathbf{x}_0} \frac{d\mathbf{x}}{dt} \Big|_{t=t_0}.$$

In terms of functions, we have

$$\frac{d(\mathbf{f} \circ \mathbf{x})(t)}{dt} = \begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \end{pmatrix} \frac{d\mathbf{x}}{dt}$$

Recall from Derivation 2.51 and Theorem 2.61 that a matrix-vector product can be expressed as either a linear combination of column vectors or as a vector of dot products. We have already seen the second expression; here is the first:

$$\begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \end{pmatrix} \frac{d\mathbf{x}}{dt} = \begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \cdot \frac{d\mathbf{x}}{dt} \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \cdot \frac{d\mathbf{x}}{dt} \end{pmatrix} = \sum_{i=1}^n \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_i} \frac{dx_i(t)}{dt}.$$

Definition 6.17. (The Jacobian).

$$\text{Let } \mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix}.$$

Drawing upon the idea of the derivative of a function with respect to a function (see Definition 6.3), we define the *Jacobian matrix* $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ to be

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} := \begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \end{pmatrix} = \left(\frac{\partial f_i}{\partial x_j} \right)$$

Using the Jacobian, the multivariable chain rule for differentiable functions $\mathbb{R}^n \rightarrow \mathbb{R}^m$ is now succinctly stated as

$$\frac{d(\mathbf{f} \circ \mathbf{x})}{dt} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt}$$

Definition 6.18. (Directional derivative of a differentiable function $\mathbb{R}^n \rightarrow \mathbb{R}^m$).

The directional derivative of a differentiable function $\mathbf{f} : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ is defined analogously to that of a differentiable function $\mathbf{f} : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$. Indeed, in the special case of $m = 1$, the two definitions are equivalent.

As was done previously, let $\mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}^n$ be the curve with $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\frac{d\mathbf{x}}{dt}\big|_{t=t_0} = \mathbf{v}$. We define the *directional derivative* $\frac{\partial \mathbf{f}}{\partial \mathbf{v}}$ of \mathbf{f} in the direction of \mathbf{v} to be

$$\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\big|_{\mathbf{x}=\mathbf{x}_0} := \frac{d(\mathbf{f} \circ \mathbf{x})}{dt}\big|_{\mathbf{x}=\mathbf{x}_0} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\big|_{\mathbf{x}=\mathbf{x}_0} \frac{d\mathbf{x}}{dt}\big|_{t=t_0}$$

So this most general definition of directional derivative is expressed as

$$\begin{aligned} \frac{\partial \mathbf{f}}{\partial \mathbf{v}}\big|_{\mathbf{x}=\mathbf{x}_0} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\big|_{\mathbf{x}=\mathbf{x}_0} \frac{d\mathbf{x}}{dt}\big|_{t=t_0} \\ \frac{\partial \mathbf{f}}{\partial \mathbf{v}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt} \end{aligned}$$

Remark 6.19. Linearity with respect to the \mathbf{v} in the denominator.

6.3 Technical theorems

Theorem 6.20. (The integral is “linear” with respect to the region of integration).

If $D = \cup_{i=1}^k D_i$, where D_1, \dots, D_k are domains of integration such that $D_i \cap D_j$ has measure zero¹ for all $i \neq j$, then

$$\int_D f = \sum_{i=1}^k \int_{D_i} f.$$

¹Informally, a subset of \mathbb{R}^n has *measure zero* iff its volume is zero.

7

Basic topology

Topology in \mathbb{R}

In this chapter we present the basic concepts of *topology*. Topology is useful to us because it generalizes ideas from calculus such as *convergence* and *continuity* to an abstract setting where no concept of distance is required; instead of “distance”, topology relies on the concept of what are called *open sets*. Primarily, we require topology in order to use the topological concept of an *n-manifold*, which is a *topological space* that “looks like \mathbb{R}^n ”.

We just said that topology generalizes ideas of calculus. More accurately, topology generalizes the ideas of *real analysis*, which is the technical framework which calculus relies upon. As the name implies, real analysis is essentially the study of the real numbers \mathbb{R} . As the is not assumed to have a real analysis background or a topology background, we will introduce real analysis concepts and their generalizations in topology simultaneously. In my opinion, it is actually best to learn the introductory concepts of real analysis and topology at the same time, as the general apparatus of topology, not being bogged down by the nitty-gritty details of real analysis, helps one more quickly realize how concepts interact.

Open sets, closed sets and their characterizations

We begin our investigation of the basic concepts of topology- *open sets* and *closed sets*- by starting in the context of the real numbers \mathbb{R} .

Definition 7.1. (Open set in \mathbb{R}).

An *open set* in \mathbb{R} is an arbitrary union of open intervals in \mathbb{R} .

Recall, an open interval in \mathbb{R} is a set of the form $(a, b) = \{x \mid a < x < b, a, b \in \mathbb{R}\}$. By “arbitrary union”, we mean that there is no restriction on the cardinality of the sets in the union; sets in the union can be either finite, countably infinite, or uncountably infinite.

Definition 7.2. (Interior points and interior of $A \subseteq \mathbb{R}$).

Let $A \subseteq \mathbb{R}$ be any subset of \mathbb{R} . A point $x \in A$ is called an *interior point* of A iff there exists an open set contained in A that contains x . Symbolically, $x \in A$ is an interior point iff $\exists \text{open } U_x \ni x \text{ s.t. } U_x \subseteq A$.

We define the *interior* of A to be the set of all interior points of A ,

$$\text{int}(A) := \{x \in A \mid x \text{ is an interior point of } A\}.$$

Theorem 7.3. (Interior point characterization of open sets in \mathbb{R}).

Let $A \subseteq \mathbb{R}$ be any subset of \mathbb{R} . Then A is open iff each $x \in A$ is an interior point. Symbolically, A is open iff $\forall x \in A \exists \text{open } U_x \ni x \text{ s.t. } U_x \subseteq A$.

Equivalently, A is open iff $A = \text{int}(A)$. (Every set A satisfies $\text{int}(A) \subseteq A$. So $A = \text{int}(A)$ iff $A \subseteq \text{int}(A)$).

Proof. We show that $(A \text{ is open}) \iff (\text{each } x \in A \text{ is an interior point})$.

(\implies) . If A is open, then every $x \in A$ is contained in an open set $U_x \subseteq A$, namely, $U_x = A$ for all x .

(\impliedby) . If every $x \in A$ is contained in an open set $U_x \subseteq A$, then $A = \cup_{x \in A} U_x$. (We have $\cup_{x \in A} U_x \subseteq A$ since each $U_x \subseteq A$, and $A \subseteq \cup_{x \in A} U_x$ since each $x \in A$ is in U_x). We defined an open set in \mathbb{R} to be an arbitrary union of open sets in \mathbb{R} , so A is open. \square

Next, see what the interior point characterization of open sets implies for complements of open sets.

Derivation 7.4. (Limit points in \mathbb{R}).

Each of the lines in the following derivation are logically equivalent. Skim this derivation, but don't try to understand it on the first pass- some new notation that is used in the derivation is defined after the derivation!

$$\begin{aligned}
& U \text{ is open} \\
U = \text{int}(U) & \iff U \subseteq \text{int}(U) \iff \forall x \in U \ x \in \text{int}(U) \\
& \forall x \in U \ \exists \text{open } U_x \ni x \text{ s.t. } U_x \subseteq U \\
& \forall x \in U \ \exists \text{open } U_x \ni x \text{ s.t. } U_x \cap U^c = \emptyset \\
& \forall x \ x \notin U^c \implies \text{not}(\forall \text{open } U_x \ni x \ U_x \cap U^c \neq \emptyset) \\
\forall x \ x \notin U^c & \implies \text{not}(\forall \text{open } U_x \ni x \ U_x \cap U^c - \{x\} \neq \emptyset) \\
& \forall x \ x \notin U^c \implies x \notin (U^c)' \\
& \forall x \ x \in (U^c)' \implies x \in U^c \\
& (U^c)' \subseteq U^c.
\end{aligned}$$

Note that line 6 follows from line 5 because we can subtract x out of $U_x \cap U^c$ due to the hypothesis " $x \notin U^c$ ". Line 6 implies line 5 for the same reason.

In line 7, we define the notion of a limit point of a subset of \mathbb{R} . We say $x \in \mathbb{R}$ is a *limit point* of $A \subseteq \mathbb{R}$ iff $\forall \text{open } U_x \ U_x \cap U^c - \{x\} \neq \emptyset$. That is, $x \in \mathbb{R}$ is a *limit point* of $A \subseteq \mathbb{R}$ iff every open set containing x intersects A at a point other than x . The set of limit points of A is denoted A' .

The above shows that U is open iff $(U^c)' \subseteq U^c$. In words, a set is open iff its complement contains all of its limit points. For this reason, we define a *closed set* in \mathbb{R} to be any set which is the complement of an open set, or, equivalently, any set which contains all of its limit points. We call the fact that C is closed iff $C' \subseteq C$ the *limit point characterization of closed sets*. We repeat these definitions below.

Definition 7.5. (Limit point in \mathbb{R}).

Let $A \subseteq \mathbb{R}$ be any subset of \mathbb{R} . We say $x \in \mathbb{R}$ is a *limit point* of $A \subseteq \mathbb{R}$ iff every open set containing x intersects A at a point other than x . Symbolically, $x \in \mathbb{R}$ is a *limit point* of $A \subseteq \mathbb{R}$ iff $\forall \text{open } U_x \ U_x \cap U^c - \{x\} \neq \emptyset$.

The set of limit points of A is denoted A' .

Definition 7.6. (Closed set in \mathbb{R}).

A subset $A \subseteq \mathbb{R}$ of \mathbb{R} is said to be *closed* iff A is the complement of some open subset of \mathbb{R} . Equivalently, $A \subseteq \mathbb{R}$ is closed iff A contains all of its limit points.

Remark 7.7. (Open vs. closed).

Subsets of \mathbb{R} can be open but not closed, closed but not open, both and closed, or neither open nor closed. (This is true for general topological spaces as well).

Theorem 7.8. (Limit point characterization of closed sets in \mathbb{R}).

A subset $A \subseteq \mathbb{R}$ is closed iff A contains all of its limit points.

Remark 7.9. (Characterizations of open sets and closed sets in \mathbb{R}).

We have seen that open sets in \mathbb{R} are characterized by the "interior point characterization of open sets", while closed sets in \mathbb{R} are characterized by the "limit point characterization of closed sets".

Remark 7.10. put this somewhere else

Still looking at the above proof, we can notice that we accidentally proved $x \in \text{int}(U)$ iff $x \notin (U^c)'$. So, we proved $x \in \text{int}(U)$ iff $x \in ((U^c)')^c$, which means $\text{int}(U) = ((U^c)')^c \iff \text{int}(U)^c = (U^c)'$. The one to remember is $\text{int}(U)^c = (U^c)'$.

7.1 Topological spaces

To discover the general apparatus of topology, we investigate unions and intersections of open and closed sets in \mathbb{R} . It quickly follows from the definition of an open set as an arbitrary union of open intervals that an arbitrary union of open sets is an open set. DeMorgan's laws (recall [...]) then imply that an arbitrary intersection of closed sets is a closed set.

What about intersections of open sets- or, equivalently, by DeMorgan's laws- unions of closed sets in \mathbb{R} ? Well, any infinite union of closed sets is not necessarily closed: consider $\cup_{i=1}^{\infty} \{\frac{1}{n}\} = \{\frac{1}{n} \mid n \in \mathbb{N}\}$, which does not contain its limit point of 0. Perhaps at least finite unions of closed sets must be closed? This is indeed the case for closed subsets of \mathbb{R} .

The following definition of a topological space is motivated by these properties of unions and intersections of open and closed sets in \mathbb{R} . Before this definition, we need some more technical language.

Definition 7.11. (Covers and generating covers).

Let X be any set. A set \mathcal{C} of subsets of X is a *cover* of X iff $\cup_{C \in \mathcal{C}} C = X$.

Let τ be any set. A set \mathcal{C} *generates* τ iff each $U \in \tau$ is an arbitrary union of the elements of \mathcal{C} ; that is, iff each $U \in \tau$ is $U = \cup_{\alpha \in I} C_{\alpha}$, $\{C_{\alpha}\} \subseteq \mathcal{C}$.

We can now define topological spaces.

Definition 7.12. (Topological space).

Let (X, τ) be a tuple, where X is any set and τ is a set of subsets of X . We say that (X, τ) is a *topological space*, and that τ is a *topology* on X , iff...

1. There is a cover \mathcal{B} of X that generates τ .
 - For the same reasons as earlier, when we had $X = \mathbb{R}$ and $\tau = \{\text{open sets} \subseteq \mathbb{R}\}$, (1) is equivalent to the interior point characterization of open sets, which is in turn equivalent to the limit point characterization of closed sets. Interior points and limit points will be defined analogously as to how they were before.
 - This requirement is equivalent to: arbitrary unions of open sets are open \iff arbitrary intersections of closed sets are closed.
2. Finite unions of closed sets are closed \iff finite intersections of open sets are open.

We interpret the elements of τ as being open sets. Formally, we say that $U \subseteq X$ is an *open set* iff $U \in \tau$.

Remark 7.13. (\mathbb{R}^n with the standard topology).

In the section "Topology on \mathbb{R} ", we were really doing topology on (\mathbb{R}, std) , where *std* is the *standard topology on \mathbb{R}* , $\text{std} := \{\text{subsets of } \mathbb{R} \text{ that are arbitrary unions of open intervals in } \mathbb{R}\}$. In general, the *standard topology on \mathbb{R}^n* is $\{\text{subsets of } \mathbb{R}^n \text{ that are arbitrary unions of open intervals in } \mathbb{R}^n\}$.

Definition 7.14. (Topological space).

The above definition of a topology can quickly be seen to be equivalent to the following most common definition, which makes no mention of a generating cover. The most common definition requires τ to satisfy the following:

3. Arbitrary unions of open sets are open.
 - As noted in (1), we have that (1) and (3) are equivalent.

4. Finite intersections of open sets are open.

Remark 7.15. $(\emptyset, X \in \tau)$.

Most definitions of a topological space explicitly require that the empty set and X are open subsets of X . This requirement is implied by the above definition. X is generated by τ by definition. It's less easy to see that \emptyset is also generated by τ , but this is true too. The empty set is the empty union (which employs an empty indexing set) of any collection of open sets.

Bases

Our first definition of a topological space (X, τ) involved a cover \mathcal{B} that generated X . There was also another condition- we required finite intersections of open sets to be open. We would like to impose additional requirements on \mathcal{B} so that a topological space (X, τ) can be said to be generated by \mathcal{B} , without any other requirements on (X, τ) being necessary. In other words, we want to find requirements on \mathcal{B} such that finite intersections of open sets are guaranteed to be open just by the virtue of the requirements on \mathcal{B} . We would also like for these requirements on \mathcal{B} to be not just sufficient for this finite intersection property, but also necessary. The following theorem gives us the answer we want and need.

Theorem 7.16. (Finite intersections of open sets are open iff the cover “refines with interior points”).

If \mathcal{B} is a cover that generates τ , then finite intersections of open sets are open iff \mathcal{B} *refines with interior points*. We say that \mathcal{B} *refines with interior points*, that is, iff $\forall B_1, B_2 \in \mathcal{B} \quad x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B} \text{ s.t. } x \in B_3$.

Proof.

(\implies). Assume a finite union of closed sets is closed. Then by DeMorgan's laws, any finite intersection of open sets is open. We must show that \mathcal{B} refines with interior points. Basis elements are by definition open, so if $B_1, B_2 \in \mathcal{B}$, then $B_1 \cap B_2$ is a finite intersection of open sets, and is thus open. By the interior point characterization of open sets, $x \in B_1 \cap B_2$ implies there is some open set $U \ni x$ such that $U \subseteq B_1 \cap B_2$. Because U is open, then it is a union of basis elements, so $x \in B_3$ for some $B_3 \in \mathcal{B}$.

(\impliedby). Assume \mathcal{B} is a basis; that is, \mathcal{B} is a cover that refines with interior points. We want to show that finite unions of closed sets are closed. By DeMorgan's laws, we can instead show that finite intersections of open sets are open.

So, set $V = \bigcap_{i=1}^n U_i$, where the U_i are open. If any U_i is empty, then their intersection is \emptyset , which is open, so assume no U_i is empty. We show that V is open by showing it satisfies the interior point characterization of open sets. Consider $x \in V$. Then $x \in U_i$ for all i . Each U_i is a union of basis elements, so, for each U_i , we have $x \in B_i$ for some $B_i \in \mathcal{B}$. Thus $x \in \bigcap_{i=1}^n B_i$. Using induction on the fact that \mathcal{B} refines with interior points, there is a $B_x \in \mathcal{B}$ s.t. $x \in B_x \subseteq \bigcap_{i=1}^n B_i$. We have $\bigcap_{i=1}^n B_i = \bigcup_{x \in V} B_x$, so $\bigcap_{i=1}^n B_i$ is open. \square

Definition 7.17. (Basis for a topological space).

A cover \mathcal{B} for X is called a *basis* iff \mathcal{B} refines with interior points, that is, iff $\forall B_1, B_2 \in \mathcal{B} \quad x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B} \text{ s.t. } x \in B_3$.

Theorem 7.18. Every basis generates a topological space, and every topological space has a basis.

Proof. For the first part of the theorem, recall that the whole point of defining a basis was to find the types of generating covers that generate topological spaces.

For the second part, note that if (X, τ) is a topological space, then τ is a basis for X . \square

Basic facts about topological spaces

We now state the generalizations of the topological results we found in the context of \mathbb{R} . We do not state any proofs in this section, as the arguments given in the “Topology in \mathbb{R} ” section all generalize easily.

Definition 7.19. (Open set).

Let (X, τ) be a topological space. As was mentioned previously, an *open set* in X is an element of τ . We often refer to open sets in X as simply “open sets”.

Open sets are also often called *neighborhoods*. A *neighborhood of a point* $x \in X$ is an open set which contains x .

Definition 7.20. (Interior points and interior).

Let (X, τ) be a topological space, and $A \subseteq X$. A point $x \in A$ is called an *interior point of A* iff there exists an open set contained in A that contains x . Symbolically, $x \in A$ is an interior point iff $\exists \text{open } U_x \ni x \text{ s.t. } U_x \subseteq A$.

We define the *interior* of A to be the set of all interior points of A ,

$$\text{int}(A) := \{x \in A \mid x \text{ is an interior point of } A\}.$$

Definition 7.21. (Interior point characterization of open sets).

Let (X, τ) be a topological space, and let $A \subseteq X$. Then A is open iff each $x \in A$ is an interior point. Symbolically, A is open iff $\forall x \in A \exists \text{open } U_x \ni x \text{ s.t. } U_x \subseteq A$.

Equivalently, A is open iff $A = \text{int}(A)$. (Every set A satisfies $\text{int}(A) \subseteq A$. So $A = \text{int}(A)$ iff $A \subseteq \text{int}(A)$).

Definition 7.22. (Limit point).

Let (X, τ) be a topological space, and let $A \subseteq X$. We say $x \in A$ is a *limit point of A* iff every open set containing x intersects A at a point other than x . Symbolically, $x \in \mathbb{R}$ is a *limit point of* $A \subseteq \mathbb{R}$ iff $\forall \text{open } U_x \ U_x \cap U^c - \{x\} \neq \emptyset$.

The set of limit points of A is denoted A' .

Definition 7.23. (Closed set).

Let (X, τ) be a topological space. A subset $A \subseteq X$ is said to be *closed* iff A is the complement of some open set. Equivalently, $A \subseteq X$ is closed iff A contains all of its limit points.

Theorem 7.24. (Limit point characterization of closed sets).

Let (X, τ) be a topological space. A subset $A \subseteq X$ is closed iff A contains all of its limit points.

Hausdorff spaces

Definition 7.25. (Hausdorff space).

Theorem 7.26. (In a Hausdorff space, compact implies closed).

Metric spaces

Definition 7.27. (Metric space).

Define $B(\epsilon, x)$

Theorem 7.28. (Every metric space is Hausdorff).

The subspace topology

The following definition and theorem justify what we would might naturally assume the phrase “open set in $Y \subseteq X$ ” means.

Definition 7.29. (Subspace topology).

Let (X, τ_X) be a topological space, and let $Y \subseteq X$. The *subspace topology* τ_Y is defined to be the topology on Y whose open sets are $\tau_Y := \{U \cap Y \mid U \text{ is open in } X\}$.

If $A \subseteq Y$, then we say that A is *open in Y* iff $A \in \tau_Y$, and we say that A is *closed in Y* iff $X - A$ is open in Y .

Theorem 7.30. A set A is closed in Y iff $A = B \cap Y$, where B is closed in X .

Proof.

(\implies). Suppose A is closed in Y . Then $Y - A$ is open in Y , so $Y - A = U \cap Y$ for some U open in X . Thus $A = Y - (Y - A) = Y - (U \cap Y) = Y - U = Y \cap (X - U)$. So, $A = (X - U) \cap Y$, where $X - U$ is closed in X .

(\impliedby). Suppose $A = B \cap Y$, where B is closed in X . Now, we show that $Y - A$ is open in Y . We have $Y - A = Y - (B \cap Y) = Y \cap (X - B)$. Thus $Y - A = (X - B) \cap Y$, where $X - B$ is open in X , meaning $(X - B) \cap Y$ is open in Y . \square

Interior, closure, and boundary

For the purposes of differential forms, understanding this section in depth is not necessary.

[closure and continuity?]

Just try to get a general sense of what “boundary” is.

Theorem 7.31. (Equivalent definitions of interior).

Let (X, τ) be a topological space. The *interior* of $A \subseteq X$ is

- $\text{int}(A) := \{x \in A \mid x \text{ is an interior point of } A\}$, by definition.
- The “largest” open set contained in A ; that is, the union of all open sets contained in A .

Proof. We need to show

$$\{x \in A \mid x \text{ is an interior point of } A\} = \bigcup_{U \subseteq X, U \in \tau} U.$$

The proof of this is quick: $x \in A$ is an interior point of A iff there exists an open set $U_x \subseteq A$ containing x iff x is in the union of open sets contained in A . \square

Definition 7.32. (Closure).

Let (X, τ) be a topological space. The *closure* $\text{cl}(A)$ of $A \subseteq X$ is the “smallest” closed set that contains A ; that is, it is the intersection of all closed sets containing A .

Theorem 7.33. (Condition for being in the closure).

Let (X, τ) be a topological space, and let $A \subseteq X$. We have

$$\begin{aligned} x \in \text{cl}(A) &\iff \forall \text{closed } C \supseteq A \ x \in C \\ &\iff \forall \text{open } U \ U \cap A = \emptyset \implies x \notin U \\ &\iff \forall \text{open } U \ni x \ U \cap A \neq \emptyset. \end{aligned}$$

In the second line, we’ve used $U = C^c$. The third line follows from the second because $(P \text{ and not } Q) \iff (Q \implies P)$.

In all, we have

$$x \in \text{cl}(A) \iff \forall \text{open } U \ni x \ U \cap A \neq \emptyset.$$

Remark 7.34. (Being in the closure vs. being a limit point).

Notice that the condition for being in the closure is just a little bit weaker than the condition for being a limit point. Recall, $x \in A$ is a limit point of A iff $\forall \text{open } U \ni x \ U \cap A - \{x\} \neq \emptyset$.

Definition 7.35. (Isolated point).

Isolated points are $A - A'$?

Theorem 7.36. $\text{cl}(A) = A \cup A'$ so $\text{cl}(A) = (A - A') \sqcup A'$

Definition 7.37. (Boundary).

Let (X, τ) be a topological space. The *boundary* ∂A of $A \subseteq X$ is defined to be $\partial A := \text{cl}(A) - \text{int}(A)$.

7.2 Continuous functions, homeomorphisms, and compactness

Limits and continuity for functions $A \subseteq \mathbb{R} \rightarrow \mathbb{R}$

Definition 7.38. (Limit of a function $A \subseteq \mathbb{R} \rightarrow \mathbb{R}$).

Let $A \subseteq \mathbb{R}$, and consider a function $f : A \subseteq \mathbb{R} \rightarrow \mathbb{R}$. We write $\lim_{x \rightarrow x_0} f(x) = L$ iff $\forall \epsilon > 0 \exists \delta > 0$ s.t. $x \in B(\delta, x_0) - \{x_0\} \implies f(x) \in B(\epsilon, L)$.

Definition 7.39. (Continuity of a function $A \subseteq \mathbb{R} \rightarrow \mathbb{R}$ at a point).

Let $A \subseteq \mathbb{R}$, and consider a function $f : A \subseteq \mathbb{R} \rightarrow \mathbb{R}$. We say that f is continuous at $x_0 \in A$ iff $\forall \epsilon > 0 \exists \delta > 0$ s.t. $x \in B(\delta, x_0) \implies f(x) \in B(\epsilon, f(x_0))$

Theorem 7.40. (Condition for continuity for functions $A \subseteq \mathbb{R} \rightarrow \mathbb{R}$).

Let $A \subseteq \mathbb{R}$. A function $f : A \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is continuous at $x_0 \in A$ iff $\lim_{x \rightarrow x_0} f(x) = f(x_0)$.

Proof.

(\implies). If f is continuous at x_0 , then we immediately know $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ because $x \in B(\delta, x_0)$ is a weaker condition than $x \in B(\delta, x_0) - \{x_0\}$.

(\impliedby). Suppose that $\forall \epsilon > 0 \exists \delta > 0$ s.t. $x \in B(\delta, x_0) \cap A - \{x_0\} \implies f(x) \in B(\epsilon, f(x_0))$. We need to show that when $x = x_0$ we have $f(x) \in B(\epsilon, f(x_0))$. But this follows immediately because $f(x_0) \in B(\epsilon, f(x_0))$ for any ϵ , as $f(x_0) = f(x_0)$. □

Limits and continuity for functions on subsets of topological spaces

Definition 7.41. (Limit of a function on a subset of a topological space).

Let X and Y be topological spaces, let $A \subseteq X$, and consider a function $f : A \subseteq X \rightarrow Y$. We write $\lim_{x \rightarrow x_0} f(x) = L$ iff $\forall \text{open } V \ni L, V \subseteq Y \exists \text{open } U \ni x_0, U \subseteq X$ s.t. $f(U - \{x_0\}) \subseteq V$.

Definition 7.42. (Continuity of a function on a subset of a topological space).

Let X and Y be topological spaces, let $A \subseteq X$, and consider a function $f : A \subseteq X \rightarrow Y$. We say that f is continuous at $x_0 \in A$ iff $\forall \text{open } V \ni f(x_0), V \subseteq Y \exists \text{open } U \ni x_0, U \subseteq X$ s.t. $f(U) \subseteq V$.

Remark 7.43. (Topological limits and continuity generalize real analytical notions).

Note, the earlier definitions of the limit of a function $A \subseteq \mathbb{R} \rightarrow \mathbb{R}$ and continuity at a point for a function $A \subseteq \mathbb{R} \rightarrow \mathbb{R}$ can both be viewed as a special case of the just-stated corresponding definitions topological spaces. Specifically, these earlier definitions are the special case in which $(X, \tau_X) = (Y, \tau_Y) = (\mathbb{R}, \text{std})$, each V is $V = B(\epsilon, L)$, and each U is $U = B(\delta, x_0)$.

Theorem 7.44. (Condition for continuity for functions on subsets of topological spaces).

Let X and Y be topological spaces, and let $A \subseteq X$. A function $f : A \subseteq X \rightarrow Y$ is continuous at $x_0 \in A$ iff $\lim_{x \rightarrow x_0} f(x) = f(x_0)$.

Proof.

(\implies). If f is continuous at x_0 , then we immediately know $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ because $f(U) \subseteq V$ is a weaker condition than $f(U - \{x_0\}) \subseteq V$.

(\impliedby). Suppose that $\forall \text{open } V \ni f(x_0), V \subseteq Y \exists \text{open } U \ni x_0, U \subseteq X$ s.t. $f(U - \{x_0\}) \subseteq V$. We need to show that when $x = x_0$ we have $f(x) \in V$ for any open $V \ni x_0$. But this is true by definition of V ; we have $f(x_0) \in V$. □

We now present the classic topological interpretation of a continuous function.

Definition 7.45. (Continuous function).

Let X and Y be topological spaces, and let $A \subseteq X$. A function $f : A \subseteq X \rightarrow Y$ is called *continuous* iff it is continuous at every $x \in A$.

Theorem 7.46. (The inverse image of a continuous function preserves openness and closedness).

Let X and Y be topological spaces, and let $A \subseteq X$. A function $f : A \subseteq X \rightarrow Y$ is continuous iff the following equivalent conditions hold:

- For all open sets $V \subseteq Y$, the subset $f^{-1}(V) \subseteq X$ is open in X .
- For all closed sets $D \subseteq Y$, the subset $f^{-1}(D) \subseteq X$ is closed in X .

Proof. (Continuous \iff 1).

(1 \iff 2).

□

Homeomorphisms

Definition 7.47. (Homeomorphism).

Compactness

Definition 7.48. (Compact topological space).

Definition 7.49. (Compactness of a topological subspace).

Definition 7.50. (Compactness in \mathbb{R}^n with the standard topology).
closed and bounded

Part III

Differential forms

8

Manifolds

8.1 Introduction to manifolds

Definition 8.1. [Lee, p.2] (Manifold). A n -manifold is a topological space M that is...

- Hausdorff, or “point-separable”
- *second-countable*; that is, M has a countable basis
- *locally Euclidean of dimension n* in the sense that each point in M has a neighborhood that is homeomorphic to \mathbb{R}^n (where \mathbb{R}^n has the standard topology)

Theorem 8.2. [Lee, p.3] (Topological invariance of dimension).

A nonempty n -manifold is homeomorphic to an m -dimensional manifold if and only if $n = m$. (See [Lee] for proof).

Definition 8.3. [Lee, p.25] (Closed n -dimensional upper half-space). Consider \mathbb{R}^n with the standard topology. We define the *closed n -dimensional half space* to be the topology $\mathbb{H}^n \subseteq \mathbb{R}^n$,

$$\mathbb{H}^n := \{(x^1, \dots, x^n) \in \mathbb{R}^n \mid x^n \geq 0\},$$

where \mathbb{H}^n has the subspace topology inherited from the standard topology of \mathbb{R}^n .

The point of defining \mathbb{H}^n is to allow for a distinction between what will be called *interior points of M* and *boundary points of M* . To see how involving \mathbb{H}^n facilitates this, note that the interior $\text{int}(\mathbb{H}^n)$ and boundary $\partial\mathbb{H}^n$ of \mathbb{H}^n , in the usual topological senses of “interior” and “boundary” (see Defn [...]), are

$$\begin{aligned}\text{int}(\mathbb{H}^n) &= \{(x^1, \dots, x^n) \in \mathbb{R}^n \mid x^n > 0\} \\ \partial\mathbb{H}^n &= \{(x^1, \dots, 0) \in \mathbb{R}^n\}.\end{aligned}$$

Recall [from topology] that $\mathbf{p} \in \mathbb{H}^n$ is an *interior point of \mathbb{H}^n* iff $\mathbf{p} \in \text{int}(\mathbb{H}^n)$, and $\mathbf{p} \in \mathbb{H}^n$ is a *boundary point of \mathbb{H}^n* iff $\mathbf{p} \in \partial\mathbb{H}^n$.

Now let M be an n -dimensional Hausdorff second-countable space, and suppose that for some open subset $U \subseteq M$ containing a point $\mathbf{p} \in U$, there is a homeomorphism $\mathbf{x} : U \subseteq M \rightarrow \mathbb{R}^n$, where $\mathbf{x}(\mathbf{p}) = (x^1(\mathbf{p}), \dots, x^{n-1}(\mathbf{p}), 0)$ is in $\partial\mathbb{H}^n$. Then, in analogy to the notion of “locally Euclidean” introduced in the definition of a manifold, we can say that the open subset $U \subseteq M$ “looks like a piece of the boundary of \mathbb{H}^n ”, or that “ M locally (near \mathbf{p}) looks like a piece of the boundary of \mathbb{H}^n ”. This motivates the following definition.

Definition 8.4. [Lee, p.25] (Manifold with boundary). An n -manifold with boundary is a topological space M that is...

- Hausdorff, or “point-separable”.
- *second-countable*; that is, M has a countable basis.
- *locally Euclidean of dimension n* in the sense that each point in M has a neighborhood that is homeomorphic to \mathbb{R}^n (where \mathbb{R}^n has the standard topology).
- such that “ M has a (possibly empty) manifold boundary”. That is, each point of M has a neighborhood that is either homeomorphic to an open subset of \mathbb{R}^n (where \mathbb{R}^n has the standard topology) or to an open subset \mathbb{H}^n with the subspace topology (inherited from the standard topology on \mathbb{R}^n).

Remark 8.5. [Lee, p.26] (Topological interior and boundary vs. manifold interior and boundary).

Let M be an n -manifold with boundary. We can obtain the *topological* interior and *topological* boundary of M by regarding M as a topological space and taking $\text{int}(M), \partial M$ in the usual topological senses of interior and boundary (see Definition [...] and Definition [...]). In general, the topological interior and topological boundary are *not* the same as the manifold interior and manifold boundary.

To see this, we first need to remember that the topological notions of interior and boundary are only applicable when M is a subset of some other topological space X . This is illustrated in the cases $X = M$ and $X = \mathbb{R}^m$, $m \geq n$.

- When $X = M$, the topological interior of M in the subspace topology of $X = M$ is M , and the topological boundary of M in the same subspace topology is $\text{cl}(M) - \text{int}(M) = M - M = \emptyset$. (int and cl denote topological interior and closure here). These facts conflict with the manifold interior and boundary of M : the manifold interior cannot be all of M when the manifold boundary of M is nonempty, and the manifold boundary is obviously not necessarily empty.
- When $X = \mathbb{R}^m$, then because $M \subseteq X$, we have $M = \mathbb{R}^n$ where $n \leq m$. When $n = m$, then the topological and manifold notions of interior and boundary are equivalent, but when $n < m$, the topological interior **is always empty?**.

“So for \mathbb{H}^n as a subspace of \mathbb{R}^n , there it is true that the topological boundary is the manifold boundary. And that’s the model for all manifolds with boundary. But because of global topology, it is often not possible for those local \mathbb{H}^n s to live in a global ambient \mathbb{R}^n , and without that we no longer have topological boundary = manifold boundary. (Consider a cylinder $\mathbb{I} \times S^1$, it’s not a subspace of \mathbb{R}^2)”

Definition 8.6. [Lee, p.25] (Manifold interior and boundary).

The (*manifold*) *interior* of M is the set of interior points in M , and is denoted $\text{int}(M)$. The (*manifold*) *boundary* of M is set of all boundary points in M , and is denoted ∂M .

8.2 Coordinatizing manifolds

Definition 8.7. [Lee, p.4] (Chart).

Let M be an n -manifold. A (*coordinate*) *chart* on M is a pair (U, \mathbf{x}) , where $\mathbf{x} : U \rightarrow V \subseteq \mathbb{R}^n$ is a function from an open subset $U \subseteq M$, which is called the *domain* of the chart, to an open subset $V \subseteq \mathbb{R}^n$. Since $\mathbf{x}(\mathbf{p}) = \begin{pmatrix} x^1(\mathbf{p}) \\ \vdots \\ x^n(\mathbf{p}) \end{pmatrix}$, we often refer to the component functions $\{x^i\}_{i=1}^n$ as (*local*) *coordinates*. The component functions are local in the sense that their domain is U , rather than all of M .

A coordinate chart (U, \mathbf{x}) is said to be *about* $\mathbf{p} \in M$ iff $\mathbf{p} \in U$.

Definition 8.8. [Lee, p.13] (Atlas).

Let M be an n -manifold with or without corners. An *atlas* for M is a collection of charts $\{(U_\alpha, \mathbf{x}_\alpha)\}$ whose domains cover M , $M = \cup_\alpha U_\alpha$.

Definition 8.9. (Coordinate representations of functions on manifolds).

Let M be an n -manifold, and let (U, \mathbf{x}) be a chart on M about $\mathbf{p} \in M$. The *coordinate representation of a function* $\mathbf{F} : M \rightarrow \mathbb{R}^k$ *relative to the chart* (U, \mathbf{x}) is the function $\tilde{\mathbf{F}}_{(U, \mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^k$.

Now let N be an m -manifold, consider a function $\mathbf{F} : U \subseteq M \rightarrow N$, and let (V, \mathbf{y}) be a chart about $\mathbf{F}(\mathbf{p}) \in V$. The *coordinate representation of* $\mathbf{F} : U \subseteq M \rightarrow V \subseteq N$ *relative to the charts* (U, \mathbf{x}) *and* (V, \mathbf{y}) is the function $\tilde{\mathbf{F}}_{(U, \mathbf{x})} = \mathbf{y} \circ \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Theorem 8.10. delete?

If x^i is the i th coordinate function of a chart (U, \mathbf{x}) , then the coordinate representation $\tilde{x}^i_{(U, \mathbf{x})}$ of x^i relative to the chart (U, \mathbf{x}) is $\tilde{x}^i_{(U, \mathbf{x})} = x^i \circ \mathbf{x}^{-1} = \phi^{\hat{\mathbf{e}}_i} \in (\mathbb{R}^n)^*$, where $\{\phi^{\hat{\mathbf{e}}_1}, \dots, \phi^{\hat{\mathbf{e}}_n}\}$ is the dual basis of $(\mathbb{R}^n)^*$ induced by the standard basis $\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n\}$ of \mathbb{R}^n . Recall, $\phi^{\hat{\mathbf{e}}_i}(\mathbf{p}) = [\mathbf{p}]_{\mathcal{E}}^i$.

8.3 Smooth manifolds

Definition 8.11. [Lee, p.11] (Differentiability classes, smooth functions, and diffeomorphisms).

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be of *differentiability class* C^n iff the partial derivatives of orders $0, 1, \dots, n$ are continuous. In particular, C^0 functions are real-valued continuous functions on M , and C^1 functions are real-valued continuous functions on M whose first partial derivatives are also continuous. A function $f : M \rightarrow \mathbb{R}$ has *differentiability class* C^∞ iff $f \in C^n(M)$ for all $n \in \{0, 1, \dots\}$.

A function $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^k$ is said to be of *differentiability class* C^n if each component function F_i of \mathbf{F} is of differentiability class C^n , in the previous sense of C^n for functions $\mathbb{R}^n \rightarrow \mathbb{R}$. Similarly a function $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^k$ is said to be of *differentiability class* C^∞ if each component function F_i of \mathbf{F} is of differentiability class C^∞ , in the previous sense of C^∞ for functions $\mathbb{R}^n \rightarrow \mathbb{R}$.

We define $C^n(\mathbb{R}^n \rightarrow \mathbb{R}^k)$ to be the set of functions of differentiability class C^n , and define $C^\infty(\mathbb{R}^n \rightarrow \mathbb{R}^k) := \cup_{n=1}^\infty C^n(M)$. We will use the word “smooth” to mean C^∞ . Following this convention, the set $C^\infty(\mathbb{R}^n \rightarrow \mathbb{R}^k)$ is called the set of *smooth real-valued functions on* \mathbb{R}^k .

Lastly, we say that a function $\mathbb{R}^n \rightarrow \mathbb{R}^k$ is a *diffeomorphism* iff it is smooth, bijective, and has a smooth inverse.

Smooth manifolds with boundary

Derivation 8.12. (Smooth manifold with boundary).

Let M be an n -manifold with or without boundary. We say that a function $\mathbf{F} : M \rightarrow \mathbb{R}^k$ is *smooth relative to a chart* (U, \mathbf{x}) of M iff the coordinate representation $\tilde{\mathbf{F}}_{(U, \mathbf{x})} : \mathbf{x}(U \cap V) \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$ is a smooth function.

Consider two charts (U, \mathbf{x}) and (V, \mathbf{y}) . If $U \cap V \neq \emptyset$, then the *transition map* $\mathbf{y} \circ \mathbf{x}^{-1} : \mathbf{x}(U \cap V) \rightarrow \mathbf{y}(U \cap V)$ is a homeomorphism, so the domains of the coordinate representations $\tilde{\mathbf{F}}_{(U, \mathbf{x})}$ and $\tilde{\mathbf{F}}_{(V, \mathbf{y})}$ are homeomorphic. We would like to define a notion of *smooth manifold with or without boundary* to be such that, if M is a smooth manifold with or without boundary, then

$$\mathbf{F} : M \rightarrow \mathbb{R}^k \text{ is smooth relative to some chart } (U, \mathbf{x})$$

$$\implies$$

$$\mathbf{F} \text{ is smooth relative to all other charts } (V, \mathbf{y}) \text{ that overlap } (U, \mathbf{x}), U \cap V \neq \emptyset.$$

In our present situation, this is not the case. If (U, \mathbf{x}) and (V, \mathbf{y}) intersect, $U \cap V \neq \emptyset$, so that the domains of $\tilde{\mathbf{F}}_{(U, \mathbf{x})}$ and $\tilde{\mathbf{F}}_{(V, \mathbf{y})}$ are homeomorphic, it is still possible for \mathbf{F} to be smooth relative to (U, \mathbf{x}) but not relative to (V, \mathbf{y}) . To see why, express $\tilde{\mathbf{F}}_{(U, \mathbf{x})}$ as $\tilde{\mathbf{F}}_{(U, \mathbf{x})} = \tilde{\mathbf{F}}_{(V, \mathbf{y})} \circ (\mathbf{y} \circ \mathbf{x}^{-1})$. If $\tilde{\mathbf{F}}_{(V, \mathbf{y})}$ is smooth, then $\tilde{\mathbf{F}}_{(U, \mathbf{x})}$ is not guaranteed to be smooth, since composing with a homeomorphism does not preserve smoothness. (In fact, there is always a choice of \mathbf{x} for which $\tilde{\mathbf{F}}_{(U, \mathbf{x})}$ is *not* smooth!).

Since smoothness *is* preserved by composing with a diffeomorphism, we define two charts to be *smoothly compatible* iff the transition map between them is a diffeomorphism, and define a *smooth*

atlas for M to be one in which any two charts are smoothly compatible¹. A chart that is an element of a smooth atlas is called a *smooth chart*.

Definition 8.13. (Smooth functions on manifolds).

$$C^n(M), C^\infty(M)$$

Example 8.14. [Lee, p.20] (Graph of smooth functions into \mathbb{R}^k are smooth manifolds).

If $U \subseteq \mathbb{R}^n$ is an open subset and $\mathbf{F} : U \rightarrow \mathbb{R}^k$ is a smooth function, we have already observed above (Example 1.3) that the graph of \mathbf{F} is a topological n -manifold in the subspace topology. Since $\text{graph}(\mathbf{F})$ is covered by the single graph coordinate chart $\mathbf{x} : \text{graph}(\mathbf{F}) \rightarrow U$ (the restriction of π_1), we can put a canonical smooth structure on $\text{graph}(\mathbf{F})$ by declaring the graph coordinate chart $(\text{graph}(\mathbf{F}), \mathbf{x})$ to be a smooth chart.

Example 8.15. [Lee, p.5] (Graphs of continuous functions into \mathbb{R}^k are manifolds).

Let $U \subseteq \mathbb{R}^n$ be open. The *graph* of a function $f : U \rightarrow \mathbb{R}^k$ is the subset $\text{graph}(f) := U \times f(U)$ of $\mathbb{R}^n \times \mathbb{R}^k$. The graph of a continuous function is a manifold. (See [Lee] for proof).

Example 8.16. [Lee, p.20]

Level sets are also smooth manifolds.

Manifolds with or without boundary or corners

Definition 8.17. [Lee, p.415] (Manifold with or without boundary or corners).

Just as we used \mathbb{H}^n to construct a notion of “manifold boundary”, we will use the topological closure of \mathbb{H}^n to construct a notion of “corners on a manifold”. Observe that the closure of \mathbb{H}^n is

$$\text{cl}(\mathbb{H}^n) = \{(x^1, \dots, x^n) \in \mathbb{R}^n \mid x^i \geq 0 \text{ for all } i\}.$$

We say that a topological space M is a *manifold with or without boundary or corners* iff

- Hausdorff, or “point-separable”.
- *second-countable*; that is, M has a countable basis.
- *locally Euclidean of dimension n* in the sense that each point in M has a neighborhood that is homeomorphic to \mathbb{R}^n (where \mathbb{R}^n has the standard topology).
- such that “ M has a (possibly empty) manifold boundary and M has (possibly no) corners”. That is, each point of M has a neighborhood that is either homeomorphic to an open subset of \mathbb{R}^n (where \mathbb{R}^n has the standard topology), to an open subset \mathbb{H}^n with the subspace topology (inherited from the standard topology on \mathbb{R}^n), or to an open subset of $\text{cl}(\mathbb{H}^n)$ with the subspace topology (inherited from the standard topology on \mathbb{R}^n , and where cl denotes topological closure).

Remark 8.18. [Lee, p.27] (Empty boundary and no corners).

Every smooth manifold is a smooth manifold with boundary, where the manifold boundary is empty. Every smooth manifold with boundary is a smooth manifold with corners, where the set of corner points is empty.

Theorem 8.19. [Lee, p.26] [Lee, p.416] (Topological invariance of interior, boundary, and corner points).

If $\mathbf{p} \in M$ is an interior, boundary, or corner point in some chart, then it is an interior, boundary, or corner point, respectively, in all charts. Furthermore, every interior point is neither a boundary point nor a corner point. (See [Lee] for the proofs).

Definition 8.20. [Lee, p.4] [Lee, p.415] (Classification of charts and points).

We classify a chart (U, \mathbf{x}) on M as follows. (U, \mathbf{x}) is a/an...

¹Note, the empty function is a diffeomorphism, so this definition covers the case in which $U \cap V = \emptyset$ and the transition map is the empty function.

- *Interior chart* iff $\mathbf{x}(U)$ is an open subset of \mathbb{R}^n .
- *Boundary chart* iff $\mathbf{x}(U)$ contains a boundary point of \mathbb{H}^n , i.e., iff $\mathbf{x}(U)$ is an open subset of \mathbb{H}^n that intersects the boundary of \mathbb{H}^n , $\mathbf{x}(U) \cap \partial\mathbb{H}^n \neq \emptyset$.
- *Chart with corners* iff $\mathbf{x}(U)$ contains a *corner point* of $\text{cl}(\mathbb{H}^n)$; a point $\mathbf{p} \in \text{cl}(\mathbb{H}^n)$ is a *corner point* of $\text{cl}(\mathbb{H}^n)$ iff more than one of the coordinate functions $x^i|_{\mathbf{p}}$ evaluated at \mathbf{p} vanish.

We classify a point $\mathbf{p} \in M$ according to the type of chart it lies in. $\mathbf{p} \in M$ is a/an...

- *Interior point* iff there is an interior chart about \mathbf{p} .
- *Boundary point* iff there is a boundary chart about \mathbf{p} .
- *Corner point* iff there is a chart with corners about \mathbf{p} .

Definition 8.21. (Smooth manifolds with or without boundary or corners).

Remark 8.22. [Lee, p.415] (Manifolds with corners are *topologically* the same as manifolds with boundary).

The title of this theorem is true because $\text{cl}(\mathbb{H}^n)$ is homeomorphic to \mathbb{H}^n . *Smooth* manifolds with corners are different than *smooth* manifolds with boundary because [...]

8.4 Tangent vectors

No way to directly pull back geometric vectors from $\mathbb{R}_{\mathbf{F}(\mathbf{p})}^n$ to M . We can however pull back derivations at $\mathbf{F}(\mathbf{p}) \in \mathbb{R}^n$.

Definition 8.23. (Derivation at $\mathbf{p} \in \mathbb{R}^n$, tangent space to \mathbb{R}^n).

A *derivation at $\mathbf{p} \in \mathbb{R}^n$* is a function $v_{\mathbf{p}} : C^\infty(\mathbb{R}^n \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$ which satisfies

$$\begin{aligned} v_{\mathbf{p}}(f + g) &= v_{\mathbf{p}}(f) + v_{\mathbf{p}}(g) \\ v_{\mathbf{p}}(fg) &= v_{\mathbf{p}}(f)g(\mathbf{p}) + f(\mathbf{p})v_{\mathbf{p}}(g), \end{aligned}$$

for all $f, g \in C^\infty(\mathbb{R}^n \rightarrow \mathbb{R})$. Due to these conditions, we will see in Theorem [...] that every derivation is a directional derivative.

The set of derivations at $\mathbf{p} \in \mathbb{R}^n$ is called the *tangent space to \mathbb{R}^n at \mathbf{p}* , and is denoted $T_{\mathbf{p}}(\mathbb{R}^n)$.

Theorem 8.24. (Basis of $T_{\mathbf{p}}(\mathbb{R}^n)$).

When $M = \mathbb{R}^n$, the directional derivatives $\left\{ \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right\}_{i=1}^n$ form a basis for $T_{\mathbf{p}}(\mathbb{R}^n)$.

Proof. Take the standard basis $\{\hat{\mathbf{e}}_{\mathbf{p}}^i\}_{i=1}^n$ for $\mathbb{R}_{\mathbf{p}}^n$. Since $\mathbf{v}_{\mathbf{p}} \mapsto \frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$ is an isomorphism $\mathbb{R}_{\mathbf{p}}^n \rightarrow T_{\mathbf{p}}(\mathbb{R}^n)$, then the images of the $\hat{\mathbf{e}}_{\mathbf{p}}^i$ under this isomorphism are a basis for $T_{\mathbf{p}}(\mathbb{R}^n)$. Thus $\left\{ \frac{\partial}{\partial \hat{\mathbf{e}}^i} \Big|_{\mathbf{p}} \right\}_{i=1}^n = \left\{ \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right\}_{i=1}^n$ is a basis for $T_{\mathbf{p}}(\mathbb{R}^n)$. \square

Lemma 8.25. (Properties of derivations at $\mathbf{p} \in \mathbb{R}^n$). Suppose $\mathbf{p} \in \mathbb{R}^n$, $v_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^n)$ is a derivation at \mathbf{p} , and $f, g \in C^\infty(\mathbb{R}^n \rightarrow \mathbb{R})$. Then

- If f is a constant function, then $v_{\mathbf{p}}(f) = 0$.
- If $f(\mathbf{p}) = g(\mathbf{p}) = 0$, then $v_{\mathbf{p}}(fg) = 0$.

Proof.

- First set $f_1 \equiv 1$ and use the product rule with $f_1 \cdot f_1 = f_1$ to show $v_{\mathbf{p}}(f_1) = 0$. Any other constant function $f_c \equiv c$ is of the form $f_c = cf_1$, so $v_{\mathbf{p}}(f_c) = v_{\mathbf{p}}(cf_1) = cv_{\mathbf{p}}(f_1) = 0$ by linearity.
- Use the product rule.

□

Theorem 8.26. (Derivations and directional derivatives are in one-to-one correspondence).

The set of directional derivatives at $\mathbf{p} \in \mathbb{R}^n$ is equal to the set of derivations at \mathbf{p} :

$$\left\{ \frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}} \mid \mathbf{v} \in \mathbb{R}^n \right\} = T_{\mathbf{p}}(\mathbb{R}^n).$$

Proof.

(\subseteq). Showing that every directional derivative at $\mathbf{p} \in \mathbb{R}^n$ is a derivation at $\mathbf{p} \in \mathbb{R}^n$ follows straightforwardly from the definition of “derivation at $\mathbf{p} \in \mathbb{R}^n$ ”.

(\supseteq). Let $\mathbf{p} \in \mathbb{R}^n$ and let $v_{\mathbf{p}} : C^\infty(\mathbb{R}^n \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$ be a derivation. We must show that $v_{\mathbf{p}}$ is a directional derivative at \mathbf{p} , and find a vector $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}_{\mathbf{p}}^n$ for which $v_{\mathbf{p}} = \frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$. So, we will find a $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}_{\mathbf{p}}^n$ for which

$$v_{\mathbf{p}}(f) = \left(\frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}} \right)(f) \text{ for all } f \in C^\infty(\mathbb{R}^n \rightarrow \mathbb{R}).$$

We use Taylor’s theorem (cite) to write $f \in C^\infty(\mathbb{R}^n \rightarrow \mathbb{R})$ as

$$f(\mathbf{x}) = f(\mathbf{p}) + \sum_{i=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} (x^i - p^i) + \sum_{i,j=1}^n (x^i - p^i)(x^j - p^j) \int_0^1 (1-t) \frac{\partial^2 f}{\partial x^i \partial x^j} \Big|_{\mathbf{p}+t(\mathbf{x}-\mathbf{p})} dt,$$

where $x^i = [\mathbf{x}]_E^i$, $p^i = [\mathbf{p}]_E^i$.

Now we produce the vector $\mathbf{v}_{\mathbf{p}}$. Let $\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n\}$ be the standard basis for \mathbb{R}^n and let $\{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ be the induced dual basis for $(\mathbb{R}^n)^*$. We set $\mathbf{v}_{\mathbf{p}} = \sum_{i=1}^n v_{\mathbf{p}}(\phi^{\mathbf{e}_i}) \mathbf{e}_i$. (Note that applying the derivation $v_{\mathbf{p}}$ to $\phi^{\mathbf{e}_i}$ makes sense because $\phi^{\mathbf{e}_i}$, being the “ i th coordinate function on \mathbb{R}^n ”, is a smooth function $\mathbb{R}^n \rightarrow \mathbb{R}$). This choice of $\mathbf{v}_{\mathbf{p}}$ makes more sense in hindsight after reading Lemma [coords]. (That lemma logically depends on this result, though, so we must make this relatively unmotivated choice of $\mathbf{v}_{\mathbf{p}}$!)

Apply $v_{\mathbf{p}}$ to f and use Lemma [...] to obtain

$$v_{\mathbf{p}}(f) = v_{\mathbf{p}} \left(\sum_{i,j=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} (x^i - p^i) \right) = \sum_{i=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} (v_{\mathbf{p}}(x^i) - v_{\mathbf{p}}(p^i)) = \sum_{i=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} v_{\mathbf{p}}(x^i).$$

When we apply $v_{\mathbf{p}}$ to x^i in the above, we are using a slight abuse of notation and interpreting x^i to be the coordinate function $\mathbf{x} \mapsto x^i$ evaluated at \mathbf{x} . Thinking of x^i in this way, we have $x^i = \phi^{\mathbf{e}_i}$, so the above further simplifies to

$$\sum_{i=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} v_{\mathbf{p}}(\phi^{\mathbf{e}_i}) = (\nabla \mathbf{F})_{\mathbf{p}} \cdot \mathbf{v}_{\mathbf{p}} = \frac{\partial f}{\partial \mathbf{v}} \Big|_{\mathbf{p}}.$$

Thus, $v_{\mathbf{p}} = \frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$.

□

Theorem 8.27. (Geometric tangent vectors are naturally isomorphic to derivations).

The map $\mathbf{v}_{\mathbf{p}} \mapsto \frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$ is a natural linear isomorphism $\mathbb{R}_{\mathbf{p}}^n \cong T_{\mathbf{p}}(\mathbb{R}^n)$.

Proof. We need to show $\mathbf{v} \mapsto \frac{\partial}{\partial \mathbf{v}}$ is linear, injective and surjective. The naturality of the isomorphism is immediate because it is a basis-independent definition.

Linearity follows immediately from the fact that the directional derivative $\frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$ is linear with respect to \mathbf{v} (see Remark 6.19). The map $\mathbf{v}_{\mathbf{p}} \mapsto \frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$ is surjective because every derivation at $\mathbf{p} \in \mathbb{R}^n$ is a directional derivative at $\mathbf{p} \in \mathbb{R}^n$, so every derivation at $\mathbf{p} \in \mathbb{R}^n$ is of the form $\frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$ for some $\mathbf{v} \in \mathbb{R}^n$.

It remains to show injectivity. We show the map $\mathbf{v}_{\mathbf{p}} \mapsto \frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$ has a trivial kernel. So, assume $\mathbf{v}_{\mathbf{p}}$ is sent to the zero function; we need to show $\mathbf{v}_{\mathbf{p}} = \mathbf{0}_{\mathbf{p}}$.

Let $E = \{\mathbf{e}^1, \dots, \mathbf{e}^n\}$ be a basis for \mathbb{R}^n , and let $E^* = \{\phi^{\mathbf{e}^1}, \dots, \phi^{\mathbf{e}^n}\}$ be its dual basis for $(\mathbb{R}^n)^*$. Note that since $\phi^{\mathbf{e}^i} : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined by $\phi^{\mathbf{e}^i}(\mathbf{p}) = [\mathbf{p}]_E^i$, each $\phi^{\mathbf{e}^i}$ is a C^∞ function on \mathbb{R}^n .

At $\mathbf{p} \in \mathbb{R}^n$, the i th coordinate of the zero map is $\frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}(\phi^{\mathbf{e}^i})$ (see Theorem 4.26). All coordinates of the zero function relative to any basis must be zero, so, using the linearity of the directional derivative with respect to $\mathbf{v}_{\mathbf{p}}$ (again, see Remark 6.19), we have

$$0 = \frac{\partial}{\partial \mathbf{v}} \Big|_{\mathbf{p}}(\phi^{\mathbf{e}^i}) = \frac{\partial}{\partial(\sum_j [\mathbf{v}]_E^j \mathbf{e}^j)} \Big|_{\mathbf{p}}(\phi^{\mathbf{e}^i}) = \sum_j [\mathbf{v}]_E^j \left(\frac{\partial}{\partial \mathbf{e}^j} \Big|_{\mathbf{p}} \right) (\phi^{\mathbf{e}^i}).$$

Since $\frac{\partial}{\partial \mathbf{e}^j} \Big|_{\mathbf{p}} = \frac{\partial}{\partial x^j} \Big|_{\mathbf{p}}$, the above becomes

$$0 = \sum_j [\mathbf{v}]_E^j \delta_j^i = [\mathbf{v}]_E^i.$$

We see $[\mathbf{v}]_E^i = 0$ for all i , so $\mathbf{v} = \mathbf{0}$, which means $\mathbf{v}_{\mathbf{p}} = \mathbf{0}_{\mathbf{p}}$. □

Definition 8.28. (Derivation at $\mathbf{p} \in M$, tangent space to a manifold).

Let M be a smooth n -manifold with or without boundary. A *derivation at $\mathbf{p} \in M$* is a linear function $v_{\mathbf{p}} : C^\infty(M \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$ that satisfies the product rule,

$$v_{\mathbf{p}}(fg) = v_{\mathbf{p}}(f)g(\mathbf{p}) + f(\mathbf{p})v_{\mathbf{p}}(g) \text{ for all } f, g \in C^\infty(M \rightarrow \mathbb{R}).$$

The set of derivations at $\mathbf{p} \in M$ is called the *tangent space to M at \mathbf{p}* , and is denoted $T_{\mathbf{p}}(M)$.

Theorem 8.29. (Properties of derivations at $\mathbf{p} \in M$).

Let M be a smooth n -manifold with or without boundary, let $\mathbf{p} \in \mathbb{R}^n$, $v_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^n)$ be a derivation at \mathbf{p} , and $f, g \in C^\infty(\mathbb{R}^n \rightarrow \mathbb{R})$. Then

- If f is a constant function, then $v_{\mathbf{p}}(f) = 0$.
- If $f(\mathbf{p}) = g(\mathbf{p}) = 0$, then $v_{\mathbf{p}}(fg) = 0$.

Proof. The proof is analogous to the proof of Lemma [...]. □

Definition 8.30. [War, p.14, 15] (Basis of $T_{\mathbf{p}}(M)$ and its abuse of notation).

Let M be a smooth n -manifold with or without boundary, let (U, \mathbf{x}) be a smooth chart on M , and let $\mathbf{p} \in M$.

In an abuse of notation, we define $\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}}$, $i \in \{1, \dots, n\}$, to be the derivations at $\mathbf{p} \in U \subseteq M$ for which

$$\left(\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right) (f) = \left(\frac{\partial}{\partial x^i} \Big|_{\mathbf{x}(\mathbf{p})} \right) (f \circ \mathbf{x}^{-1}),$$

where the $\frac{\partial}{\partial x^i}\big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial}{\partial \mathbf{e}^i}\big|_{\mathbf{x}(\mathbf{p})}$ on the right hand side are directional derivatives taking in smooth functions on \mathbb{R}^n as their arguments. The $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}$ on the left hand side are what we are defining in our abuse of notation, and (we will see) are derivations $C^\infty(M \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$; that is, they are elements of $T_{\mathbf{p}}(M)$.

Importantly, $\left\{\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}\right\}_{i=1}^n$ is a basis of $T_{\mathbf{p}}(M)$. This follows because the $\frac{\partial}{\partial x^i}\big|_{\mathbf{x}(\mathbf{p})}$ on the right hand side are a basis for $T_{\mathbf{x}(\mathbf{p})}(\mathbb{R}^n)$ (see Theorem 8.24).

Lastly, if we consider the coordinate representation $\tilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1}$ of \mathbf{F} relative to the chart (U, \mathbf{x}) , then the above definition becomes

$$\left(\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}\right)(f) = \frac{\partial}{\partial x^i}\big|_{\mathbf{x}(\mathbf{p})} \tilde{\mathbf{F}}_{(U,\mathbf{x})}.$$

Proof. We need to show that the $\underbrace{\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}}_{\text{LHS}} : C^\infty(M \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$ from the left side of the condition of the

definition are derivations at $\mathbf{p} \in M$. In this proof, we put an “LHS” under $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}$ when we mean $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}$ to be from the left hand side of the condition in the above definition. All other occurrences of $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}$ in this proof are directional derivatives.

We need to show that $\underbrace{\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}}_{\text{LHS}}$ are linear and follow the product rule. Linearity follows easily from

the linearity of the directional derivative with respect to its argument from $C^\infty(\mathbb{R}^n \rightarrow \mathbb{R})$. We show the product rule holds:

$$\begin{aligned} \underbrace{\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}}_{\text{LHS}}(fg) &= \frac{\partial(fg \circ \mathbf{x}^{-1})}{\partial x^i}\big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\big|_{\mathbf{x}^{-1}(\mathbf{x}(\mathbf{p}))} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\big|_{\mathbf{x}(\mathbf{p})} \\ &= \left(\frac{\partial f}{\partial x^i}\big|_{\mathbf{p}} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p})) \frac{\partial g}{\partial x^i}\big|_{\mathbf{p}}\right) \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\big|_{\mathbf{x}(\mathbf{p})} \\ &= \frac{\partial f}{\partial x^i}\big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\big|_{\mathbf{x}(\mathbf{p})} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p})) \frac{\partial g}{\partial x^i}\big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\big|_{\mathbf{x}(\mathbf{p})} \\ &= \frac{\partial(f \circ \mathbf{x})}{\partial x^i}\big|_{\mathbf{p}} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p})) \frac{\partial(g \circ \mathbf{x})}{\partial x^i}\big|_{\mathbf{p}} \\ &= \underbrace{\left(\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}\right)}_{\text{LHS}}(f) g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p})) \underbrace{\left(\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}\right)}_{\text{LHS}}(g). \end{aligned}$$

Here, we used the chain and product rules and then reversed the product and chain rules. \square

The differential of a smooth function $M \rightarrow N$

Definition 8.31. (Differential of a smooth function $\mathbb{R}^n \rightarrow \mathbb{R}^m$).

Consider a smooth function $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$. We define the *differential* $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(\mathbb{R}^n) \rightarrow T_{\mathbf{F}(\mathbf{p})}(\mathbb{R}^m)$ of \mathbf{F} at \mathbf{p} to be the linear function whose matrix relative to the bases $\{\frac{\partial}{\partial x^i}\}$ for $T_{\mathbf{p}}(\mathbb{R}^n)$ and $\{\frac{\partial}{\partial y^i}\}$ for $T_{\mathbf{p}}(\mathbb{R}^m)$ is the Jacobian matrix of \mathbf{F} at \mathbf{p} :

$$d\mathbf{F}_{\mathbf{p}} := \frac{\partial \mathbf{F}}{\partial \mathbf{x}}\big|_{\mathbf{p}} = \left(\frac{\partial F^i}{\partial x^j}\big|_{\mathbf{p}}\right).$$

(Here, F^i is the i th component function of \mathbf{F}).

(We will show that if $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$, then $d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})$ is a derivation at $\mathbf{F}(\mathbf{p})$ in $T_{\mathbf{F}(\mathbf{p})}(N)$ later). This means that $d\mathbf{F}_{\mathbf{p}}$ acts on a basis tangent vector $\frac{\partial}{\partial x^i}$ by

$$d\mathbf{F}_{\mathbf{p}}\left(\frac{\partial}{\partial x^i}\right) = \sum_{j=1}^m a_i^j \frac{\partial}{\partial y^j} = \sum_{j=1}^m \frac{\partial F^j}{\partial x^i} \Big|_{\mathbf{p}} \frac{\partial}{\partial y^j} \Big|_{\mathbf{F}(\mathbf{p})},$$

where $(a_i^j) = \left[d\mathbf{F}_{\mathbf{p}}\left(\left\{\frac{\partial}{\partial x^i}\right\}\right) \right]_{\left\{\frac{\partial}{\partial y^j}\right\}}$ is the matrix of $d\mathbf{F}_{\mathbf{p}}$ relative to the bases $\left\{\frac{\partial}{\partial x^i}\right\}$ and $\left\{\frac{\partial}{\partial y^j}\right\}$.

Using linearity, we see $d\mathbf{F}_{\mathbf{p}}\left(\frac{\partial}{\partial x^i}\right)$ acts on $f \in C^\infty(U \subseteq \mathbb{R}^n \rightarrow \mathbb{R})$ by

$$\begin{aligned} d\mathbf{F}_{\mathbf{p}}\left(\frac{\partial}{\partial x^i}\right)(f) &= \left(\sum_{j=1}^m \frac{\partial F^j}{\partial x^i} \Big|_{\mathbf{p}} \frac{\partial}{\partial y^j} \Big|_{\mathbf{F}(\mathbf{p})} \right)(f) = \sum_{j=1}^m \left[\left(\frac{\partial F^j}{\partial x^i} \Big|_{\mathbf{p}} \frac{\partial}{\partial y^j} \Big|_{\mathbf{F}(\mathbf{p})} \right)(f) \right] \\ &= \sum_{j=1}^m \frac{\partial F^j}{\partial x^i} \Big|_{\mathbf{p}} \frac{\partial f}{\partial y^j} \Big|_{\mathbf{F}(\mathbf{p})} = \sum_{j=1}^m \frac{\partial f}{\partial y^j} \Big|_{\mathbf{F}(\mathbf{p})} \frac{\partial F^j}{\partial x^i} \Big|_{\mathbf{p}} = \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} (f \circ \mathbf{F}). \end{aligned}$$

Note, this last $\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}}$ is a directional derivative, $\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} = \frac{\partial}{\partial \hat{\mathbf{e}}^i} : C^\infty(\mathbb{R}^n \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$.

The above shows

$$d\mathbf{F}_{\mathbf{p}}\left(\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}}\right)(f) = \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} (f \circ \mathbf{F}).$$

Since $d\mathbf{F}_{\mathbf{p}}$ is linear and $\left\{\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}}\right\}$ is a basis for $T_{\mathbf{p}}(\mathbb{R}^n)$, the above condition extends to any $v_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^n)$:

$$d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})(f) = v_{\mathbf{p}}(f \circ \mathbf{F}).$$

This characterization of the differential is coordinate free, and therefore provides an easy way to define the differential in a more general setting. We do this in the next definition.

Definition 8.32. (Differential of a smooth function $\mathbf{F} : M \rightarrow N$).

Let M and N be smooth n - and m - dimensional manifolds with or without boundary.

We define the *differential* $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \rightarrow T_{\mathbf{F}(\mathbf{p})}(N)$ of \mathbf{F} at \mathbf{p} by

$$d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})(f) = v_{\mathbf{p}}(f \circ \mathbf{F}),$$

where $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ and $f \in C^\infty(N)$.

Proof. We need to check that $d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})$ is a derivation at $\mathbf{F}(\mathbf{p}) \in N$. To do so, follow the proof of Theorem [...], which showed $\underbrace{\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}}}_{\text{LHS}}$ is a derivation at $\mathbf{x}(\mathbf{p}) \in \mathbb{R}^n$. \square

Theorem 8.33. (Properties of differentials). Let M, N , and P be smooth manifolds with or without boundary, let $\mathbf{F} : M \rightarrow N$ and $\mathbf{G} : N \rightarrow P$ be smooth functions, and let $\mathbf{p} \in M$.

- (Chain rule). $d(\mathbf{G} \circ \mathbf{F})_{\mathbf{p}} = d\mathbf{G}_{\mathbf{F}(\mathbf{p})} \circ d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \rightarrow T_{(\mathbf{G} \circ \mathbf{F})(\mathbf{p})}(P)$.
- (Differential of the identity). $d(\mathbf{I}_M)_{\mathbf{p}} = \mathbf{I}_M$, where $\mathbf{I}_M : T_{\mathbf{p}}(M) \rightarrow T_{\mathbf{p}}(M)$ is the identity on M .
- If \mathbf{F} is a diffeomorphism, then $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \rightarrow T_{\mathbf{F}(\mathbf{p})}(N)$ is a linear isomorphism, and $(d\mathbf{F}_{\mathbf{p}})^{-1} = d(\mathbf{F}^{-1})_{\mathbf{F}(\mathbf{p})}$.

Proof. See Exercise 3.7 of Lee. [cite more] \square

Tangent covectors

Theorem 8.34. [Lee, p.281] (Differential of a smooth function $M \rightarrow \mathbb{R}$).

What happens when we take the differential of a smooth function $f : M \rightarrow \mathbb{R}$? Well, by definition of the differential of a smooth function $M \rightarrow \mathbb{R}$, we have

$$df_{\mathbf{p}}(v_{\mathbf{p}})(g) = v_{\mathbf{p}}(g \circ f).$$

We have $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \rightarrow T_{f(\mathbf{p})}(\mathbb{R})$. Notice that since $T_{f(\mathbf{p})}(\mathbb{R})$ is 1-dimensional, then $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$. There is therefore an induced function $\tilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \rightarrow \mathbb{R}$ for which the following diagram commutes:

$$\begin{array}{ccc} T_{\mathbf{p}}(M) & \xrightarrow{v_{\mathbf{p}} \mapsto df_{\mathbf{p}}(v_{\mathbf{p}})} & T_{f(\mathbf{p})}(\mathbb{R}) \\ & \searrow \tilde{d}f_{\mathbf{p}} & \downarrow w \mapsto \sum_{i=1}^n w(x^i) \hat{\mathbf{e}}_i \\ & & \mathbb{R} \end{array}$$

We think of \tilde{d} as the differential that is induced by the identification $T_{\mathbf{q}}(\mathbb{R}) \cong \mathbb{R}$.

From the diagram, we see that the map $\tilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \rightarrow \mathbb{R}$ acts on a tangent vector $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by

$$\tilde{d}f_{\mathbf{p}}(v_{\mathbf{p}}) = \sum_{i=1}^1 df_{\mathbf{p}}(v_{\mathbf{p}})(x^i) \hat{\mathbf{e}}_i = df_{\mathbf{p}}(v_{\mathbf{p}})(x^1) \hat{\mathbf{e}}_1 = v_{\mathbf{p}}(x^1 \circ f) \hat{\mathbf{e}}_1.$$

Since \mathbb{R}^1 is one-dimensional, then $\mathbb{R}^1 \cong \mathbb{R}$. As we switch from \mathbb{R}^1 to \mathbb{R} , the coordinate function $x^1 : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ becomes the identity on \mathbb{R} and $\hat{\mathbf{e}}_1$ becomes the scalar $1 \in \mathbb{R}$. These final identifications² give

$$\tilde{d}f_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f).$$

In practice, we write $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \rightarrow \mathbb{R}$ to mean $\tilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \rightarrow \mathbb{R}$. (This includes “ dx^i ”; whenever x^i is a coordinate function, we write dx^i to mean $\tilde{d}x^i$, where \tilde{d} is the differential obtained by identifying $T_{\mathbf{q}}(\mathbb{R}) \cong \mathbb{R}$). So, the above characterizing condition of the differential of $f : M \rightarrow \mathbb{R}$ is restated as

$$\boxed{df_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)}$$

Theorem 8.35. [Lee, p.281] (Differential of a smooth function $M \rightarrow \mathbb{R}$ in coordinates).

Let M be a smooth n -manifold with or without boundary or corners, and let (U, \mathbf{x}) be a smooth chart on M , with x^i being the i th coordinate function of \mathbf{x} . Then the differential of a smooth function $f : M \rightarrow \mathbb{R}$ obtained by identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ (see the previous theorem) is given by

$$\boxed{df_{\mathbf{p}} = \sum_{i=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}}$$

Note, the d on the right hand side is the differential of the i th coordinate function x^i of \mathbf{x} after identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ (see the previous theorem).

²To be very formal, we could write the above as $\tilde{\tilde{d}}f_{\mathbf{p}}(v_{\mathbf{p}})$ to indicate that $\tilde{\tilde{d}}$ is the differential obtained from \tilde{d} by identifying $\mathbb{R}^1 \cong \mathbb{R}$, but this identification is so trivial that we do not do this.

Proof. Notice that since $\mathbf{v} \mapsto df_{\mathbf{p}}(\mathbf{v})$ is a linear function, $df_{\mathbf{p}}$ is an element of $T_{\mathbf{p}}(M)^*$. (We will explore this fact more later, in Section 8.5).

Since $df_{\mathbf{p}} \in T_{\mathbf{p}}(M)^*$, we can decompose $df_{\mathbf{p}}$ by using the dual basis $\{\lambda^i|_{\mathbf{p}}\}_{i=1}^n$ for $T_{\mathbf{p}}(M)^*$ induced by the basis $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}_{i=1}^n$ for $T_{\mathbf{p}}(M)$. (The notation $\{\lambda^i|_{\mathbf{p}}\}_{i=1}^n$ is temporary; by the end of this proof, we will have a more meaningful notation for this dual basis). Thus, we can write

$$df_{\mathbf{p}} = \sum_{i=1}^n C_i(\mathbf{p}) \lambda^i|_{\mathbf{p}}$$

for some smooth functions $C_i : U \rightarrow \mathbb{R}$.

To determine the C_i , recall that Theorem 4.26 stated that if V is a finite-dimensional vector space, $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for V , and $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ is the basis for V^* induced by E , then for any $\phi \in V^*$, we have $([\phi]_{E^*})_i = \phi(\mathbf{e}_i)$. Applying this theorem and using $df_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)$, we have

$$C_i(\mathbf{p}) = df_{\mathbf{p}}\left(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right) = \left(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right)(f) = \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}.$$

Thus $df_{\mathbf{p}}$ is

$$df_{\mathbf{p}} = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \lambda^i|_{\mathbf{p}}.$$

It remains to determine the $\lambda^i|_{\mathbf{p}}$. We claim that $\lambda^i|_{\mathbf{p}} = dx^i|_{\mathbf{p}}$, where x^i is the i th coordinate function of \mathbf{x} , and where d of a smooth function $M \rightarrow \mathbb{R}$ obtained by identifying $T_{\mathbf{q}}(\mathbb{R}) \cong \mathbb{R}$ (so d is same differential that we've been using in this proof; it is the differential \tilde{d} of the previous theorem). To prove the claim, we compute

$$dx^i|_{\mathbf{p}} = \sum_{j=1}^n \frac{\partial x^i}{\partial x^j}\Big|_{\mathbf{p}} \lambda^j|_{\mathbf{p}}.$$

In the sum, $\frac{\partial x^i}{\partial x^j}$ is the partial derivative of the i th coordinate function $x^i : U \rightarrow \mathbb{R}$ relative to x^j . This partial derivative is $\frac{\partial x^i}{\partial x^j} = \delta_j^i$, so we obtain

$$dx^i|_{\mathbf{p}} = \sum_{j=1}^n \delta_j^i \lambda^j|_{\mathbf{p}} = \lambda^i|_{\mathbf{p}}.$$

□

Remark 8.36. (Differential of a smooth function $\mathbb{R}^n \rightarrow \mathbb{R}$ is the directional derivative).

When $M = \mathbb{R}^n$, we can use the first line in the boxed equation to compute

$$df_{\mathbf{p}} = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}.$$

Since $M = \mathbb{R}^n$, then $df_{\mathbf{p}} : T_{\mathbf{p}}(\mathbb{R}^n) \rightarrow \mathbb{R}$. Notice that we have $T_{\mathbf{q}}(\mathbb{R}^n) \cong \mathbb{R}^n$ because $T_{\mathbf{p}}(\mathbb{R}^n)$ and $\mathbb{R}_{\mathbf{p}}^n$ are both n -dimensional, so we get from $df_{\mathbf{p}} : T_{\mathbf{p}}(\mathbb{R}^n) \rightarrow \mathbb{R}$ the induced function $\tilde{d}f_{\mathbf{p}} : \mathbb{R}_{\mathbf{p}}^n \rightarrow \mathbb{R}$, where \tilde{d} is thought of as the “induced differential”. Above, we see that our original $df_{\mathbf{p}}$ is a linear combination of $dx^i|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. Thus, $\tilde{d}f_{\mathbf{p}}$ is a linear combination of $\tilde{d}x^i|_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^n)$ with the same weights:

$$\sum_{i=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} \tilde{dx}^i$$

In the proof of the previous theorem, we saw that $dx^i|_{\mathbf{p}}$ is the i th coordinate function on $T_{\mathbf{p}}(M)$. Since $T_{\mathbf{p}}(\mathbb{R}^n) \cong \mathbb{R}_{\mathbf{p}}^n$, then \tilde{dx}^i must be the i th coordinate function on $\mathbb{R}_{\mathbf{p}}^n$. This implies that $\tilde{df}_{\mathbf{p}}$ acts on $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}_{\mathbf{p}}^n$ by

$$\tilde{df}_{\mathbf{p}}(\mathbf{v}_{\mathbf{p}}) = \sum_{i=1}^n \frac{\partial f}{\partial x^i} \tilde{dx}^i(\mathbf{v}) = \sum_{i=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} [\mathbf{v}_{\mathbf{p}}]_{\mathcal{E}}^i = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} = \frac{\partial f}{\partial \mathbf{v}} \Big|_{\mathbf{p}},$$

where $\frac{\partial f}{\partial \mathbf{v}} \Big|_{\mathbf{p}}$ is the directional derivative of f at \mathbf{p} in the direction of \mathbf{v} . This is to be expected because we defined the differential of a smooth function of smooth manifolds (see Definition [...]) so that its coordinate representation is represented by the Jacobian relative to the coordinate bases. (Here, the row-matrix of partial derivatives is the Jacobian matrix of f . Recall from Theorem [jacobian] that the Jacobian is used to express the directional derivative of a smooth function $\mathbb{R}^n \rightarrow \mathbb{R}^m$. In this case we have $m = 1$).

8.5 Tangent vectors and tangent covectors with coordinates

Definition 8.37. (Cotangent space to a manifold).

Let M be a smooth n -manifold, and let $\mathbf{p} \in M$. The *cotangent space* $T_{\mathbf{p}}^*(M)$ to M at \mathbf{p} is the dual vector space to the tangent space at \mathbf{p} . That is, ${}_{\mathbf{p}}^*(M) := T_{\mathbf{p}}(M)^*$.

An element $\phi_{\mathbf{p}} \in T_{\mathbf{p}}^*(M)$ of the cotangent space at \mathbf{p} is called a *tangent covector*, or a *covector*, for short.

Theorem 8.38. (Induced bases in a chart).

Let M be a smooth n -manifold with or without boundary and let (U, \mathbf{x}) be a smooth chart on M about $\mathbf{p} \in M$. Consider the set $C^\infty(U \subseteq M \rightarrow \mathbb{R})$ of smooth real-valued functions defined on $U \subseteq M$ as a vector space over \mathbb{R} . Then

1. $\{x^i|_{\mathbf{p}}\}_{i=1}^n$ is a basis for $C^\infty(U \subseteq M \rightarrow \mathbb{R})$, where x^i is the i th coordinate function of \mathbf{x} , and where we've denoted $x^i|_{\mathbf{p}} := x^i(\mathbf{p})$.
2. The dual basis for $T_{\mathbf{p}}(M) = C^\infty(U \subseteq M \rightarrow \mathbb{R})^*$ induced by $\{x^i\}_{i=1}^n$ is $\left\{ \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right\}$.
3. The dual basis for $T_{\mathbf{p}}^*(M) = T_{\mathbf{p}}(M)^* = C^\infty(U \subseteq M \rightarrow \mathbb{R})^{**}$ induced by $\left\{ \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right\}$ is $dx^i|_{\mathbf{p}}$, where x^i is the i th coordinate function of \mathbf{x} , and where the d here is *not* obtained by identifying $T_{\mathbf{p}}(\mathbb{R}^n) \cong \mathbb{R}$, but by “leaving $T_{\mathbf{p}}(\mathbb{R}^n)$ alone” (recall Theorem 8.34 to see happens to d when we identify $T_{\mathbf{q}}(\mathbb{R}) \cong \mathbb{R}$).

Proof.

1. hmm
2. $\left(\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right) (x^j) = \delta_i^j$
3. proved as part of showing $df_{\mathbf{p}} = \sum_{i=1}^n \frac{\partial f}{\partial x^i} \Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}$ in Theorem 8.34

□

Theorem 8.39. (Coordinates of tangent vectors and dual vectors).

Theorem 4.26 stated that if V is a finite-dimensional vector space over K , $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for V , and $E^* = \{\phi^{\mathbf{e}_1}, \dots, \phi^{\mathbf{e}_n}\}$ is the basis for V^* induced by E , then

$$\begin{aligned} ([\mathbf{v}]_E)^i &= \phi^{\mathbf{e}_i}(\mathbf{v}) = \phi_{\mathbf{v}}(\Phi_{\mathbf{e}_i}) \\ ([\phi]_{E^*})_i &= \phi(\mathbf{e}_i), \end{aligned}$$

where $\Phi_{\mathbf{v}} \in V^{**}$ is the linear function $V^* \rightarrow K$ defined by $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$.

We can apply this theorem to the pairs of bases and induced dual bases from the last theorem. Let M be a smooth n -manifold with or without boundary and let (U, \mathbf{x}) be a smooth chart on M about $\mathbf{p} \in M$. Then the i th coordinate of a tangent vector $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ relative to $\left\{ \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right\}_{i=1}^n$ and the i th coordinate of a tangent covector $\phi_{\mathbf{p}} \in T_{\mathbf{p}}^*(M)$ are

$$\boxed{\begin{aligned} ([v_{\mathbf{p}}]_{\left\{ \frac{\partial}{\partial x^j} \Big|_{\mathbf{p}} \right\}_{j=1}^n})^i &= v_{\mathbf{p}}(x^i) = \phi_{x^i}(v_{\mathbf{p}}) \\ ([\phi_{\mathbf{p}}]_{\left\{ dx^j \Big|_{\mathbf{p}} \right\}_{j=1}^n})_i &= \phi_{\mathbf{p}}\left(\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right) \end{aligned}}$$

In the second equation of the first line, ϕ_{x^i} is the element of $C^\infty(U \subseteq M \rightarrow \mathbb{R})^{**} = T_{\mathbf{p}}^*(M)$ that is identified with the i th coordinate function $x^i \in C^\infty(U \subseteq M \rightarrow \mathbb{R})$ of \mathbf{x} . Recall Theorem 3.23 to see that $\phi_f : C^\infty(U \subseteq M \rightarrow \mathbb{R})^* = T_{\mathbf{p}}(M) \rightarrow \mathbb{R}$, where $f \in C^\infty(U \subseteq M \rightarrow \mathbb{R})$, is defined by $\phi_f(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)$.

The second equation of the first line is not of much practical use, but it helps formalize the precise way in which $x^i|_{\mathbf{p}}$ and $dx^i|_{\mathbf{p}}$ are “the same”: $dx^i|_{\mathbf{p}} = (\phi_{x^i})|_{\mathbf{p}}$.

Remark 8.40. (Interpretations of x^i).

Let M be a smooth n -manifold, and consider a smooth chart (U, \mathbf{x}) about $\mathbf{p} \in M$.

When considering the basis $\left\{ \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right\}$ for $T_{\mathbf{p}}(M)$, the x^i in the “denominator” is *not* a coordinate function of the smooth chart \mathbf{x} that is involved in the definition $\left(\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right)(f) = \frac{\partial(f \circ \mathbf{x}^{-1})}{\partial x^i} \Big|_{\mathbf{x}(\mathbf{p})}$. (Recall Definition [...]). In this context, the x^i in the “denominator” on the left hand side is simply notation that evokes the mental imagery of the meaning of the x^i on the right hand side (on the right hand side, the x^i in the “denominator” is used in the notation $\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} := \frac{\partial}{\partial \hat{\mathbf{e}}^i} \Big|_{\mathbf{p}}$ for directional derivatives which act on smooth functions defined on \mathbb{R}^n). The purpose of the function \mathbf{x} , from the smooth chart (U, \mathbf{x}) , is unrelated to this mental imagery; \mathbf{x} serves the different purpose of producing the coordinate representation $\tilde{\mathbf{F}}_{(U, \mathbf{x})} = (\mathbf{F} \circ \mathbf{x}^{-1}) : C^\infty(\mathbb{R}^n \rightarrow \mathbb{R}) \rightarrow \mathbb{R}$ of \mathbf{F} .

However, we have also seen that it is useful to use x^i to denote a coordinate function of \mathbf{x} when we are interested in the coordinates of $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ relative to $\left\{ \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right\}$, since $x^i(v_{\mathbf{p}})$ is the i th coordinate of $v_{\mathbf{p}}$ relative to $\left\{ \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right\}$.

A general rule is that when x^i appears in a “numerator” or “by itself”, then x^i is a coordinate function that is the argument of a directional derivative, and, when x^i appears in a “denominator”, it is because that “denominator” is part of the basis vector $\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}}$ of $T_{\mathbf{p}}(M)$. (In the special case of $M = \mathbb{R}^n$, then $\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}}$ is a directional derivative that acts on smooth functions defined on \mathbb{R}^n . In this special case, the “mental imagery” mentioned above has been realized, because the chart \mathbf{x} is the identity).

Theorem 8.41. (Change of coordinates for tangent vectors in terms of basis vectors of $T_{\mathbf{p}}(M)$).

Theorem 2.67 stated that if V is a finite-dimensional vector space with bases E and F , then

$$\mathbf{f}_i = \sum_{j=1}^n ([\mathbf{f}_i]_E)_j \mathbf{e}_j = \sum_{j=1}^n ([\mathbf{F}]_E)_{ji} \mathbf{e}_j.$$

Let M be a smooth n -manifold, and consider smooth charts (U, \mathbf{x}) and (V, \mathbf{y}) , where $\mathbf{p} \in U \cap V$. Let x^i and y^i denote the i th coordinate functions of \mathbf{x} and \mathbf{y} , respectively. Applying the above theorem to the vector space $T_{\mathbf{p}}(M)$ and its bases $E = \left\{ \left. \frac{\partial}{\partial x^j} \right|_{\mathbf{p}} \right\}_{j=1}^n$ and $F = \left\{ \left. \frac{\partial}{\partial y^j} \right|_{\mathbf{p}} \right\}_{j=1}^n$, we have

$$[\mathbf{F}]_E = \begin{pmatrix} \left[\left. \frac{\partial}{\partial y^1} \right|_{\mathbf{p}} \right]_F & \cdots & \left[\left. \frac{\partial}{\partial y^n} \right|_{\mathbf{p}} \right]_F \end{pmatrix} = \begin{pmatrix} \left. \frac{\partial x^1}{\partial y^1} \right|_{\mathbf{p}} & \cdots & \left. \frac{\partial x^1}{\partial y^n} \right|_{\mathbf{p}} \\ \vdots & & \vdots \\ \left. \frac{\partial x^n}{\partial y^1} \right|_{\mathbf{p}} & \cdots & \left. \frac{\partial x^n}{\partial y^n} \right|_{\mathbf{p}} \end{pmatrix} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}},$$

where $\left. \frac{\partial x^i}{\partial y^j} \right|_{\mathbf{p}} = \left(\left. \frac{\partial}{\partial y^j} \right|_{\mathbf{p}} \right)(x^i)$, and where x^i is the i th coordinate function of \mathbf{x} . The matrix $\frac{\partial \mathbf{x}}{\partial \mathbf{y}}$ is the Jacobian matrix described in Theorem [jacobian].

Applying the fact $\mathbf{f}_i = \sum_{j=1}^n ([\mathbf{F}]_E)_{ji} \mathbf{e}_j$ from above, we have

$$\boxed{\left. \frac{\partial}{\partial y^i} \right|_{\mathbf{p}} = \sum_{j=1}^n \left. \frac{\partial x^j}{\partial y^i} \right|_{\mathbf{p}} \left. \frac{\partial}{\partial x^j} \right|_{\mathbf{p}}.}$$

This change of basis equation strongly resembles the chain rule, and indeed simplifies to the chain rule when $M = \mathbb{R}^n$. In the general case when M is not necessarily $M = \mathbb{R}^n$, be sure to interpret the x^i 's and y^i 's as described in the previous remark.

8.6 Vector fields and covector fields

The tangent, cotangent, and tensor bundles

Definition 8.42. (Tangent bundle).

Let M be a smooth n -manifold. The *tangent bundle* of M is the set $T(M) := \bigsqcup_{\mathbf{p} \in M} T_{\mathbf{p}}(M)$, where \bigsqcup denotes a disjoint³ union.

Theorem 8.43. [Lee, p.66] (The tangent bundle is a smooth $2n$ -manifold).

For any smooth n -manifold M ; the tangent bundle $T(M)$ has a natural topology and smooth structure that make it into a $2n$ -dimensional smooth manifold.

Proof idea. Follow the special case of $M = \mathbb{R}^n$:

$$T(\mathbb{R}^n) = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} T_{\mathbf{p}}(\mathbb{R}^n) = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} \mathbb{R}_{\mathbf{p}}^n = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} (\mathbb{R}^n \times \{\mathbf{p}\}) = \mathbb{R}^n \times \mathbb{R}^n.$$

(Figure out why last equality works).

Definition 8.44. (Cotangent bundle).

Let M be a smooth n -manifold. The *cotangent bundle* of M is the set $T^*(M) := \bigsqcup_{\mathbf{p} \in M} T_{\mathbf{p}}^*(M)$, where \bigsqcup again denotes a disjoint union.

[cotangent bundle is a smooth $(2n?)$ manifold]

Definition 8.45. (Tensor bundle).

Let M be a smooth n -manifold. The $\binom{p}{q}$ *tensor bundle* of M is the set $T_q^p(T(M)) := \bigsqcup_{\mathbf{p} \in M} T_q^p(T_{\mathbf{p}}(M))$.
[is a manifold]

³Note that a disjoint union is *not* a “pairwise disjoint union”. We say $A = \bigsqcup_{\alpha} A_{\alpha}$ iff $A = \bigsqcup_{\alpha} A_{\alpha}$ and $\cap_{\alpha} A_{\alpha} = \emptyset$.

Remark 8.46. (Tangent bundle and cotangent bundle special cases of the tensor bundle).

Have $T_0^p(T(M)) = \bigsqcup_{\mathbf{p} \in M} (T_{\mathbf{p}}(M))^{\otimes p}$ and $T_q^0(T(M)) = \bigsqcup_{\mathbf{p} \in M} (T_{\mathbf{p}}^*(M))^{\otimes p}$. So in particular, $T_0^1(T(M)) = T(M)$, $T_1^0(T(M)) = T^*(M)$.

What about $T_0^0(M)$?

Definition 8.47. (Vector fields, covector fields, and $\binom{p}{q}$ tensor fields).

Let M be a smooth n -manifold WWBOC, and let $A \subseteq M$. A *vector field on A* is a continuous function $A \subseteq M \rightarrow T(M)$, a *covector field on A* is a continuous function $A \subseteq M \rightarrow T^*(M)$, and a $\binom{p}{q}$ *tensor field on A* is a continuous function $A \subseteq M \rightarrow T_q^p(T(M))$. *Smooth vector fields, smooth covector fields, and smooth tensor fields* are smooth such functions.

Frames and coframes

Definition 8.48. [Lee, p.178, 278] (Frames and coframes).

Let M be a smooth n -manifold WWBOC, and let (U, \mathbf{x}) be a smooth chart for M .

A *local frame on (U, \mathbf{x})* is a set $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$ of vector fields on U , where, at each $\mathbf{p} \in U \subseteq M$, the \mathbf{E}_i 's are linearly independent and span $T_{\mathbf{p}}(M)$. A local frame $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$ is a *smooth frame* iff each \mathbf{E}_i is smooth, and is a *global frame* iff $U = M$. A local frame is essentially a “basis” of vector fields.

A *local coframe on (U, \mathbf{x})* is a set of covector fields $\{\mathcal{E}^1, \dots, \mathcal{E}^n\}$ on U , where, at each $\mathbf{p} \in U \subseteq M$, the \mathcal{E}^i 's are linearly independent and span $T_{\mathbf{p}}^*(M)$. A local coframe $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$ is a *smooth coframe* iff each \mathbf{E}^i is smooth, and is a *global coframe* iff $U = M$. A local coframe is essentially a “basis” of covector fields.

Definition 8.49. [Lee, p.176, 278] (Coordinate vector fields and coordinate covector fields).

Let M be a smooth n -manifold WWBOC, and let (U, \mathbf{x}) be a smooth chart on M .

The smooth⁴ vector field defined by $\mathbf{p} \mapsto \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}}$ is denoted (in an abuse of notation) by $\frac{\partial}{\partial x^i}$ and is called the *i th local coordinate vector field (on (U, \mathbf{x}))*.

Recall that $\{dx_{\mathbf{p}}^i\}_{i=1}^n$ is the dual basis of $T_{\mathbf{p}}^*(M)$ induced by the basis $\{\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}}\}_{i=1}^n$ for $T_{\mathbf{p}}(M)$. The smooth⁵ covector field defined by $\mathbf{p} \mapsto dx^i|_{\mathbf{p}}$ is denoted (in an abuse of notation) by dx^i and is called the *i th local coordinate covector field (on (U, \mathbf{x}))*.

Definition 8.50. [Lee, p.178, 278] (Coordinate frames and coordinate coframes).

Let M be a smooth manifold WWBOC, and let (U, \mathbf{x}) be a smooth chart on M .

The smooth⁶ local frame $\{\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}\}$, where $\frac{\partial}{\partial x^i}$ denotes the i th local coordinate vector field, is called a *local coordinate frame (on (U, \mathbf{x}))*.

The smooth⁷ coframe $\{dx_1, \dots, dx_n\}$, where dx_i denotes the i th local coordinate covector field, is called the *local coordinate coframe (on (U, \mathbf{x}))*.

Definition 8.51. [Lee, p.278] (Duality of frames and coframes).

Let M be a smooth n -manifold WWBOC, and let (U, \mathbf{x}) be a smooth chart on M .

A local frame $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$ on (U, \mathbf{x}) and local coframe $\{\mathcal{E}^1, \dots, \mathcal{E}^n\}$ on (U, \mathbf{x}) are said to be *dual* to each other iff, at each $\mathbf{p} \in U$, the basis $\{\mathcal{E}^1|_{\mathbf{p}}, \dots, \mathcal{E}^n|_{\mathbf{p}}\}$ is the dual basis induced for $T_{\mathbf{p}}^*(M)$ induced by the basis $\{\mathbf{E}_1|_{\mathbf{p}}, \dots, \mathbf{E}_n|_{\mathbf{p}}\}$ for $T_{\mathbf{p}}(M)$.

As an example, $\frac{\partial}{\partial x^i}$ and dx^i are dual to each other.

Vector fields and covector fields in coordinates

Theorem 8.52. (Coordinates of frames and coframes).

Basically slightly generalize “coordinates of tangent vectors and dual vectors”

⁴The i th coordinate vector field is smooth because its component functions are constants.

⁵The i th coordinate covector field is smooth because its component functions are constants.

⁶The local coordinate frame is smooth because the i th local coordinate vector field is smooth.

⁷It can be shown that the local coordinate coframe is smooth because it is *dual* to the local coordinate frame, in the sense of the next definition.

“The components of a covector field ϕ at \mathbf{p} relative to $\{\lambda_1, \dots, \lambda_n\}$ are the covector fields $([\phi]_{\{\lambda_1, \dots, \lambda_n\}})_i$ defined by $([\phi]_{\{\lambda_1, \dots, \lambda_n\}})_i|_{\mathbf{p}} = \phi_{\mathbf{p}}\left(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right)$ [Lee, p.277] ”

Theorem 8.53. (Change of coordinates for frames in terms of coordinate frames).
is “in terms of coordinate frames” correct?

$$\frac{\partial}{\partial y^i} = \sum_{j=1}^n \frac{\partial x^j}{\partial y^i} \frac{\partial}{\partial x^j}$$

[Lee, p.275]

Theorem 8.54. (Differential of a function $U \subseteq M \rightarrow \mathbb{R}$ after identifying $T_{\mathbf{f}(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$).

Recall thm which says $df_{\mathbf{p}} = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}$. This implies

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x^i} dx^i$$

where d is the differential after identifying $T_{\mathbf{q}}(\mathbb{R}^n) \cong \mathbb{R}^n$.

Given a smooth chart (U, \mathbf{x}) on M , the differential of a smooth function $f : M \rightarrow \mathbb{R}$ is a smooth covector field. [Lee, p.281] .

8.7 Oriented manifolds and their oriented boundaries

Definition 8.55. [Lee, p.380] (Oriented manifolds).

Define pointwise orientation of a manifold in terms of orientation of its tangent spaces.

...

Define oriented manifold

Definition 8.56. [Lee, p.381, 382] (Orientation of a smooth chart on an oriented manifold).

oriented chart

Misc.

- $\dim(\partial M) = \dim(M) - 1$. We say ∂M has a *codimension* of 1.
- If M is a smooth manifold with boundary, then the boundary ∂M of M is a smooth manifold properly embedded in M . (A manifold N is said to be *embedded* in a manifold M iff there exists an embedding, which is an injective smooth immersion, $N \rightarrow M$. An *immersion* is a smooth function between manifolds whose derivative is everywhere injective. See Lee Theorem 5.11).
- “slice charts for embedded submanifolds”

Inward- and outward- pointing vectors

Definition 8.57. (Inward- and outward- pointing tangent vectors).

Let (U, \mathbf{x}) be a smooth chart on ∂M with $\mathbf{p} \in U$. We classify tangent vectors in $T_{\mathbf{p}}(M)$ as follows.

- A tangent vector is *inward-pointing* (on ∂M) iff it has positive- x^n component.
- A tangent vector is *tangent to* ∂M iff it has an x^n -component of zero.
- A tangent vector is *outward-pointing* (on ∂M) iff it has negative- x^n component, i.e., $\mathbf{v} \in T_{\mathbf{p}}(M)$ is outward pointing iff $-\mathbf{v}$ is inward pointing.

Example 8.58. (Inward- and outward- pointing tangent vectors).

Consider the unit disk D in \mathbb{R}^2 , and pick a point \mathbf{p} on the boundary. Any vector thought of as anchored at \mathbf{p} is either in $T_{\mathbf{p}}(D)$, which is a line, or in one of the two half-spaces resulting from the splitting of \mathbb{R}^2 by the line $T_{\mathbf{p}}(\partial D)$. Both halves of \mathbb{R}^2 are homeomorphic to half planes. One of these half planes contains $D - \{\mathbf{p}\}$ and the other one does not, so it makes sense to call vectors in the first half-plane “inward pointing” and vectors in the second half-plane “outward pointing.”

Lemma 8.59. (Existence of inward- and outward-pointing vector fields on M).

There exists a global smooth vector field on M whose restriction to ∂M is everywhere inward-pointing, and one whose restriction to ∂M is everywhere outward-pointing. (See Lee problem 8-4).

Theorem 8.60. (Induced orientation on the boundary).

Let ω be an orientation form on M in the actual function sense. Since there exists a vector field \mathbf{N} that is nowhere tangent to ∂M , then there is an induced *induced orientation form* ω' on the boundary (due to interior multiplication⁸) defined by $\omega'(\mathbf{v}_1, \dots, \mathbf{v}_{n-1}) = \omega(\mathbf{N}(\mathbf{p}), \mathbf{v}_1, \dots, \mathbf{v}_{n-1})$. (Lee Proposition 15.21). The orientation induced by ω' does not depend on the vector field \mathbf{N} that is nowhere tangent to ∂M . (Lee Proposition 15.24).

Proof. We need to show (1) that ω' is indeed an orientation form on ∂M and (2) that the orientation induced by ω' is independent of the choice of the nowhere tangent vector field \mathbf{N} .

1. We show $\omega' > 0$ on [tangent bundle of ∂M]; it suffices to show $\omega' \neq 0$ on [tangent bundle] because $\omega > 0$ on [tangent bundle of M].

Let $\{\mathbf{e}_1, \dots, \mathbf{e}_{n-1}\}$ be an ordered basis for $T_{\mathbf{p}}(\partial M)$. For all $\mathbf{p} \in M$, $\mathbf{N}(\mathbf{p})$ is not tangent to M , so $\{\mathbf{N}(\mathbf{p}), \mathbf{e}_1, \dots, \mathbf{e}_{n-1}\}$ is a basis for $T_{\mathbf{p}}(M)$, since $T_{\mathbf{p}}(M)$ is identifiable with \mathbb{R}^n . ω is nonvanishing, so it cannot be zero on any ordered basis. This means ω' nonvanishing as well.

2. Let \mathbf{N} and \mathbf{N}' be two vector fields that are both nowhere tangent to ∂M . We need to show that the ordered bases $E = \{\mathbf{N}(\mathbf{p}), \mathbf{v}_1, \dots, \mathbf{v}_{n-1}\}$ and $F = \{\mathbf{N}'(\mathbf{p}), \mathbf{v}_1, \dots, \mathbf{v}_{n-1}\}$ have the same orientation. To do so, we prove that the determinant of the change of basis matrix between the two ordered bases is positive.

\mathbf{N} and \mathbf{N}' are both outward-pointing, so the n th component of $\mathbf{N}(\mathbf{p})$ relative to E and the n th component of $\mathbf{N}'(\mathbf{p})$ relative to F are both negative; denote these n th components by $(\mathbf{N}(\mathbf{p}))_n$ and $(\mathbf{N}'(\mathbf{p}))_n$, respectively. Relative to the bases E, F , the change of basis matrix between E and F has a first column whose n th entry is $\frac{(\mathbf{N}'(\mathbf{p}))_n}{(\mathbf{N}(\mathbf{p}))_n}$, and for $i > 1$, the i th column of this matrix is $\hat{\mathbf{e}}_i$. The change of basis matrix is therefore upper triangular, so its determinant is the product of the diagonal entries, i.e., the determinant is $\frac{(\mathbf{N}'(\mathbf{p}))_n}{(\mathbf{N}(\mathbf{p}))_n} > 0$.

□

Theorem 8.61. [HH, p.542 - 544] (Oriented boundary of a k -parallelipiped).

Proof. Start with using induced boundary orientation from previous thm. Then use Hubbards' arg □

8.8 Riemannian manifolds

Definition 8.62. [Lee, p.327] (Riemannian manifold).

Let M be a smooth manifold WWBOC. A *Riemannian metric tensor* g on M is a smooth symmetric $\binom{0}{2}$ tensor field on M that is positive-definite at each point. Thus, g is an inner product on $T_{\mathbf{p}}(M)$ for each $\mathbf{p} \in M$.

A *Riemannian manifold WWBOC* is a tuple (M, g) , where M is a smooth manifold WWBOC and g is a Riemannian metric. An *oriented Riemannian manifold WWBOC* is a Riemannian manifold WWBOC that has been given an orientation.

⁸We could have put $\mathbf{N}(\mathbf{p})$ in any of ω 's n argument slots, but we chose to use the first. This choice corresponds to the operation called *interior multiplication*, which you can read about in Lee's book. See p. 358 and Corollary 14.21 on p. 362.

Theorem 8.63. • [Lee, p.329] “Every smooth manifold with or without boundary admits a Riemannian metric”. But there is no natural choice of metric

- [Lee, p.332] flat metrics; not every metric is flat

Definition 8.64. [Lee, p.389] (Riemannian volume form).

Let (M, g) be a smooth oriented Riemannian manifold, let $\{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ be an ordered⁹ orthonormal frame for M , and let $\{\mathbf{U}^1, \dots, \mathbf{U}^n\}$ be the induced ordered orthonormal coframe. The *Riemannian volume form* ω_g corresponding to $\{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ is defined to be

$$\omega_g := \mathbf{U}^1 \wedge \dots \wedge \mathbf{U}^n.$$

Theorem 8.65. [Lee, p.389-390] (Riemannian volume form in local coordinates).

Let (M, g) be a smooth oriented Riemannian manifold, let $\{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ be an ordered orthonormal frame for M , and let $\{\mathbf{U}^1, \dots, \mathbf{U}^n\}$ be the induced ordered orthonormal coframe. If (U, \mathbf{x}) is a smooth chart on M , then the Riemannian volume form is on U expressed as

$$\omega_g = \det \left(\frac{\partial \mathbf{U}_j}{\partial x^i} \right) dx^1 \wedge \dots \wedge dx^n.$$

Proof. Using Theorem 8.53, which expresses the change of coordinates for frames in terms of coordinate frames, we have

$$\frac{\partial}{\partial x^i} = \sum_{j=1}^n \frac{\partial \mathbf{U}_j}{\partial x^i} \frac{\partial}{\partial \mathbf{U}_j}.$$

Since $\omega_g := \mathbf{U}^1 \wedge \dots \wedge \mathbf{U}^n$ is a smooth differential n -form on U , then $\omega_g = f dx^1 \wedge \dots \wedge dx^n$ for some smooth $f : U \subseteq M \rightarrow \mathbb{R}^n$. We will show that $f = \det(g_{ij})$.

We have **explain the contraction**

$$f = C \left(\omega_g, \frac{\partial}{\partial x^1} \wedge \dots \wedge \frac{\partial}{\partial x^n} \right) = C(\mathbf{U}^1 \wedge \dots \wedge \mathbf{U}^n, \frac{\partial}{\partial x^1} \wedge \dots \wedge \frac{\partial}{\partial x^n}) = \det \left(\mathbf{U}^j \left(\frac{\partial}{\partial x^i} \right) \right).$$

The last equality follows from the fact that the pushforward on the top exterior power is multiplication by the determinant. (Use Theorem 5.37, where \mathbf{f} sends $\frac{\partial}{\partial x^i} \mapsto \mathbf{U}^j$).

By Theorem 8.52 [...] = $\mathbf{U}^j \left(\frac{\partial}{\partial x^i} \right)$, so

$$f = \det \left(\frac{\partial \mathbf{U}_j}{\partial x^i} \right).$$

□

Theorem 8.66. [Lee, p.389-390] (Riemannian volume form in local coordinates in terms of the Riemannian metric).

Let (M, g) be a smooth oriented Riemannian manifold, let $\{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ be an positively oriented orthonormal frame for M , and let $\{\mathbf{U}^1, \dots, \mathbf{U}^n\}$ be the induced ordered orthonormal coframe. If (U, \mathbf{x}) is a smooth chart on M , then we have $\det \left(\frac{\partial \mathbf{U}_j}{\partial x^i} \right) = \sqrt{\det(g_{ij})}$, so it follows from the previous theorem that the Riemannian volume form is expressed on U as

$$\omega_g = \sqrt{\det(g_{ij})} dx^1 \wedge \dots \wedge dx^n.$$

⁹We define an *ordered frame*, analogous to an ordered basis, to be a frame in which the order of the vector fields matter.

Proof. We show $\det\left(\frac{\partial \mathbf{u}_j}{\partial x^i}\right) = \sqrt{\det(g_{ij})}$. To do so, we compute $g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)$. Using the expression for $\frac{\partial}{\partial x^i}$ from the previous proof that was obtained by applying Theorem 8.53, we have

$$\begin{aligned} g_{ij} &= g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = g\left(\sum_{k=1}^n \frac{\partial \mathbf{u}_k}{\partial x^i} \frac{\partial}{\partial \mathbf{u}_k}, \sum_{\ell=1}^n \frac{\partial \mathbf{u}_\ell}{\partial x^j} \frac{\partial}{\partial \mathbf{u}_\ell}\right) = \sum_{k=1}^n \sum_{\ell=1}^n \frac{\partial \mathbf{u}_k}{\partial x^i} \frac{\partial \mathbf{u}_\ell}{\partial x^j} g\left(\frac{\partial}{\partial \mathbf{u}_k}, \frac{\partial}{\partial \mathbf{u}_\ell}\right) \\ &= \sum_{k=1}^n \sum_{\ell=1}^n \frac{\partial \mathbf{u}_k}{\partial x^i} \frac{\partial \mathbf{u}_\ell}{\partial x^j} \delta_\ell^k = \sum_{k=1}^n \frac{\partial \mathbf{u}_k}{\partial x^i} \frac{\partial \mathbf{u}_k}{\partial x^j}. \end{aligned}$$

(Note, we have $g\left(\frac{\partial}{\partial \mathbf{u}_k}, \frac{\partial}{\partial \mathbf{u}_\ell}\right) = \delta_\ell^k$ because the bases $\{\mathbf{U}_1|_{\mathbf{p}}, \dots, \mathbf{U}_n|_{\mathbf{p}}\}$ and $\{\mathbf{u}^1|_{\mathbf{p}}, \dots, \mathbf{u}^n|_{\mathbf{p}}\}$ are orthonormal).

This last sum is the ij entry of the matrix product $\left(\frac{\partial \mathbf{u}_j}{\partial x^i}\right)^\top \left(\frac{\partial \mathbf{u}_j}{\partial x^i}\right)$. Since $\det(\mathbf{A}^\top) = \det(\mathbf{A})$, then $\det(g_{ij}) = \det\left(\left(\frac{\partial \mathbf{u}_j}{\partial x^i}\right)^\top \left(\frac{\partial \mathbf{u}_j}{\partial x^i}\right)\right) = \det\left(\frac{\partial \mathbf{u}_j}{\partial x^i}\right)^2$, which means $\det\left(\frac{\partial \mathbf{u}_j}{\partial x^i}\right) = \pm \sqrt{\det(g_{ij})}$. The sign of the square root must be positive because, since $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$ is positively oriented, $\{\boldsymbol{\varepsilon}^1, \dots, \boldsymbol{\varepsilon}^n\}$ [must be] as well. \square

Differential forms on manifolds

9.1 Differential forms

Definition 9.1. [Lee, p.360] (Differential k -form).

Let M be a smooth n -manifold WWBOC.

The subset of $T_k^0(T(M))$ consisting of alternating tensors is $\Lambda^k(T^*(M)) = \Lambda^k(\bigsqcup_{\mathbf{p} \in M} T^*(M)) \cong \bigsqcup_{\mathbf{p} \in M} \Lambda^k(T_{\mathbf{p}}^*(M))$. Is $\binom{n}{k}$ dimensional, and has basis [...].

A *differential k -form on M* is a continuous function $\lambda^k(T^*(M)) \rightarrow M$. So, you might say that a differential k -form is a “alternating $\binom{0}{k}$ tensor field” (remember, all tensor fields are continuous maps). The vector space of differential k -forms on M is denoted $\Omega^k(M)$.

Theorem 9.2. [Lee, p.360] (Differential k -form expressed relative to a coordinate chart).

Let M be a smooth n -manifold. Given any smooth chart (U, \mathbf{x}) on M , where x^i is the i th component function of \mathbf{x} , a differential k -form ω can be expressed as

$$\omega = \sum_{1 \leq i_1 < \dots < i_k \leq n} f_{i_1 \dots i_k} dx^{i_1} \wedge \dots \wedge dx^{i_k},$$

where each $f_{i_1 \dots i_k} : U \rightarrow \mathbb{R}$.

Theorem 9.3. [Lee, p.360] (Smoothness of a differential form).

When thought of as a tensor field, a differential k -form is smooth iff its component functions are smooth.

Remark 9.4. (Differential 0-forms are functions).

9.2 Integration of differential forms on manifolds

This section follows Ch. 16 of [Lee].

There is not always a integrate *functions* on manifold in a coordinate-independent way. For example, if B is the unit ball in \mathbb{R}^n and $f \equiv 1$, then $\int_B f dV = \text{vol}(B)$; using the change of coordinates $\mathbf{x} \mapsto c\mathbf{x}$, we see the integral changes to $c \text{vol}(B)$. (Or $c^n \text{vol}(B)$)?

Differential forms will be the objects whose integrals on manifolds are coordinate independent.

We will consider only compactly supported differential forms; this means the integrals we will consider are analogous to “proper” integrals of functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Recall, we say that an integral of a smooth function $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ is *improper* iff D does not contain all of its limit points (so D wouldn’t be closed, and thus wouldn’t be compact due to Theorem [...]) or is unbounded (so D wouldn’t be compact due to Theorem [...]).

[Lee, p.422] We can define a coordinate independent integral of a smooth function $\mathbf{F} : M \rightarrow \mathbb{R}$ a Riemannian manifold (M, g) : define $\int_M f := \int_M f \omega_g$, where ω_g is the Riemannian volume form induced by g . Riemannian volume form is denoted ds , dA , or dV . (But the d next to s , A , or V does not indicate that ω_g can be interpreted as an exterior derivative!)

In order to define a coordinate-independent notion of integrating differential forms on manifolds, we will first need to compute the pullback of a differential k -form.

Definition 9.5. [GP74, p.163-165] (Pullback of a differential k -form).

Let M and N be n -smooth manifolds, let (U, \mathbf{x}) and (V, \mathbf{y}) be smooth charts on M and N , respectively, and let $\mathbf{F} : U \subseteq M \rightarrow V \subseteq N$ be a smooth function. Consider a differential k -form $\omega \in \Omega^k(V)$, $k \leq n$, on $V \subseteq N$.

In Theorem [...], we showed how to pull back an element of $\Lambda^k(W)$ to obtain an element of $\Lambda^k(V)$, where V and W are vector spaces. Since a differential form evaluated at a point yields an element of an exterior power, we can pull back differential forms by using the pullback map on exterior powers.

We define the *pullback* $\Omega^k \mathbf{F}^* : \Omega^k(V) \rightarrow \Omega^k(U)$ of a differential k -form ω on V by

$$\left((\Omega^k \mathbf{F}^*)(\omega) \right)_{\mathbf{p}} := \left(\Lambda^k(d\mathbf{F}_{\mathbf{p}})^* \right) (\omega_{\mathbf{F}(\mathbf{p})}).$$

On the right hand side, $(d\mathbf{F}_{\mathbf{p}})^* : T_{\mathbf{F}(\mathbf{p})}(V)^* \rightarrow T_{\mathbf{p}}(U)^*$ is the dual of the differential $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(U) \rightarrow T_{\mathbf{F}(\mathbf{p})}(V)$, which is a linear map, and $\Lambda^k(d\mathbf{F}_{\mathbf{p}})^*$ is the pullback from the k th exterior power $\Lambda^k(T_{\mathbf{F}(\mathbf{p})}(V))$ to the k th exterior power $\Lambda^k(T_{\mathbf{p}}(U))$. (For the exact meaning of $\Lambda^k(d\mathbf{F}_{\mathbf{p}})^*$, recall Definition 5.24).

Theorem 9.6. (Basic properties of pullbacks of differential forms).

Let M and N be smooth n -manifolds, let $U \subseteq M$ be open, and consider a smooth function $\mathbf{F} : U \subseteq M \rightarrow N$.

1. For any smooth function $f : \mathbb{R}^n \rightarrow U \subseteq M$, we have

$$\Omega^k \mathbf{F}^*(f) = (d\mathbf{F})^*(f),$$

where $(d\mathbf{F})^*$ is the dual of the linear map $d\mathbf{F}$.

2. If ω and η are differential k -forms, $k \leq n$, on \mathbb{R}^n , then

$$\Omega^k \mathbf{F}^*(\omega \wedge \eta) = \Omega^k \mathbf{F}^*(\omega) \wedge \Omega^k \mathbf{F}^*(\eta).$$

- 3.

$$(\Omega^k \mathbf{F})^* = \Omega^k \mathbf{F}^*,$$

where $\Omega^k \mathbf{F}$ is defined “pointwise” in the way you’d expect, $(\Omega^k \mathbf{F})^*|_{\mathbf{p}} := ((\Omega^k \mathbf{F})_{\mathbf{p}})^*$.

Proof. All items follow straightforwardly from the definition of $\Omega^k \mathbf{F}^*$. □

Theorem 9.7. (The differential commutes with the pullback by $d\mathbf{F}_{\mathbf{p}}$ of a diffeomorphism).

use notation $\mathbf{F}_{\mathbf{p}}$ and $d\mathbf{F}_{\mathbf{p}}$?

Let M and N be smooth n -manifolds, let $U \subseteq M$ be open, and consider a smooth function $\mathbf{F} : U \subseteq M \rightarrow N$. Additionally, let $\mathbf{p} \in U \subseteq M$, and let $d_{\mathbf{p}}$ denote the map $f \mapsto d_{\mathbf{p}} f$. Then for any smooth function $f : \mathbb{R}^n \rightarrow U \subseteq M$, we have

$$\Omega^k \mathbf{F}^* \circ d_{\mathbf{p}} = d_{\mathbf{p}} \circ \Omega^k \mathbf{F}^* \iff \Omega^k \mathbf{F}^*(d_{\mathbf{p}} f) = d_{\mathbf{p}}(\Omega^k \mathbf{F}^*(f)).$$

Here, the differential d can be interpreted in both of the ways we have mentioned earlier. (That is, d can be considered to be the differential which results from identifying $T_{\mathbf{q}}(\mathbb{R}) \cong \mathbb{R}$ or the differential in which this identification is not performed, and $T_{\mathbf{p}}(\mathbb{R})$ is “left alone” (recall Theorem 8.34 and Definition 8.32, respectively).

Proof. For this proof, we will assume that d is the differential in which $T_{\mathbf{q}}(\mathbb{R}) \cong \mathbb{R}$ is “left alone”, so that $df_{\mathbf{p}}$ acts on a smooth function $u_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by $df_{\mathbf{p}}(u_{\mathbf{p}}) = u_{\mathbf{p}}(f)$ (recall Theorem 8.34). (After we have proven the theorem for this interpretation of d , the theorem holds for when d is obtained by identifying $T_{\mathbf{q}}(\mathbb{R}) \cong \mathbb{R}$ precisely because we have an isomorphism $T_{\mathbf{q}}(\mathbb{R}) \cong \mathbb{R}$).

Let $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. We will show $\Omega^k \mathbf{F}^*(df_{\mathbf{p}})(v_{\mathbf{p}}) = d(\Omega^k \mathbf{F}^*(f))_{\mathbf{p}}$.

By property (1) of the previous theorem, $\Omega^k \mathbf{F}^*(df_{\mathbf{p}})(v_{\mathbf{p}}) = (d\mathbf{F}_{\mathbf{p}})^*(df_{\mathbf{p}})(v_{\mathbf{p}})$. Then $(d\mathbf{F}_{\mathbf{p}})^*(df_{\mathbf{p}})(v_{\mathbf{p}}) = (df_{\mathbf{p}} \circ d\mathbf{F}_{\mathbf{p}})(v_{\mathbf{p}}) = df_{\mathbf{p}}(d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}}))$. Since $df_{\mathbf{p}}(u_{\mathbf{p}}) = u_{\mathbf{p}}(f)$, we use $u_{\mathbf{p}} = d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})$ to obtain $df_{\mathbf{p}}(d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})) = (d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}}))(f)$. By the definition of the differential $d\mathbf{F}_{\mathbf{p}}$ (recall Definition 8.32), $(d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}}))(f) = v_{\mathbf{p}}(f \circ d\mathbf{F}_{\mathbf{p}})$. Then $v_{\mathbf{p}}(f \circ d\mathbf{F}_{\mathbf{p}}) = v_{\mathbf{p}}((d\mathbf{F}_{\mathbf{p}})^*(f)) = v_{\mathbf{p}}(\Omega^k \mathbf{F}^*(f))$. Lastly, we notice that because $v_{\mathbf{p}}(g) = dg_{\mathbf{p}}(v_{\mathbf{p}})$, we use $g = \Omega^k \mathbf{F}^*(f)$ to obtain $v_{\mathbf{p}}(\Omega^k \mathbf{F}^*(f)) = d(\Omega^k \mathbf{F}^*(f))_{\mathbf{p}}(v_{\mathbf{p}})$. \square

Theorem 9.8. (Pullback of a top degree differential form).

Consider the same situation as described in the previous definition, except now suppose that $\omega \in \Omega^n(V)$ is a differential n -form, i.e., ω is a *top degree* differential form on V . (Recall $\dim(M) = \dim(N) = n$). We can write ω as

$$\omega = f \, dx^1 \wedge \dots \wedge dx^n,$$

where x^i is the i th coordinate function of \mathbf{x} (recall [...]).

Since the pullback on the top exterior power (here, the top exterior power in question is $\Lambda^n T_{\mathbf{p}}(V)$) is multiplication by the determinant, we have

$$\left((\Omega^k \mathbf{F}^*)(\omega) \right)_{\mathbf{p}} = \det((d\mathbf{F}_{\mathbf{p}})^* \mathbf{F}^*(f)(\mathbf{p})) \left((d\mathbf{F}_{\mathbf{p}})^*(dx^1)(\mathbf{F}(\mathbf{p})) \right) \wedge \dots \wedge \left((d\mathbf{F}_{\mathbf{p}})^*(dx^n)(\mathbf{F}(\mathbf{p})) \right).$$

(On the right hand side, $(d\mathbf{F}_{\mathbf{p}})^*(dx^i)(\mathbf{F}(\mathbf{p}))$ denotes the function $(d\mathbf{F}_{\mathbf{p}})^*(dx^i)$ evaluated on $\mathbf{F}(\mathbf{p})$).

If y^i is the i th coordinate function of \mathbf{y} , then, since $\mathbf{F} : U \subseteq M \rightarrow V \subseteq N$ **we have** $dy^i|_{\mathbf{p}} = (d\mathbf{F}_{\mathbf{p}})^*(dx^i)(\mathbf{F}(\mathbf{p}))$. Using this with the previous two properties of the pullback of a differential form, we have

$$\left((\Omega^k \mathbf{F}^*)(\omega) \right)_{\mathbf{p}} = \det(d\mathbf{F}_{\mathbf{p}})(f \circ \mathbf{F})(\mathbf{p}) \, dy^1|_{\mathbf{p}} \wedge \dots \wedge dy^n|_{\mathbf{p}}.$$

Suppressing dependence on \mathbf{p} , we have

$$(\Omega^k \mathbf{F}^*)(\omega) = \det(d\mathbf{F})(f \circ \mathbf{F}) \, dy^1 \wedge \dots \wedge dy^n.$$

Here, $d\mathbf{F}$ denotes the function which sends $\mathbf{p} \mapsto d\mathbf{F}_{\mathbf{p}}$.

Using the convention of denoting $\mathbf{F}^* := \overline{\Omega^k \mathbf{F}^*}$, the above is restated as

$$(\Omega^k \mathbf{F}^*)(\omega) = \det(d\mathbf{F})(f \circ \mathbf{F}) \, dy^1 \wedge \dots \wedge dy^n.$$

That is,

$$\boxed{(\Omega^k \mathbf{F}^*)(f \, dx^1 \wedge \dots \wedge dx^n) = \det(d\mathbf{F})(f \circ \mathbf{F}) \, dy^1 \wedge \dots \wedge dy^n.}$$

Remark 9.9. (Star notation for the pullback of a differential form).

In Remark 5.25, we introduced the practice of using \mathbf{F}^* to denote a pullback. From now on, we denote the pullback $\Omega^k \mathbf{F}^* : T_{\mathbf{F}(\mathbf{p})}(V \subseteq N) \rightarrow T_{\mathbf{p}}(U \subseteq M)$ by $\mathbf{F}^* : T_{\mathbf{F}(\mathbf{p})}(V \subseteq N) \rightarrow T_{\mathbf{p}}(U \subseteq M)$.

Definition 9.10. [Lee, p.653] (Domain of integration in \mathbb{R}^n). A *domain of integration* in \mathbb{R}^n is a bounded subset of \mathbb{R}^n whose boundary has an n -dimensional measure of zero.

Definition 9.11. (Support of a function on a manifold).

Let M be a (not necessarily smooth) manifold and $\mathbf{F} : M \rightarrow \mathbb{R}^n$. The *support* of \mathbf{F} is defined to be the closure of the set of points where \mathbf{F} is nonzero, $\text{supp}(\mathbf{F}) := \text{cl}(M - \mathbf{F}^{-1}(\mathbf{0}))$. Iff $\text{supp}(\mathbf{F}) \subseteq A$, then we say \mathbf{F} is *supported in* A .

We say \mathbf{F} is *compactly supported* iff $\text{supp}(\mathbf{F})$ is compact.

Can similarly (by extension?) define the support of a differential k -form.

Derivation 9.12. [GP74, p.166] [Lee, p.403] (Change of variables theorem in light of the pullback).

The change of variables theorem says that if D and E are open domains of integration in \mathbb{R}^n or \mathbb{H}^n and $\mathbf{F} : \text{cl}(D) \rightarrow \text{cl}(E)$ is a smooth map whose restriction $\mathbf{F}|_D : D \rightarrow E$ is a diffeomorphism, then, for every continuous function $f : \text{cl}(E) \rightarrow \mathbb{R}$, we have

$$\int_D f = \int_E (f \circ \mathbf{F}) |\det(d\mathbf{F})|.$$

Here, $\det(d\mathbf{F})$ denotes the function [...]

Notice that, when $|\det(d\mathbf{F})| = \det(d\mathbf{F})$, i.e., when $\det(d\mathbf{F}) > 0$, then the integrand of the right hand side is exactly the pullback of $f dy^1 \wedge \dots \wedge dy^k$. When $\det(d\mathbf{F}) < 0$, the integrand of the right hand side is the negation of this pullback. So, *in the case that \mathbf{F} is orientation-preserving or orientation-reversing*, the change of variables theorem can be restated as

$$\int_D f = \begin{cases} \int_E \mathbf{F}^*(f dy^1 \wedge \dots \wedge dy^n) & \mathbf{F} \text{ is orientation-preserving} \\ -\int_E \mathbf{F}^*(f dy^1 \wedge \dots \wedge dy^n) & \mathbf{F} \text{ is orientation-reversing} \end{cases}.$$

(It is possible for \mathbf{F} to be neither orientation-preserving nor orientation-reversing. In this case the integral of the pullback over V is likely unrelated to the integral of \mathbf{F} over U).

Definition 9.13. (Integral of a differential form on a domain of integration in \mathbb{R}^n or \mathbb{H}^n).

Let D be a domain of integration in \mathbb{R}^n or \mathbb{H}^n . Due our discovery in the previous derivation, we define the *integral of a compactly supported n -form $\omega = f dx^1 \wedge \dots \wedge dx^n$ over D* , to be

$$\boxed{\int_D \omega = \int_D f dx^1 \wedge \dots \wedge dx^n := \int_D f}$$

This definition us to concisely state the change of variables theorem by involving the integrals and pullbacks of differential forms.

Theorem 9.14. (Change of variables theorem for differential forms on open domains of integration).

Let D and E be open domains of integration in \mathbb{R}^n or \mathbb{H}^n , and let ω be a compactly supported differential n -form on $\text{cl}(E)$. Given the above definition, the restatement of the change of variables theorem presented at the end of the previous derivation is further restated as

$$\boxed{\int_D \omega = \begin{cases} \int_E \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-preserving} \\ -\int_E \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-reversing} \end{cases}}$$

This most recent boxed equation is not merely a restatement of the change of variables theorem, but a generalization. Previously, we only had a change of variables theorem for real-valued functions defined on open domains of integration in \mathbb{R}^n or \mathbb{H}^n ; now, we have a change of variables theorem for differential forms defined on open domains of integration in \mathbb{R}^n or \mathbb{H}^n .

Remark 9.15. (The meaning of $dx_1 \dots dx_n$).

We should note that our recent definition of the integral of a differential form gives meaning to the $dx_1 \dots dx_n$ that is used as a placeholder in an integral. If we use the convention of writing the placeholder $dx_1 \dots dx_n$ after the integrand, so that

$$\int_D f = \int_D f \, dx_1 \dots dx_n,$$

then the definition of the integral of a differential n -form on an open domain of integration becomes

$$\int_D f \, dx_1 \wedge \dots \wedge dx_n := \int_D f \, dx^1 \dots dx^n.$$

In some sense, the placeholder $dx_1 \dots dx_n$ is “secretly” $dx^1 \wedge \dots \wedge dx^n$. So, while the definition technically defines the left hand side in terms of the right hand side, you might think of it as giving algebraic meaning to the old placeholder notation of the right hand side.

Remark 9.16. Could discover def of differential forms by starting with the change of variables theorem and trying to contrive the notion of a pullback

At this point, we know how to integrate differential forms over open domains of integration in \mathbb{R}^n or \mathbb{H}^n . As a stepping stone to defining integration on manifolds, we now define the notion of integrating differential forms over open subsets of \mathbb{R}^n or \mathbb{H}^n .

Definition 9.17. [Lee, p.404] (Integral of a differential n -form over an open subset of \mathbb{R}^n or \mathbb{H}^n).

Let U be an open subset of \mathbb{R}^n or \mathbb{H}^n . We define the *integral of a differential n -form ω that is compactly supported on U* to be

$$\int_U \omega := \int_D \omega,$$

where D is any domain of integration in \mathbb{R}^n or \mathbb{H}^n containing $\text{supp}(\omega)$. This definition does not depend on the choice of D ; see [Lee, p. 403] for the details. The right hand side is interpreted with the definition of the previous derivation.

We now state a slight generalization of the change of variables theorem that relies on the previous definition.

Theorem 9.18. [Lee, p.404] (Change of variables theorem for differential forms on open subsets of \mathbb{R}^n or \mathbb{H}^n).

Let U and V be open subsets of \mathbb{R}^n or \mathbb{H}^n . If $\mathbf{F} : U \rightarrow V$ is a diffeomorphism and $\omega = f \, dx^1 \wedge \dots \wedge dx^n$ is a compactly supported differential n -form on V , then

$$\int_U \omega = \begin{cases} \int_V \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-preserving} \\ - \int_V \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-reversing} \end{cases}$$

As this is our final generalization of the change of variables theorem, it’s worth examining this theorem after expanding out the pullback $\mathbf{F}^*(\omega) = \mathbf{F}^*(f \, dx^1 \wedge \dots \wedge dx^n)$:

$$\int_U f \, dx^1 \wedge \dots \wedge dx^n = \pm \int_V f \circ d\mathbf{F} \, (dx^1 \circ d\mathbf{F}) \wedge \dots \wedge (dx^n \circ d\mathbf{F}),$$

where $d\mathbf{F}$ denotes the map $\mathbf{q} \mapsto d\mathbf{F}|_{\mathbf{q}}$, and where the \pm sign depends on whether \mathbf{F} is orientation-preserving or orientation-reversing.

Proof. The proof is a matter of using technical properties of diffeomorphisms. See [Lee, p. 404]. \square

As promised, we are now ready to define integration on smooth manifolds.

Definition 9.19. [Lee, p.404] (Integral of a differential n -form that is compactly supported in a single chart over a smooth n -manifold).

Let M be an oriented smooth n -manifold with or without boundary, let (U, \mathbf{x}) be a smooth oriented chart on M , and let ω be a differential n -form on M with compact support in U . We define the *integral of ω over M* to be

$$\int_M \omega := \begin{cases} \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega) & (U, \mathbf{x}) \text{ is positively oriented} \\ - \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega) & (U, \mathbf{x}) \text{ is negatively oriented} \end{cases}.$$

(Recall from Definition 8.56 that a smooth chart is positively oriented iff ... and ...).

Here, have used the pullback $(\mathbf{x}^{-1})^*$ to produce a differential n -form $(\mathbf{x}^{-1})^*(\omega)$ that is compactly supported on the open subset $\mathbf{x}(U)$ of \mathbb{R}^n or \mathbb{H}^n . The previous definition allows us to evaluate the integral of such a differential n -form, which the right hand side.

Note, this definition does not depend on the smooth chart (U, \mathbf{x}) for which $\text{supp}(\omega) \subseteq U$; see [Lee, p. 404] for the proof of this.

The previous theorem has sound algebraic intuition, but we have not yet seen any geometric intuition. To find geometric insight, we consider following special case.

Theorem 9.20. (Integral of a compactly supported differential n -form on a submanifold of \mathbb{R}^n)

Let M be a smooth submanifold of \mathbb{R}^n , let (M, \mathbf{x}) be a smooth positively oriented global chart on M , and consider a smooth differential n -form $\omega = f dx^1 \wedge \dots dx^n$ that is compactly supported in M . By the previous definition,

$$\begin{aligned} \int_U \omega &= \int_U f dx^1 \wedge \dots dx^n = \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(f dx^1 \wedge \dots dx^n) \\ &= \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(f) (\mathbf{x}^{-1})^*(dx^1) \wedge \dots \wedge (\mathbf{x}^{-1})^*(dx^n). \end{aligned}$$

Recall that $(\mathbf{x}^{-1})^*$ is not the dual transformation¹ of \mathbf{x}^{-1} , but the pullback by the differential $d\mathbf{x}^{-1}$ (recall Theorem [...]). So the above is

$$\int_U \omega = \int_{\mathbf{x}(U)} (d\mathbf{x}^{-1})^*(f) (d\mathbf{x}^{-1})^*(dx^1) \wedge \dots \wedge (d\mathbf{x}^{-1})^*(dx^n).$$

Here, $(d\mathbf{x}^{-1})^*$ denotes the map $\mathbf{q} \mapsto (d\mathbf{x}_{\mathbf{q}}^{-1})^*$, where $(d\mathbf{x}_{\mathbf{q}}^{-1})^*$ is the dual of the linear map $d\mathbf{x}_{\mathbf{q}}^{-1} : T_{\mathbf{q}}(\mathbb{R}^n) \rightarrow T_{\mathbf{x}^{-1}(\mathbf{q})}(M)$. Evaluating $(d\mathbf{x}_{\mathbf{q}}^{-1})^*$ on a function is the same as precomposing by $d\mathbf{x}_{\mathbf{q}}^{-1}$, so we have

$$\int_U \omega = \int_{\mathbf{x}(U)} f \circ d\mathbf{x}^{-1} (dx^1 \circ d\mathbf{x}^{-1}) \wedge \dots \wedge (dx^n \circ d\mathbf{x}^{-1}).$$

same (convention?) as before, so $dy^i|_{\mathbf{q}} = (d\mathbf{x}_{\mathbf{p}}^{-1})^*(dx^i)|_{\mathbf{q}}$, where $\mathbf{x}^{-1}(\mathbf{q}) = \mathbf{p}$. Suppressing dependence on \mathbf{p}, \mathbf{q} , get $dy^i := dx^i \circ d\mathbf{x}^{-1}$

$$\int_U \omega = \int_{\mathbf{x}(U)} f \circ d\mathbf{x}^{-1} dy^1 \wedge \dots \wedge dy^n.$$

The right hand side is an integral over a subset of \mathbb{R}^n . We know that to compute such an integral, we “erase the wedges”. Thus

¹This would not even make sense, because a function must be linear in order for us to take its dual. In general, \mathbf{x}^{-1} is not even a map of vector spaces.

$$\int_U \omega = \int_{\mathbf{x}(U)} f \circ d\mathbf{x}^{-1},$$

where the right hand side is now a “standard” integral of a real valued function.

Thus far, the steps we have taken have just been applications of definitions. We will make use of a new observation now. Define the *contraction* C of *antisymmetric tensor fields* on elementary tensor fields by

$$C\left(dy^{i_1} \wedge \dots \wedge dy^{i_k}, \frac{\partial}{\partial y^{j_1}} \wedge \dots \wedge \frac{\partial}{\partial y^{j_k}}\right) = dy^{j_1} \left(\frac{\partial}{\partial y^{i_1}}\right) \dots dy^{j_k} \left(\frac{\partial}{\partial y^{i_k}}\right),$$

and extend the definition of C with linearity and the antisymmetry of \wedge . (Recall Definition 4.42 for the earlier notions of tensor contraction).

Since $dy^i \left(\frac{\partial}{\partial y^i}\right) = 1$, our most recent “standard” integral is the same as

$$\begin{aligned} \int_{\mathbf{x}(U)} f \circ d\mathbf{x}^{-1} &= \int_{\mathbf{x}(U)} C\left(f \circ d\mathbf{x}^{-1} dy^1 \wedge \dots \wedge dy^n, \frac{\partial}{\partial y^1} \wedge \dots \wedge \frac{\partial}{\partial y^n}\right) \\ &= \int_{\mathbf{x}(U)} C(\omega, \frac{\partial}{\partial y^1} \wedge \dots \wedge \frac{\partial}{\partial y^n}) \end{aligned}$$

is this something like the following?

$$\int_{\mathbf{x}(U)} C\left((\mathbf{x}^{-1})^*(\omega), (\mathbf{x}^{-1})^*\left(\frac{\partial}{\partial x^1} \wedge \dots \wedge \frac{\partial}{\partial x^n}\right)\right)$$

Theorem 9.21. (Integral of a differential n -form over a smooth chart is “linear” with respect to the region of integration).

Let M be a smooth n -manifold, let (U, \mathbf{x}) be a smooth oriented chart on M , and consider a smooth differential n -form ω that is compactly supported in $\bigcup_{i=1}^k U_i$, where each $U_i \subseteq U$, and where $U_i \cap U_j$ has measure zero² for all $i \neq j$. Then

$$\int_U \omega = \sum_{i=1}^k \int_{U_i} \omega$$

Proof. By definition,

$$\int_U \omega = \pm \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega),$$

where the \pm sign depends on the orientation of (U, \mathbf{x}) . Since $\text{supp}(\omega) \subseteq \bigcup_{i=1}^k U_i$, then

Since $\text{supp}(\omega) \subseteq \bigcup_{i=1}^k U_i$, then $\text{supp}((\mathbf{x}^{-1})^*(\omega)) \subseteq \bigcup_{i=1}^k \mathbf{x}(U_i)$. Additionally, since \mathbf{x} is a smooth map, then $U_i \cap U_j$ having measure zero for all $i \neq j$ implies³ that $\mathbf{x}(U_i) \cap \mathbf{x}(U_j)$ has measure zero for all $i \neq j$. Thus, recalling Theorem 6.20, we see $\mathbf{x}(U_1), \dots, \mathbf{x}(U_k)$ satisfy the conditions that are required to “break up” an integral. Applying this same theorem, we have

$$\pm \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega) = \sum_{i=1}^k \pm \int_{\mathbf{x}(U_i)} (\mathbf{x}^{-1})^*(\omega) = \sum_{i=1}^k \int_{U_i} \omega,$$

as claimed. □

²Informally, a subset of \mathbb{R}^n has *measure zero* iff its volume is zero.

³*Sard’s theorem* is what is used to formally prove this implication.

The previous theorem is generalized by the following definition.

Definition 9.22. [Lee, p.408] (Integral of a differential n -form on a smooth n -manifold).

Let M be an oriented smooth n -manifold with or without corners and let ω be a compactly supported n -form on M . Since $\text{supp}(\omega)$ is compact, there is a finite collection $\{(U_i, \mathbf{x}_i)\}_{i=1}^k$ of charts on M for which $\{U_i\}_{i=1}^k$ is a pairwise nonoverlapping **use a different word- be more precise (pairwise disjoint and boundaries have measure zero)** open cover of $\text{supp}(\omega)$.

We define *the integral of ω over M to be*

$$\int_M \omega := \sum_{i=1}^k \int_{U_i} \omega$$

Note that each integral in the sum on the right hand side is interpreted with the previous definition.

prove that doesn't depend on open cover chosen; probably is a result of the nonoverlappingness

9.3 The exterior derivative

Definition 9.23. [HH, p.545] (The exterior derivative).

Recall defn of div and curl

If V is a compact region in \mathbb{R}^3 with volume $|V|$, then the divergence div of a smooth $[?]$ vector field $\mathbf{F} : V \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is defined as flux per “infinitesimal” volume:

$$\text{div}(\mathbf{F}) := \lim_{|V| \rightarrow 0} \frac{1}{|V|} \int_{\partial V} \mathbf{F} \cdot \hat{\mathbf{n}} dS.$$

[use \mathbf{V} for vector field?]

Similarly, if C is a closed simple curve in \mathbb{R}^3 , $A(C)$ is the oriented area enclosed by C , and $\hat{\mathbf{n}}$ is the positively oriented unit vector that is normal to $A(C)$, then $\text{curl}(\mathbf{F})$ is defined as work per “infinitesimal” surface area:

$$\text{curl}(\mathbf{F}) \cdot \hat{\mathbf{n}} := \lim_{A(C) \rightarrow 0} \frac{1}{A(C)} \int_C \mathbf{F} \cdot d\mathbf{r} = \lim_{A(C) \rightarrow 0} \frac{1}{A(C)} \int_C \mathbf{F} \cdot \frac{d\mathbf{r}}{dt} dt.$$

$$(\mathbf{v}_1|_{\mathbf{p}}, \dots, \mathbf{v}_{k+1}|_{\mathbf{p}}) \mapsto P_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1})$$

Let M be a smooth n -manifold, let (U, \mathbf{x}) be a smooth chart on M , and let ω be a smooth differential k -form with compact support in U . Let $P_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1})$ denote the subset of $T_{\mathbf{p}}(U)$ spanned by $\mathbf{v}_1|_{\mathbf{p}}, \dots, \mathbf{v}_{k+1}|_{\mathbf{p}} \in T_{\mathbf{p}}(U)$. We call $P_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1})$ a $(k+1)$ -parallelapiped.

We define the *exterior derivative* $d\omega$ of ω to be the differential $(k+1)$ -form defined at $\mathbf{p} \in M$ by

$$d\omega_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1}) := \lim_{h \rightarrow 0} \frac{1}{h^{k+1}} \int_{\partial P_{\mathbf{p}}(h\mathbf{v}_1, \dots, h\mathbf{v}_{k+1})} \omega.$$

In words, the exterior derivative $d\omega$ is evaluated at \mathbf{p} on $k+1$ tangent vectors $\mathbf{v}_1|_{\mathbf{p}}, \dots, \mathbf{v}_{k+1}|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by considering the “infinitesimally small” $(k+1)$ -parallelapiped in $T_{\mathbf{p}}(M)$ spanned by these vectors, and then integrating ω over the faces of the $(k+1)$ -parallelapiped, **each of which is an “infinitesimally small” k -parallelapiped.**

As is noted in [HH, p. 545], it is not immediately obvious why the limit in the definition of the exterior derivative exists. Since the integral in the limit can be broken up into a sum of integrals over the faces of the $(k+1)$ -parallelapiped, and as the i th edge of the $(k+1)$ -parallelapiped is $h\mathbf{v}_i|_{\mathbf{p}}$, it seems that each integral in the sum is dominated by $h^{(k+1)-1} = h^k$. The apparent problem comes

when we move $\frac{1}{h^{k+1}}$ inside the limit, so that each integral is now dominated by $\frac{h^k}{h^{k+1}} = \frac{1}{h}$. This seems problematic because $\lim_{h \rightarrow 0} \frac{1}{h}$ does not exist. We will see in the proof below that the limit *does* exist, because the faces come in pairs with opposite orientation, which causes terms of order h^k from each pair to cancel, [leaving something of order h^{k+1} .]

Theorem 9.24. [HH, p.652 - 655] (Computing the exterior derivative).

Let M be a smooth n -dimensional manifold with or without boundary or corners, and let (U, \mathbf{x}) be a smooth chart on M , where x^i is the i th component function of \mathbf{x} .

For any smooth function $f : U \rightarrow \mathbb{R}$,

1. The exterior derivative df of f is equal to the differential df of f (where the differential of f is the differential described by Theorem [...]).
2. $d(f \, dx^{i_1} \wedge \dots \wedge dx^{i_k}) = df \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k}$.

Notice that (1) implies that the exterior derivative of any constant differential form is the zero differential form $\mathbf{0}$.

Proof.

1. $df(P_{\mathbf{p}}(\mathbf{v})) = \lim_{h \rightarrow 0} \frac{1}{h} (f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x}))$. Recalling Theorem 6.13, we see that this expression is exactly the directional derivative of f at \mathbf{p} in the direction of \mathbf{v} .
2. We need to compute

$$\lim_{h \rightarrow 0} \frac{1}{h^{k+1}} \int_{\partial P_{\mathbf{p}}(h\mathbf{v}_1, \dots, h\mathbf{v}_{k+1})} f \, dx^{i_1} \wedge \dots \wedge dx^{i_k}.$$

The boundary $\partial P_{\mathbf{p}}(h\mathbf{v}_1, \dots, h\mathbf{v}_{k+1})$ is the disjoint union of the $2(k+1)$ faces of the $(k+1)$ -parallelapiped $P_{\mathbf{p}}(h\mathbf{v}_1, \dots, h\mathbf{v}_{k+1})$. We will construct charts $\{(U_{fi}, \mathbf{x}_{fi})\}_{i=1}^{k+1}$ that parameterize the $k+1$ faces on the “front” of this $(k+1)$ -parallelapiped and charts $\{(U_{bi}, \mathbf{x}_{bi})\}_{i=1}^{k+1}$ that parameterize the remaining $k+1$ faces on its “back”. Specifically, for $\mathbf{q} \in [0, h]^n \subseteq \mathbb{R}^n$, we define

$$\begin{aligned} \mathbf{x}_{fi}^{-1}(\mathbf{q}) &:= q^1 \mathbf{v}_1|_{\mathbf{p}} + \dots + q^{i-1} \mathbf{v}_{i-1}|_{\mathbf{p}} + \cancel{q^i \mathbf{v}_i|_{\mathbf{p}}} + q^{i+1} \mathbf{v}_{i+1}|_{\mathbf{p}} + \dots + q^{k+1} \mathbf{v}_{k+1}|_{\mathbf{p}} \\ \mathbf{x}_{bi}^{-1}(\mathbf{q}) &:= q^1 \mathbf{v}_1|_{\mathbf{p}} + \dots + q^{i-1} \mathbf{v}_{i-1}|_{\mathbf{p}} + h\mathbf{v}_i|_{\mathbf{p}} + q^{i+1} \mathbf{v}_{i+1}|_{\mathbf{p}} + \dots + q^{k+1} \mathbf{v}_{k+1}|_{\mathbf{p}}. \end{aligned}$$

Then, using Theorem [...], the integral inside the limit becomes

$$\int_{\partial P_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1})} f \, dx^{i_1} \wedge \dots \wedge dx^{i_k} = \underbrace{\sum_{i=1}^{k+1} \int_{\mathbf{x}_{fi}(U_{fi})} (\mathbf{x}_{fi}^{-1})^* (f \, dx^{i_1} \wedge \dots \wedge dx^{i_k})}_{\text{sum of integrals over front faces}} + \underbrace{\sum_{i=1}^{k+1} \int_{\mathbf{x}_{bi}(U_{bi})} (\mathbf{x}_{bi}^{-1})^* (f \, dx^{i_1} \wedge \dots \wedge dx^{i_k})}_{\text{sum of integrals over back faces}}$$

Expanding out the pullbacks $(\mathbf{x}_{fi}^{-1})^*$ and $(\mathbf{x}_{bi}^{-1})^*$ (recall Theorem [...]), we have

$$\begin{aligned} \int_{\partial P_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1})} f \, dx^{i_1} \wedge \dots \wedge dx^{i_k} &= \sum_{i=1}^{k+1} \int_{\mathbf{x}_{fi}(U_{fi})} \det(d\mathbf{x}_{fi}^{-1})(f \circ \mathbf{x}_{fi}^{-1}) \, dy^{i_1} \wedge \dots \wedge dy^{i_k} \\ &\quad + \\ &\quad \sum_{i=1}^{k+1} \int_{\mathbf{x}_{bi}(U_{bi})} \det(d\mathbf{x}_{bi}^{-1})(f \circ \mathbf{x}_{bi}^{-1}) \, dy^{i_1} \wedge \dots \wedge dy^{i_k} \end{aligned}$$

Now we evaluate $\det(d\mathbf{x}_{f_i}^{-1})$ and $\det(d\mathbf{x}_{b_i}^{-1})$. Looking at the definitions of $\mathbf{x}_{f_i}^{-1}$ and $\mathbf{x}_{b_i}^{-1}$, we calculate $d\mathbf{x}_{f_i}^{-1}$ and $d\mathbf{x}_{b_i}^{-1}$ by applying Theorem 8.35, which formalizes the familiar formula of the directional derivative. We find that these differentials are the same:

$$d\mathbf{x}_{f_j}^{-1} = d\mathbf{x}_{b_j}^{-1} = \begin{cases} \mathbf{v}_j|_{\mathbf{p}} dx^j & j \neq i \\ 0 & j = i \end{cases}.$$

1. Reconcile that dx^j is basis of $T_{\mathbf{p}}^*(M)$ - in order to take the determinant, should be taking coordinates relative to a basis of $T_{\mathbf{p}}(M)$. Once this is taken care of, the fact that the i th column is zero or gone means that we can swap cols of the det to get $(-1)^{i+1} \det(\mathbf{I}) = (-1)^{i+1}$.

2. For the above to work, the determinant should somehow only take in k vectors, rather than $k+1$ vectors.

$$\begin{aligned} \int_{\partial P_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_{k+1})} f &= \sum_{i=1}^{k+1} (-1)^{i+1} \int_{\mathbf{x}_{f_i}(U_{b_i})} (f \circ \mathbf{x}_{f_i}) dy^{i_1} \wedge \dots \wedge dy^{i_k} \\ &\quad + \\ &\quad \sum_{i=1}^{k+1} (-1)^i \int_{\mathbf{x}_{b_i}(U_{b_i})} (f \circ \mathbf{x}_{b_i}) dy^{i_1} \wedge \dots \wedge dy^{i_k} \\ &= \sum_{i=1}^{k+1} (-1)^{i+1} \int_{\mathbf{x}_{f_i}(U_{f_i}) \sqcup \mathbf{x}_{b_i}(U_{b_i})} f \circ (\mathbf{x}_{f_i} - \mathbf{x}_{b_i}) dy^{i_1} \wedge \dots \wedge dy^{i_k} \end{aligned}$$

Can change domain of integration to $\mathbf{x}_{f_i}(U_{f_i}) \sqcup \mathbf{x}_{b_i}(U_{b_i})$ because $\mathbf{x}_{f_i}(U_{f_i})$ and $\mathbf{x}_{b_i}(U_{b_i})$ are pairwise nonoverlapping

□

Theorem 9.25. [Lee, p.364] ($d \circ d = 0$).

Theorem 9.26. [Lee, p.366] (d commutes with the pullback of a smooth function $M \rightarrow N$).

<https://math.stackexchange.com/questions/1302562/proving-that-the-pullback-map-commutes-with>

Theorem 9.27. [HH, p.652 - 655] (Linearity of the exterior derivative).

The exterior derivative is linear: $d(f_1\omega + f_2\eta) = f_1d\omega + f_2d\eta$.

Hodge duality

Definition 9.28. (Hodge dual).

Let V be an n -dimensional vector space. The *Hodge dual* (on V) is the map $*$: $\Lambda^k(V) \rightarrow \Lambda^{n-k}(V)$, that is defined on elementary k -wedges, and extended with linearity and the antisymmetry of \wedge , by

$$(\mathbf{v}_{i_1} \wedge \dots \wedge \mathbf{v}_{i_k}) \wedge *(\mathbf{v}_{i_1} \wedge \dots \wedge \mathbf{v}_{i_k}) = \mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n.$$

Equivalently,

$$*(\mathbf{v}_{i_1} \wedge \dots \wedge \mathbf{v}_{i_k}) = \text{sgn}(\sigma) \mathbf{v}_{j_1} \wedge \dots \wedge \mathbf{v}_{j_{n-k}},$$

where $\sigma = (i_1, \dots, i_k, j_1, \dots, j_{n-k})$, i.e., $i_\ell = \sigma(\ell)$ and $j_\ell = \sigma(\ell + k)$.

Optional Hodge theory

Lemma 9.29. A metric g determines an inner product \tilde{g} on $T_{\mathbf{p}}^*(M)$.

$$\tilde{g}(\omega, \eta) = g(\omega^{\flat}, \eta^{\flat})$$

then identify $T_{\mathbf{p}}^{**}(M) \cong T_{\mathbf{p}}(M)$

Theorem 4.5

Lemma 9.30. [Lee, p.437 - 438]

A metric g determines an inner product, also denoted g , on $\Lambda^k(T_{\mathbf{p}}^*(M))$ defined by

$$g(\omega^1 \wedge \dots \wedge \omega^k, \nu^1 \wedge \dots \wedge \nu^k) = \det(\tilde{g}(\omega^i, \eta^j)),$$

where \tilde{g} is from the previous lemma.

Theorem 9.31.

$$\omega \wedge * \eta = g(\omega, \eta) dV_g,$$

where g from the previous theorem

Definition 9.32. (Hodge inner product).

$$\langle\langle \omega, \eta \rangle\rangle := \int_M g(\omega, \eta) dV_g = \int_M \omega \wedge * \eta$$

Theorem 9.33.

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = *(\mathbf{v}_1^{\flat} \wedge *(\mathbf{v}_2^{\flat}))$$

$$\mathbf{v}_1 \times \mathbf{v}_2 = (*(\mathbf{v}_1^{\flat} \wedge \mathbf{v}_2^{\flat}))^{\sharp}$$

$$\nabla f = (df)^{\sharp}$$

$$\text{curl}(\mathbf{V}) = (*d(\mathbf{V}^{\flat}))^{\sharp}$$

$$\text{div}(\mathbf{V}) = \text{curl}(\mathbf{R}_{\frac{\pi}{2}} \mathbf{V}) = *d * (\mathbf{V}^{\flat})$$

“In \mathbb{R}^3 , div , grad curl are just d applied to 0, 1, 2 forms, resp.”

$$\alpha = \sum_i f_i dx^i \implies \alpha^{\sharp} = \sum_i f^i \frac{\partial}{\partial x^i}$$

Let (U, \mathbf{x}) be a smooth chart on \mathbb{R}^3 . \flat and \sharp are the musical isomorphisms induced by the choice of basis $\left\{ \frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right\}_{i=1}^n$, i.e., induced by the inner product $\langle \cdot, \cdot \rangle$ on $T_{\mathbf{p}}(M)$ specified in [...] [?]

should remove \natural and replace with \flat

Old notes

- Volume form μ

- Define volume form as n -form μ for which $\mu(\mathbf{X}_1, \dots, \mathbf{X}_n) = \det(\mathbf{X}_1, \dots, \mathbf{X}_n)$, where the \det on the RHS is the function on $\mathfrak{X}(M)^{\times n} \times M$ sending $(\mathbf{X}_1, \dots, \mathbf{X}_n, \mathbf{x}) \mapsto \det(\mathbf{X}_1(\mathbf{x}), \dots, \mathbf{X}_n(\mathbf{x}))$. Use “8.15. Proposition” of Sjmaar.

- Sjmaar p. 108 examples about how μ relates to ds , dA , dV

- Use Lee Prop. 15.29 and 15.31 give formulas about volume forms

- Hodge duality.

- Let V be a finite-dimensional vector space and consider an ordered basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for V . Let $\sigma_W \in S_k$ and $\sigma_{W^\perp} \in S_{n-k}$. Set W to be the oriented subspace of V spanned by the ordered basis $\{\mathbf{e}_{\sigma(1)}, \dots, \mathbf{e}_{\sigma(k)}\}$, and give $W^\perp \subseteq V$ the orientation specified by $\{\mathbf{e}_{\sigma(k+1)}, \dots, \mathbf{e}_{\sigma(n)}\}$.

Equivalently, the orientations chosen on W and W^\perp can be represented as elements of $\Lambda^k(W) \subseteq \Lambda^k(V)$ and $\Lambda^{n-k}(W^\perp) \subseteq \Lambda^{n-k}(V)$, respectively. The orientation on W is specified by $\mathbf{e}_{\sigma(1)} \wedge \dots \wedge \mathbf{e}_{\sigma(k)}$ and the orientation of W^\perp is specified by $\mathbf{e}_{\sigma(k+1)} \wedge \dots \wedge \mathbf{e}_{\sigma(n)}$.

The *Hodge dual*, or *Hodge star*, is the map $*$: $\Lambda^k(W) \rightarrow \Lambda^{n-k}(W^\perp)$ which acts on the orientation chosen for a subspace $W \subseteq V$ to the orientation chosen for W^\perp .

It is enough to say that $*$ “acts” on $\mathbf{e}_{\sigma(1)} \wedge \dots \wedge \mathbf{e}_{\sigma(k)}$ as the action on this element of $\Lambda^k(W)$ is implied: $*$ is a map of one-dimensional vector spaces, so we know

$$*(\mathbf{e}_{\sigma(1)} \wedge \dots \wedge \mathbf{e}_{\sigma(k)}) =$$

9.4 The generalized Stokes’ theorem

[in practice vs. theoretical niceties]

Definition 9.34. [Lee, p.43] (Partitions of unity).

If $\{U_\alpha\}$ is an open cover of M and $\{f_\alpha\}$ is a set of functions $M \rightarrow \mathbb{R}$, we say that $\{f_\alpha\}$ is a *partition of unity of M (subordinate to $\{U_\alpha\})$* iff

- $\text{supp}(f_\alpha) \subseteq U_\alpha$ for all α
- $f_\alpha(\mathbf{p})$ is nonzero for only finitely many α , for each $\mathbf{p} \in M$
- $(\sum_\alpha f_\alpha(\mathbf{p}) = 1 \text{ for all } \mathbf{p} \in M) \iff (\sum_\alpha f \text{ is the multiplicative identity of } \{\text{functions } M \rightarrow \mathbb{R}\}).$

A *smooth partition of unity* is a partition of unity in which each function f_α is smooth.

Every open cover of a smooth manifold with or without corners admits a smooth partition of unity ([Lee, p. 43]). Since every manifold has an open cover, this means that every smooth manifold with or without corners admits a smooth partition of unity.

Theorem 9.35. (Existence of partitions of unity).

Theorem 9.36. (Integral of a differential n -form on a smooth n -manifold via partition of unity).

Let M be an oriented smooth n -manifold with or without corners and let ω be a compactly supported n -form on M . Since $\text{supp}(\omega)$ is compact, there is a finite collection $\{(U_i, \mathbf{x}_i)\}_{i=1}^k$ of charts on M for which $\{U_i\}_{i=1}^k$ is a pairwise nonoverlapping open cover of $\text{supp}(\omega)$. Additionally, let $\{f_i\}_{i=1}^k$ be a smooth partition of unity subordinate to $\{U_i\}_{i=1}^k$.

Simply repeating the previous definition, we have,

$$\int_M \omega = \sum_{i=1}^k \int_{U_i} \omega.$$

Now, since each $f_i \omega$ is supported in U_i , we have

$$\int_{U_i} \omega = \int_{U_i} f_i \omega = \int_M f_i \omega$$

The last equality of the line above is just the definition of a differential form that is compactly supported in a single chart (see Definition [...]).

Summing over i , we obtain

$$\sum_{i=1}^k \int_{U_i} \omega = \sum_{i=1}^k \int_M f_i \omega.$$

Therefore the integral of a differential form over a manifold can be computed with use of the partition of unity:

$$\boxed{\int_M \omega = \sum_{i=1}^k \int_M f_i \omega}$$

what guarantees that $\mathbf{x}_i(U_i)$ is a domain of integration? (and this definitely came up in previous proofs

Remark 9.37. (Integration with partitions of unity).

Integration of differential forms on manifolds is most often *defined* in terms of partitions of unity. We prefer to view the partition of unity method as a consequence of the definition $\int_M \omega := \sum_{i=1}^k \int_{U_i} \omega$ (see Definition 9.22), as this definition is an intuitive starting point.

Theorem 9.38. [Lee, p.407] (Properties of integrals of differential forms).

Definition 9.39. [HH, p.356] (Dyadic paving).

also, assumption on Riemann sums

Definition 9.40. (Integral of a k -form, $k \leq n$, over an n -manifold)

Use inclusion map

Theorem 9.41. (The generalized Stokes' theorem on a single smooth chart).

Let M be a smooth n -manifold and consider a smooth chart (U, \mathbf{x}) on M , where U has the standard orientation of \mathbb{R}^k [?] and ∂U has the induced boundary [Stokes'?] orientation [shouldn't this be given to $\mathbf{x}(U)$?]. Let ω be a differential $(k-1)$ -form with compact support in U of differentiability class C^2 . Then

$$\int_M d\omega = \int_{\partial M} \omega.$$

Intuitively, this theorem is true because

$$\int_M d\omega \approx \sum_i d\omega(P_i) \approx \sum_i \int_{\partial P_i} \omega \approx \int_{\partial M} \omega.$$

The first approximation holds because integrals are limits of Riemann sums. The second approximation holds because the definition of the exterior derivative implies $d\omega(C) \approx \int_{\partial C} \omega$, where C is one of the k -cubes in the previous Riemann sum (a k -cube is a k -parallelapiped, and differential forms evaluate k -parallelapipeds). The third approximation holds because each boundary ∂P_i in the sum of integrals corresponds to exactly one oppositely oriented boundary $\partial P_i = -\partial P_j$ that occupies the same space: we have $\sum_{C \in D_N(\text{cl}(\mathbb{H}^k))} \int_{\partial C} \omega = \sum_{C' \in D_N(\partial \text{cl}(\mathbb{H}^k))} \int_{C'} \omega$, which is equal to $\int_{\partial \text{cl}(\mathbb{H}^k)} \omega$ by Definition [...].

Proof. Let $\mathbf{x}(U)$ be an open subset of \mathbb{H}^k . ??

Each “approximation” above loosely translates to a statement of the form $f(M, \omega, N_1) \approx g(M, \omega, N_2)$. A statement such as this further translates to the formal statement (for all M, ω, N_1 and for all $\epsilon > 0$, there exists an $N_2 \in \mathbb{N}$ for which $|f(M, \omega, N_1) - g(M, \omega, N_2)| < \epsilon$). It suffices to prove each formal statement individually, because approximation treated this way is associative. So, overall, we are showing [...]

$(\int_U d\omega \approx \sum_i d\omega(P_i))$. Let $\epsilon > 0$. Then there exists an N large enough such that the dyadic paving of U of fineness 2^{-N} ensures $|\int_U d\omega - \sum_{C_i \in D_N(\text{cl}(\mathbb{H}^k))} d\omega(P_i)| < \epsilon$.

$(\sum_i d\omega(P_i) \approx \sum_i \int_{\partial P_i})$. Let $\epsilon > 0$ (forget about the previous ϵ). Take the dyadic decomposition of U of fineness 2^{-N} , so that U is a countable disjoint union of k -cubes with side length 2^{-N} . Each k -cube is a k -parallelapiped of the form $C_i = P_{\mathbf{p}}(h\hat{\mathbf{e}}_1, \dots, h\hat{\mathbf{e}}_k)$ for some \mathbf{x} , and where $h = 2^{-N}$. By the definition of the exterior derivative (Definition [...]), there exist $K, \delta > 0$ such that when $|h| = 2^{-N} < \delta$, we have $|\int_{C_i} d\omega - \sum_i d\omega(C_i)| < Kh^{k+1}$. By taking N sufficiently large, $|h|$ becomes sufficiently small, and we get $|\int_{C_i} d\omega - \sum_i d\omega(C_i)| < Kh^{k+1} < \epsilon$.

$(\sum_i \int_{\partial P_i} \omega \approx \int_{\partial U} \omega)$. Let $\epsilon > 0$ (forget about the previous ϵ).

Biggest thing I need to show is

$$\sum_{C \in D_N(\text{cl}(\mathbb{H}^k))} \int_{\partial C} \omega = \sum_{C' \in D_N(\partial \text{cl}(\mathbb{H}^k))} \int_{C'} \omega$$

□

Remark 9.42. The above theorem is the real “heart” of Stokes’ theorem. The proof of the full-blown Stokes’ theorem just stitches together the application of Stokes’ theorem to all charts [in the smooth structure].

[HH, p.543] might be helpful, talks about boundary of a cube

Will also have to use notion of integrating over oriented manifold.

Theorem 9.43. [Lee, p.419] (The generalized Stokes’ theorem).

Let M be an oriented smooth manifold with corners, and let ω be a compactly supported smooth [?] differential $(n - 1)$ form on M . Then

$$\int_M d\omega = \int_{\partial M} \omega.$$

Proof. [HH, p.661 - 665]

[Partitions of unity make this proof nice.]

Using the theorem which expresses $\int_M \omega$ in terms of a partition of unity, and using the linearity of the exterior derivative, we have

$$\int_M d\omega = \int_M \sum_{i=1}^k f_i d\omega = \int_M \sum_{i=1}^k d(\underbrace{f_i \omega}_{\omega_i}) = \int_M \sum_{i=1}^k d\omega_i.$$

Here, we have⁴ defined $\omega_i := f_i \omega$.

Using the linearity of the integral, we have

$$\int_M \sum_{i=1}^k d\omega_i = \sum_{i=1}^k \int_M d\omega_i.$$

Recalling Definition [...], this becomes

$$\sum_{i=1}^k \int_M d\omega_i = \sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^* (d\omega_i).$$

Because $d \circ \mathbf{F}^* = \mathbf{F}^* \circ d$ [...], we get

⁴This notation reflects that, in some sense, using a partition of unity on a topological manifold is similar to choosing a basis for a vector space. If we *really* wanted to emphasize this fact, we might have denoted $f_i \omega$ by $([\omega]_{\{f_i\}_{i=1}^k})_i$.

$$\sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(d\omega_i) = \sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} d((\mathbf{x}_i^{-1})^*(\omega_i)).$$

We apply the previous theorem- Stokes' theorem on a single smooth chart- to the integral inside the sum to obtain

$$\sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} d((\mathbf{x}_i^{-1})^*(\omega_i)) = \sum_{i=1}^k \int_{\partial \mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(\omega_i).$$

We use that $\partial \mathbf{x}_i(U_i) = \mathbf{x}_i(\partial U_i)$ and apply Definition [...] again to get

$$\sum_{i=1}^k \int_{\partial \mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(\omega_i) = \sum_{i=1}^k \int_{\mathbf{x}_i(\partial U_i)} (\mathbf{x}_i^{-1})^*(\omega_i) = \sum_{i=1}^k \int_{\partial U_i} \omega_i.$$

Finally, since $\{U_i\}_{i=1}^k$ is a finite pairwise nonoverlapping open cover for $\text{supp}(\omega) \subseteq M$, then $\{\partial U_i\}_{i=1}^k$ is a finite pairwise nonoverlapping open cover for $\partial \text{supp}(\omega) \subseteq \partial M$. Thus, this last expression is

$$\sum_{i=1}^k \int_{\partial U_i} \omega_i = \sum_{i=1}^k \int_{\partial M} f_i \omega = \int_{\partial M} \omega.$$

The last equality follows by Theorem [...]. □

Remark 9.44. empty boundary?
pullback of inclusion?

Remark 9.45. (d and ∂ are dual).

$$\langle d\omega, M \rangle = \langle \omega, \partial M \rangle$$

<https://math.stackexchange.com/questions/939599/is-the-product-rule-for-the-boundary-of-a>
<https://mathoverflow.net/questions/46252/is-the-boundary-partial-s-analogous-to-a-derivativ>

Examples

11-27, 14-18, 16-9

9.5 Etc.

Geometric interpretation of linear functionals

Given a basis E for a finite-dimensional vector space V , a dual vector $\phi \in V^*$ acts on a vector by $\phi(\mathbf{v}) = [\phi]_E \cdot \mathbf{v}$. This is reminiscent of the equation for a plane.

Recall, the plane that contains the point $\mathbf{x}_0 \in V$ and has (not necessarily unit-length) normal \mathbf{n} is the set $\{\mathbf{x} \mid (\mathbf{x} - \mathbf{x}_0) \cdot \mathbf{n} = 0\}$, which is the same set as $\{\mathbf{x} \mid \mathbf{x} \cdot \mathbf{n} = \mathbf{x}_0 \cdot \mathbf{n}\}$. This means that if we consider the level planes $\{\phi(\mathbf{v}) = [\phi]_E \cdot \mathbf{v} = k \mid k \in K\}$, then $[\phi]_E$ takes the role of \mathbf{x} , \mathbf{v} takes the role of \mathbf{n} , and k takes the role of $\mathbf{x}_0 \cdot \mathbf{n}$ (for some \mathbf{x}_0).

Noticing this implies the following facts about $\{\phi(\mathbf{v}) = k \mid k \in K\}$.

1. The level planes are parallel to each other, because their normal vectors are all scalar multiples of \mathbf{v} .
2. $[\phi]_E$ is perpendicular to each level plane. (Consider that $[\phi]_E$ is perpendicular to the level plane corresponding to $k = 0$, since $[\phi]_E \cdot \mathbf{v} = 0$. This level plane is parallel to every other level plane, so $[\phi]_E$ is perpendicular to every level plane.)

Now, we construct a set of level planes analogous to a coordinate system. Consider the set of level planes for which $k = 0$, $k = 1$, and $k = 1 \cdot j$ for $j \in \mathbb{Z}$ (the homomorphism $j \mapsto j \cdot 1_F$ relates to ring characteristic!). Then to evaluate $h(\mathbf{v}) = [\phi]_E \cdot \mathbf{v}$, we count the number of these level planes that $[\phi]_E$ pierces.

This interpretation emphasizes thinking of dual vectors as corresponding to row-vectors, since row-vector times column-vector is the same as the dot product between the two corresponding column vectors.

Piponi visualization of diff forms

1-forms are integrated along paths.

To evaluate a linear functional on a vector, think of the linear functional as represented by a set of parallel $(n - 1)$ -dimensional planes, each a unit distance apart, and count how many planes the vector goes through.

To integrate a differential 1-form over a curve, think of the diff 1-form as represented by curving $(n - 1)$ -dimensional “leaves.” Take the directed count of how many leaves the curve passes through.

A diff k -form on an n -manifold is represented by $(n - k)$ -dimensional leaves.

A nonzero exterior product of a diff k form with a diff ℓ form is a $k + \ell$ form because intersecting a collection of $(n - k)$ -dimensional leaves with a collection of $(n - \ell)$ -dimensional leaves gives a collection of $(n - (k + \ell))$ -dimensional leaves.

A differential n -form on an n -manifold is a “top” form whose leaves are points. Such a diff form is a volume density.

To integrate $dx \wedge dy = \epsilon^1 \wedge \epsilon^2$ over an area, draw vertical grid lines (corresponding to ϵ^1), draw horizontal grid lines (corresponding to ϵ^2),

Part IV

Computational applications of differential forms

10

Discrete differential geometry and computer graphics

10.1 Discrete differential geometry

This section follows Section 4.8 of [Cra20].

Will be considering manifolds embedded in \mathbb{R}^n .

Definition 10.1. (k -cell in \mathbb{R}^n).

A k -cell is a subset of \mathbb{R}^n that is a Cartesian product of k closed intervals.

Definition 10.2. (Oriented k -simplex in \mathbb{R}^n).

Consider $k \leq n$. An *oriented k -simplicial complex in \mathbb{R}^n* , more commonly known as a *oriented k -simplex in \mathbb{R}^n* , is tuple (K, Or) , where $K \subseteq \mathbb{R}^n$ is a subset of the form

$$\left\{ \sum_{i=1}^k c_i \mathbf{q}_i \mid \sum_{i=1}^k c_i = 1, c_i \geq 0 \right\},$$

and where Or is the standard orientation of $\{\mathbf{p}_1, \dots, \mathbf{p}_k\}$ inherited from \mathbb{R}^k .

If K' and K are two k -simplices with $K' \subseteq K$, then K' is said to be a *face* of K .

Example 10.3. (0-, 1-, and 2- simplices).

A 0-simplex is a vertex (a point with positive orientation), a 1-simplex is an edge (an oriented line segment), a 2-simplex is an oriented triangle, and a 3-simplex is an oriented tetrahedron.

“Finding a discrete counterpart to the notion of differential forms is a delicate matter. If one was to represent differential forms using their coordinate values and approximate the exterior derivative using finite differences, basic theorems such as Stokes’ theorem would not hold numerically.” (<http://www.geometry.caltech.edu/pubs/DKT05.pdf>)

Definition 10.4. (Discrete differential form).

Let K be a $(k+1)$ -simplex. We define the *discrete differential k -form $\hat{\omega}$ on K* (corresponding to a differential ℓ -form ω) to be the function which integrates a differential k -form on K over an arbitrary set of k -simplices (i.e. edges) in K . Specifically, the *evaluation $\hat{\omega}|_{\{e_i\}}$ of $\hat{\omega}$ on a set of k -simplices $\{e_i\}$ in K* is defined to be

$$\hat{\omega}|_{\{e_i\}} := \int_{\{e_i\}} \omega = \sum_i \int_{e_i} \omega.$$

Here, the hat $\hat{}$ is meant to denote “discrete” (not “normalized”).

Note that a discrete 0-forms is a 0-form (i.e. a function) that’s been integrated over every 0-simplex in a 1-simplex. The integral of a function “over” a point is zero, but we will use the convention of treating the value of a discrete 0-form $\hat{f}|_{\mathbf{p}_i}$ at a point to be the value of the corresponding function at that point, $\hat{f}|_{\mathbf{p}_i} := f|_{\mathbf{p}_i}$.

Definition 10.5. (Discrete exterior derivative).

\hat{d} acts on $\hat{\omega}|_{\mathbf{p}_i}$ by

$$\hat{d}\hat{\omega}|_{\mathbf{p}_i} := \int_{(K, \text{Or})} d\omega,$$

where (K, Or) is the k -simplex whose orientation Or is the standard orientation of $\{\mathbf{p}_i, \mathbf{p}_{i+1}, \dots, \mathbf{p}_1, \dots, \mathbf{p}_{i-1}\}$.

Theorem 10.6. (Stokes' theorem for discrete differential forms).

If a k -simplex is composed of edges e_i , then by definition, the evaluation of a discrete differential form $\hat{\omega}$ at a vertex \mathbf{p}_i is

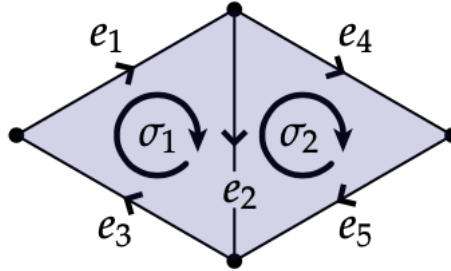
$$\hat{d}\hat{\omega}|_{\mathbf{p}_i} = \int_{(K, \text{Or})} d\omega = \int_{(\partial K, \text{Or})} \omega = \sum_i \int_{\partial e_i} \omega = \sum_i \hat{\omega}_{e_i}$$

(Note, we have used the fact that the boundary of a manifold inherits the manifold's orientation; $\partial(K, \text{Or}) = (\partial K, \text{Or})$).

Therefore we recover a discrete version of Stokes' theorem:

$$\boxed{\hat{d}\hat{\omega}|_{\mathbf{p}_i} = \sum_i \hat{\omega}_{e_i}}$$

Example 10.7. (Computing the discrete exterior derivative).



In the above figure, take \mathbf{p}_1 to be the leftmost vertex and \mathbf{p}_2 to be the topmost vertex. We have

$$\begin{aligned} \hat{d}\hat{\omega}|_{\mathbf{p}_1} &= \hat{\omega}_1 + \hat{\omega}_2 + \hat{\omega}_3 \\ \hat{d}\hat{\omega}|_{\mathbf{p}_2} &= \hat{\omega}_4 + \hat{\omega}_5 - \hat{\omega}_1. \end{aligned}$$

Definition 10.8. (Mesh duality).

primal mesh, dual mesh

“Let us introduce the notion of dual mesh of triangulated manifolds, as we will see that it is one of the key components of our discrete calculus. The main idea is to associate to each primal k -simplex a dual $(n-k)$ -cell. For example, consider the tetrahedral mesh in Figure 13, we associate a dual 3-cell to each primal vertex (0-simplex), a dual polygon (2-cell) to each primal edge (1-simplex), a dual edge (1-cell) to each primal face (2-simplex), and a dual vertex (0-cell) to the primal tet (3-simplex). By construction, the number of dual $(n-k)$ -cells is equal to that of primal k -simplices. The collection of dual cells is called a cell complex, which need not be a simplicial complex in general. Yet, this dual complex inherits several properties and operations from the primal simplicial complex. Most important is the notion of incidence. For instance, if two primal edges are on the same primal face, then the corresponding dual faces are incident, that is, they share a common dual edge (which is the dual of the primal common face). As a result of this incidence property, one may easily derive”

(<http://www.geometry.caltech.edu/pubs/DKT05.pdf>)

$d_{\text{dual}}^{n-k} = (-1)^k (d_{\text{primal}}^{k-1})^*$, where $*$ is dual as in transpose

Primal Hodge-dual, dual Hodge-dual

Definition 10.9. (Diagonal Hodge-dual)

Multiple types of Hodge dual: <https://graphics.stanford.edu/courses/cs468-13-spring/assets/lecture14-saenz.pdf>

Theorem 10.10. (Discrete Laplacian).

Let M be a Riemannian manifold WWBOC, and let (U, \mathbf{x}) be a smooth chart on M . The *Laplacian* is

define on a manifold?

Regular Laplacian is

$$\Delta = *d * d$$

Discrete Laplacian is therefore

$$\hat{\Delta} = \hat{*}_{d \rightarrow p} \hat{d}_{\text{dual}} \hat{*}_{p \rightarrow d} \hat{d}_{\text{primal}},$$

where \hat{d}_{primal} and $\hat{*}_{p \rightarrow d}$ are the discrete exterior derivative and diagonal Hodge-dual on the primal mesh, and where \hat{d}_{dual} and $\hat{*}_{d \rightarrow p}$ are the discrete exterior derivative and diagonal Hodge-dual on the dual mesh.

Let \hat{f} be a discrete differential 0-form. First, we evaluate $\hat{d}\hat{f}$ on an arbitrary edge; let (e_{ij}, Or) be the edge connecting vertices \mathbf{p}_i and \mathbf{p}_j , where Or is the standard orientation of $\{\mathbf{p}_i, \mathbf{p}_j\}$. To take the discrete exterior derivative, we integrate along the single edge e_{ij} :

$$\hat{d}_{\text{primal}} \hat{f}|_{e_{ij}} = \int_{(e_{ij}, \text{Or})} df = \int_{(\partial e_{ij}, \text{Or})} f = f|_{\mathbf{p}_j} - f|_{\mathbf{p}_i}$$

Then, applying the diagonal Hodge-dual $\hat{*}_{p \rightarrow d}$, we have

$$(\hat{*}_{p \rightarrow d} \hat{f})|_{e_{ij}} = \frac{|e_{ij}^*|}{|e_{ij}|} (f|_{\mathbf{p}_j} - f|_{\mathbf{p}_i})$$

Here, e_{ij}^* is the $(n-k)$ -cell that is dual to the k -cell (the edge) e_{ij} . Evaluating the discrete exterior derivative at a vertex \mathbf{q}_i in the dual mesh requires that we integrate over the cell $C_{\mathbf{q}_i}$ in the dual mesh which contains \mathbf{q}_i :

$$(\hat{d}_{\text{dual}} \hat{*}_{p \rightarrow d} \hat{d}_{\text{primal}} \hat{f})|_{\mathbf{q}_i} = \int_{C_{\mathbf{q}_i}} d_{\text{dual}} *_{\text{primal}} d_{\text{primal}} f.$$

Using Stokes' theorem for discrete differential forms, this becomes

$$(\hat{d}_{\text{dual}} \hat{*}_{p \rightarrow d} \hat{d}_{\text{primal}} \hat{f})|_{\mathbf{q}_i} = \sum_j \frac{|e_{ij}^*|}{|e_{ij}|} (f|_{\mathbf{p}_j} - f|_{\mathbf{p}_i}).$$

Note that the integrand contains no hats $\hat{}$ because the integrand is a “regular” (not discrete) differential form.

The last Hodge-dual **divides** the previous result by the volume of $C_{\mathbf{q}_i}$, so in all, we have

$$(\hat{d}_{\text{dual}} \hat{*}_{p \rightarrow d} \hat{d}_{\text{primal}} \hat{f})|_{\mathbf{q}_i} = \frac{1}{|C_{\mathbf{q}_i}|} \sum_j \frac{|e_{ij}^*|}{|e_{ij}|} (f|_{\mathbf{p}_j} - f|_{\mathbf{p}_i})$$

\mathbf{p}_i vs. \mathbf{q}_i

10.2 Smoothing with the Laplacian

Outline

- (End goal). Solve the problem $\frac{\partial \mathbf{f}}{\partial t} = \Delta \mathbf{f}$. Application to graphics: “smooth out geometric detail on a surface mesh (also known as *fairing* or *curvature flow*)”
- (Theory outline). First subproblem: solve the problem $\Delta \mathbf{f} = \mathbf{g}$, where $\mathbf{f}, \mathbf{g} : M \rightarrow \mathbb{R}^3$. (The previous problem uses $\mathbf{g} = \frac{\partial \mathbf{f}}{\partial t}$. Second subproblem: solve the problem $\Delta u = f$, where $u, f : M \rightarrow \mathbb{R}$. After we solve the second subproblem, we can use its solution to solve first subproblem, because $\Delta \mathbf{f} = \Delta \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} \Delta f_1 \\ \Delta f_2 \\ \Delta f_3 \end{pmatrix}$. (Define vector of functions notation).
- Two approaches to theory. Both approaches discretize and find approximate solutions. FEM discretizes by only considering solutions that live in a *discrete approximation space*. DEC discretizes by using the discrete exterior derivative [and it probably discretizes in other ways too- investigate]. Both approaches solve the second subproblem by yielding the “cotan formula”, which is an expression for the Laplacian of $u : M \rightarrow \mathbb{R}$ at vertex i .
 - (6.2) (FEM). Let $\{\phi_{\mathbf{p}_i}\}$ be a basis for {approx space}. Let f be the true solution (f is not necessarily in the approximation space). The best approximation $u \in \{\text{approx space}\}$ is the \tilde{u} for which $\langle \Delta \tilde{u} - f, \phi_{\mathbf{p}_i} \rangle = 0$ for all i , where $\langle \cdot, \cdot \rangle$ is the L^2 inner product.
 - (6.3) (Discretization via discrete exterior calculus).
 - * $\Delta u = f$ is the same as $*d * du = f$.
 - “The Hodge star converts a circulation along the edge e_{ij} into the flux through the corresponding dual edge e_{ij}^* .”
 - * Something of relevance in this discussion is the same as discretizing the system $(d*d)u = *f$. $(d*d)$ is called the conformal Laplacian.

Basic facts about the Laplacian

Definition 10.11. (Laplacian).

Definition 10.12. (Harmonic function).

A smooth function $f : M \rightarrow \mathbb{R}$ is called *harmonic* iff f is in the kernel of the Laplacian.

Lemma 10.13. (Characterizing harmonic functions).

Let M be a smooth compact connected manifold without boundary. Then for all functions $f : M \rightarrow \mathbb{R}$, f is harmonic iff f is a constant function.

(Therefore, if f solves the Laplace equation, then so does $f + c$ for any constant function c).

Proof.

(\Leftarrow). If $f \equiv c$, then all second partial derivatives of f are the zero function, so $\Delta f = 0$.

(\Rightarrow). If f is harmonic, then $\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$. **integrate; how would this fail if we had a boundary?** \square

Theorem 10.14. On a smooth compact connected manifold without boundary, the Laplacian problem $\Delta f = g$ is not well posed if $g = g_1 + c$ for some constant function c .

Proof. Note that on a smooth compact connected manifold without boundary, constant functions are not in the image of the Laplacian. Therefore if $g = g_1 + c$, then we should have $\Delta f = g_1 + c$ \square

Definition 10.15. (L^2 inner product of functions $M \rightarrow \mathbb{R}$).

$$\langle f, g \rangle := \int_M fg.$$

Definition 10.16. (L^2 inner product for vector fields on M).

$$\langle \mathbf{V}_1, \mathbf{V}_2 \rangle := \int \mathbf{V}_1 \cdot \mathbf{V}_2,$$

where $\mathbf{V}_1 \cdot \mathbf{V}_2$ denotes the function $\mathbf{p} \mapsto \mathbf{V}_1|_{\mathbf{p}} \cdot \mathbf{V}_2|_{\mathbf{p}}$.

Theorem 10.17. (Green's first identity).

$$\langle \Delta f, g \rangle = -\langle \nabla f, \nabla g \rangle + \langle \hat{\mathbf{n}} \cdot \nabla f, g \rangle_{\partial},$$

where $\hat{\mathbf{n}}$ is the outward unit normal and $\langle \cdot, \cdot \rangle_{\partial}$ is the L^2 inner product on the boundary.

Definition 10.18. (Positive-semidefinite operator).

A function $F : C^\infty(M \rightarrow \mathbb{R}) \rightarrow C^\infty(M \rightarrow \mathbb{R})$ is said to be *positive-semidefinite* iff $\langle F(f), f \rangle \geq 0$ for all $f \in C^\infty(M \rightarrow \mathbb{R})$.

relate to positive-definiteness of inner product

Theorem 10.19. (Δ is negative-semidefinite on any compact manifold without boundary).

FEM

Derivation 10.20. Have problem

$$\Delta f = g.$$

We want to find the best approximation $\tilde{f} \in \{\text{approx space}\}$ to the (existence and uniqueness?) solution f .

Approximation space will be $C^\infty(M)$.

We will define the best approximation \tilde{f} to be the function that satisfies

$$\langle \Delta \tilde{f} - g, \phi_{\mathbf{p}_i} \rangle$$

for all $\phi_{\mathbf{p}_i}$. The function $\Delta \tilde{f} - g$ is called the *residual*.

We will use the *hat functions* for our $\phi_{\mathbf{p}_i}$.

$$\langle \Delta \tilde{f} - g, \phi_{\mathbf{p}_i} \rangle = 0$$

Derivation 10.21.

$$\langle \Delta \tilde{f}, \phi_{\mathbf{p}_i} \rangle = \sum_j \langle \Delta \tilde{f}, \phi_{\mathbf{p}_i} \rangle_{\sigma_j}.$$

Using Green's first identity inside the sum, we have

$$\langle \Delta \tilde{f}, \phi_{\mathbf{p}_i} \rangle = - \sum_j \langle \nabla \tilde{f}, \nabla \phi_{\mathbf{p}_i} \rangle_{\sigma_j} + \sum_j \langle \hat{\mathbf{n}} \cdot \nabla \tilde{f}, \phi_{\mathbf{p}_i} \rangle_{\partial \sigma_j}.$$

If the mesh has no boundary, then some cancellations happen, and we're left with

$$\langle \Delta \tilde{f}, \phi_{\mathbf{p}_i} \rangle = - \sum_j \langle \nabla \tilde{f}, \nabla \phi_{\mathbf{p}_i} \rangle_{\sigma_j} = - \langle \nabla \tilde{f}, \nabla \phi_{\mathbf{p}_i} \rangle.$$

Expanding \tilde{f} in the basis $\Phi = \{\phi_{\mathbf{p}_i}\}_{i=1}^n$ as $\tilde{f} = \sum_j ([\tilde{f}]_{\Phi})_j \phi_{\mathbf{p}_j}$, using the linearity of ∇ , and using the bilinearity of $\langle \cdot, \cdot \rangle$, we have

$$\langle \Delta \tilde{f}, \phi_{\mathbf{p}_i} \rangle = \langle \nabla \left(\sum_j ([\tilde{f}]_{\Phi})_j \phi_{\mathbf{p}_j} \right), \nabla \phi_{\mathbf{p}_i} \rangle = \sum_j ([\tilde{f}]_{\Phi})_j \langle \nabla \phi_{\mathbf{p}_j}, \nabla \phi_{\mathbf{p}_i} \rangle.$$

So it remains to compute $\langle \nabla \phi_{\mathbf{p}_j}, \nabla \phi_{\mathbf{p}_i} \rangle$.

Cotan formula

Lemma 10.22. (Cotan formula lemma 1).

The “aspect ratio” $\frac{w}{h}$ of a triangle is the sum of the cotangents of the interior angles at its base. That is,

$$\frac{w}{h} = \cot(\alpha) + \cot(\beta).$$

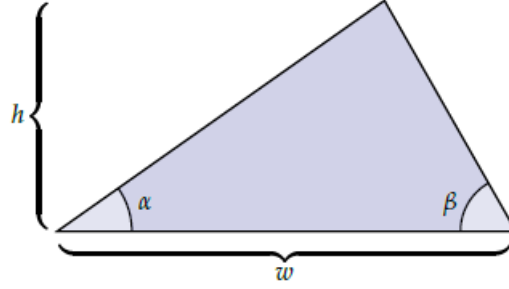


Figure 10.1: [Cra20, p. 105]

Proof. Drop an altitude from the top vertex of the triangle so that it intersects the base. Let x denote the length from this intersection point to the right vertex of the triangle. Then $\cot(\alpha) = \frac{w-x}{h}$ and $\cot(\beta) = \frac{x}{h}$, so $\cot(\alpha) + \cot(\beta) = \frac{w}{h}$. \square

Lemma 10.23. (Cotan formula lemma 2).

If \mathbf{e} is the edge vector along the base of a triangle, and \mathbf{e}_\perp is the result of rotating \mathbf{e} counterclockwise by $\frac{\pi}{2}$, then the gradient of the hat function ϕ associated with the vertex opposite the side corresponding to \mathbf{e} is

$$\nabla\phi = \frac{\mathbf{e}_\perp}{2A},$$

where A is the area of the triangle.

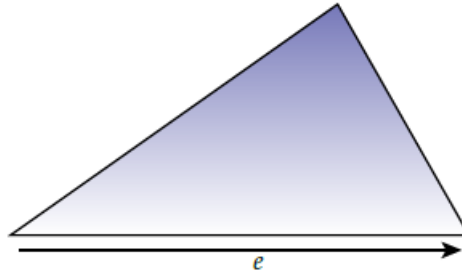


Figure 10.2: [Cra20, p. 106]

Proof. Denote the left, top, and right vertices of the triangle by \mathbf{x}_i , $i = 1, 2, 3$, respectively. Let $\phi = \phi_2$ denote the hat function associated with \mathbf{x}_2 , the top vertex. We have $\phi(\mathbf{x}_1) = \phi(\mathbf{x}_3) = 0$ and $\phi(\mathbf{x}_2) = 1$.

Since ϕ is linear, it is of the form $\phi(\mathbf{x}) = \phi(\mathbf{p}) + (\nabla_{\mathbf{x}}\phi)|_{\mathbf{p}} \cdot (\mathbf{x} - \mathbf{p})$, for any $\mathbf{p} \in \mathbb{R}^3$. (Here, \cdot is the dot product). It follows that the gradient of ϕ is the constant function $\nabla_{\mathbf{x}}\phi \equiv (\nabla_{\mathbf{x}}\phi)|_{\mathbf{p}}$. Using $\mathbf{p} = \mathbf{x}_2$, we have

$$\phi(\mathbf{x}) = 1 + (\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x} - \mathbf{x}_2).$$

Now use $\mathbf{x} = \mathbf{x}_1$ and $\mathbf{x} = \mathbf{x}_3$ to obtain

$$\begin{aligned} 0 &= 1 + (\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_1 - \mathbf{x}_2) \\ 0 &= 1 + (\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_3 - \mathbf{x}_2). \end{aligned}$$

These equations are equivalent to

$$\begin{aligned} 1 &= (\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_2 - \mathbf{x}_1) \\ 1 &= (\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_2 - \mathbf{x}_3). \end{aligned}$$

Subtract the second equation from the first and use the linearity of $\nabla_{\mathbf{x}}$ to obtain

$$(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_3 - \mathbf{x}_1) = 0.$$

Thus, $(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}$ is perpendicular to $\mathbf{e} = \mathbf{x}_3 - \mathbf{x}_1$, which is the edge vector of the base of the depicted triangle. Since $(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}$ is the direction of greatest *increase*, not decrease, in ϕ , then $(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}$ must point *into* the triangle rather than out. (ϕ increases as we go “into” the triangle, since $\phi(\mathbf{x}_1) = \phi(\mathbf{x}_3) = 0 < \phi(\mathbf{x}_2) = 1$). This implies that the direction of $(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}$ is the same as the direction of the \mathbf{e}_{\perp} described in the theorem.

To complete the proof, we will find the magnitude of $(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}$ and then show $\nabla_{\mathbf{x}}\phi \equiv (\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2} = \frac{\mathbf{e}_{\perp}}{2A}$.

Consider the previous equation

$$1 = (\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_2 - \mathbf{x}_1).$$

The dot product on the right hand side can be re-expressed as

$$1 = \|(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}\| \|\mathbf{x}_2 - \mathbf{x}_1\| \cos\left(\frac{\pi}{2} - \alpha\right),$$

where α is the angle between $(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}$ and $\mathbf{x}_2 - \mathbf{x}_1$. Equivalently, α is the angle between \mathbf{e} and the left side of the triangle. Using this reinterpretation of α , we solve for $\|(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}\|$ and get

$$\|(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}\| = \frac{1}{\|\mathbf{x}_2 - \mathbf{x}_1\| \cos\left(\frac{\pi}{2} - \alpha\right)} = \frac{1}{\|\mathbf{x}_2 - \mathbf{x}_1\| \sin(\alpha)} = \frac{1}{h},$$

where h is the height of the triangle.

Putting everything together, and using that $\|\mathbf{e}_{\perp}\| = \|\mathbf{e}\|$, we have

$$\nabla_{\mathbf{x}}\phi \equiv (\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2} = \|(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}\| \hat{\mathbf{e}}_{\perp} = \|(\nabla_{\mathbf{x}}\phi)|_{\mathbf{x}_2}\| \frac{\mathbf{e}_{\perp}}{\|\mathbf{e}_{\perp}\|} = \frac{1}{h\|\mathbf{e}\|} \mathbf{e}_{\perp} = \frac{1}{2(\frac{1}{2}h\|\mathbf{e}\|)} \mathbf{e}_{\perp} = \frac{\mathbf{e}_{\perp}}{2A}.$$

□

Lemma 10.24. (Cotan formula lemma 3).

The L^2 inner product of the gradient of a hat function ϕ corresponding to some vertex is

$$\langle \nabla \phi, \nabla \phi \rangle = \frac{1}{2} \left(\cot(\alpha) + \cot(\beta) \right),$$

where α and β are the interior angles at the other two vertices.

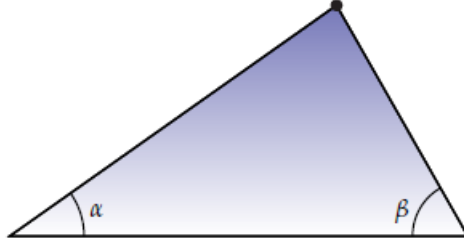


Figure 10.3: [Cra20, p. 106]

Proof. We have

$$\langle \nabla \phi, \nabla \phi \rangle = \int_{\text{triangle}} (\nabla \phi) \cdot (\nabla \phi) = \int_{\text{triangle}} \|\nabla \phi\|^2 = A \|\nabla \phi\|^2,$$

where A is the area of the triangle. In the proof of the previous theorem, we showed $\|\nabla \phi\| = \frac{1}{h}$, where h is the height of the triangle. So

$$\langle \nabla \phi, \nabla \phi \rangle = \frac{A}{h^2} = \frac{\frac{1}{2} \|\mathbf{e}\| h}{h^2} = \frac{1}{2} \frac{\|\mathbf{e}\|}{h}.$$

Apply Lemma 10.22, which states that the width-height ratio $\frac{\|\mathbf{e}\|}{h}$ is $\frac{\|\mathbf{e}\|}{h} = \cot(\alpha) + \cot(\beta)$, to obtain the result. \square

Lemma 10.25. (Cotan formula lemma 4).

Consider the hat functions $\phi_{\mathbf{p}_i}$ and $\phi_{\mathbf{p}_j}$ corresponding to vertices \mathbf{p}_i and \mathbf{p}_j of the same triangle. We have

$$\langle \nabla \phi_{\mathbf{p}_i}, \nabla \phi_{\mathbf{p}_j} \rangle = -\frac{1}{2} \cot(\theta),$$

where θ is the angle between opposing edge vectors.

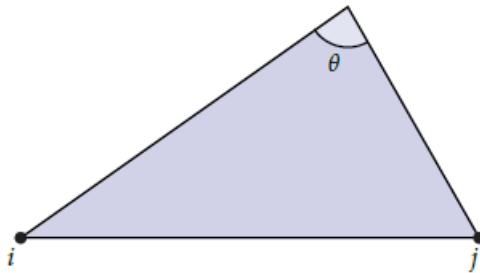


Figure 10.4: [Cra20, p. 106]

Proof. We have

$$\langle \nabla \phi_{\mathbf{p}_i}, \nabla \phi_{\mathbf{p}_j} \rangle = \int_{\text{triangle}} (\nabla \phi_{\mathbf{p}_i}) \cdot (\nabla \phi_{\mathbf{p}_i}) = A(\nabla \phi_{\mathbf{p}_i}) \cdot (\nabla \phi_{\mathbf{p}_i}),$$

where A is the area of the triangle. Applying Lemma 10.23, we have $\nabla \phi_{\mathbf{p}_i} = \frac{(\mathbf{e}_{\mathbf{p}_i})_{\perp}}{2A}$ and $\nabla \phi_{\mathbf{p}_j} = \frac{-(\mathbf{e}_{\mathbf{p}_j})_{\perp}}{2A}$, so

$$\langle \nabla \phi_{\mathbf{p}_i}, \nabla \phi_{\mathbf{p}_j} \rangle = A \frac{(\mathbf{e}_{\mathbf{p}_i})_{\perp}}{2A} \cdot \frac{-(\mathbf{e}_{\mathbf{p}_j})_{\perp}}{2A} = -\frac{\mathbf{e}_{\mathbf{p}_i} \cdot \mathbf{e}_{\mathbf{p}_j}}{4A} = -\frac{\|\mathbf{e}_{\mathbf{p}_i}\| \|\mathbf{e}_{\mathbf{p}_j}\| \cos(\theta)}{4A}.$$

To complete the proof, we do some trigonometry. Let h be the height of the triangle, let α be the angle at vertex \mathbf{p}_i and let β be the angle at vertex \mathbf{p}_j . Then

$$\langle \nabla \phi_{\mathbf{p}_i}, \nabla \phi_{\mathbf{p}_j} \rangle = -\frac{\|\mathbf{e}_{\mathbf{p}_i}\| \|\mathbf{e}_{\mathbf{p}_j}\| \cos(\theta)}{4A} = -\frac{\frac{h}{\sin(\beta)} \frac{h}{\sin(\alpha)} \cos(\theta)}{4(\frac{1}{2}wh)} = -\frac{h \cos(\theta)}{2w \sin(\alpha) \sin(\beta)}$$

where w is the length of the base of the triangle. From Lemma 10.22, we know $w = h(\cot(\alpha) + \cot(\beta))$. Thus, the above becomes

$$\begin{aligned} \langle \nabla \phi_{\mathbf{p}_i}, \nabla \phi_{\mathbf{p}_j} \rangle &= -\frac{\cos(\theta)}{2(\cot(\alpha) + \cot(\beta)) \sin(\alpha) \sin(\beta)} = -\frac{\cos(\theta)}{2\left(\frac{\cos(\alpha)}{\sin(\alpha)} + \frac{\cos(\beta)}{\sin(\beta)}\right) \sin(\alpha) \sin(\beta)} \\ &= -\frac{\cos(\theta)}{2(\cos(\alpha) \sin(\beta) + \cos(\beta) \sin(\alpha))} = -\frac{\cos(\theta)}{2 \sin(\alpha + \beta)}. \end{aligned}$$

Since $\alpha + \beta = \theta$, we obtain the claimed result. \square

Theorem 10.26. (Cotan formula).

$$(\Delta \tilde{f})|_{\mathbf{p}_i} = \frac{1}{2} \sum_j \left(\cot(\alpha_j) - \cot(\beta_j) \right) (\tilde{f}|_{\mathbf{p}_i} - \tilde{f}|_{\mathbf{p}_j})$$

Theorem 10.27. (Cotangent of the angle between vectors).

Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$, and consider the angle θ between \mathbf{v}_1 and \mathbf{v}_2 . We have

$$\cot(\theta) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1 \times \mathbf{v}_2\|}.$$

Proof. By definition of \cot , $\cot(\theta) = \frac{\|(\mathbf{v}_1)_{\parallel}\|}{\|(\mathbf{v}_1)_{\perp}\|}$. Using Lemma 2.98, we have $\|(\mathbf{v}_1)_{\parallel}\| = \|\text{proj}(\mathbf{v}_1 \rightarrow \mathbf{v}_2)\| = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|}$. Also, $\|\mathbf{v}_1 \times \mathbf{v}_2\| = \|(\mathbf{v}_1)_{\perp}\| \|\mathbf{v}_2\|$, so $\|(\mathbf{v}_1)_{\perp}\| = \frac{\|\mathbf{v}_1 \times \mathbf{v}_2\|}{\|\mathbf{v}_2\|}$. Combining these two facts, we obtain $\cot(\theta) = \frac{(\mathbf{v}_1 \cdot \mathbf{v}_2) / \|\mathbf{v}_2\|}{\|\mathbf{v}_1 \times \mathbf{v}_2\| / \|\mathbf{v}_2\|}$, so $\cot(\theta) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1 \times \mathbf{v}_2\|}$. \square

Discrete differential geometry

Theorem 10.28. (Cotan formula re-expressed with discrete differential geometry).

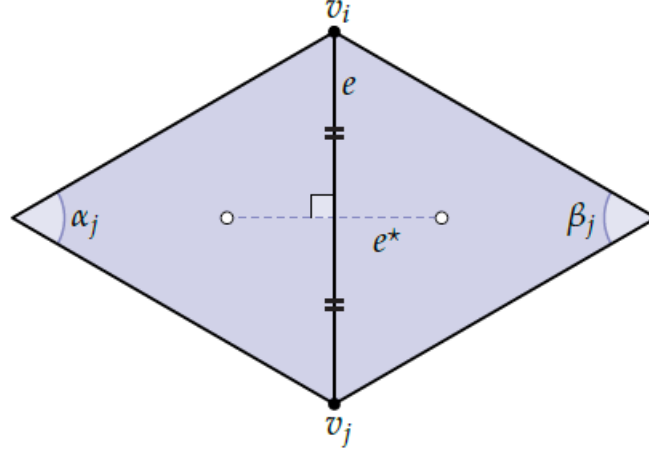


Figure 10.5: The diagram, which is from [Cra20, p. 110], uses e and e^* to denote e_{ij} and e_{ij}^* .

Since Lemma 10.22 shows that the width-height ratio $\frac{2|e_{ij}^*|}{|e_{ij}|}$ is $\frac{2|e_{ij}^*|}{|e_{ij}|} = \cot(\alpha_j) + \cot(\beta_j)$, we again obtain the “cotan formula” presented in Theorem [...]:

$$\boxed{\frac{|e_{ij}^*|}{|e_{ij}|} = \frac{1}{2}(\cot(\alpha_j) + \cot(\beta_j))}$$

Theorem 10.29. (The Laplacian of a test function via discrete differential geometry).

$$(\Delta \tilde{f})|_{\mathbf{p}_i} = \frac{1}{2} \sum_j \left(\cot(\alpha_j) - \cot(\beta_j) \right) (\tilde{f}|_{\mathbf{p}_i} - \tilde{f}|_{\mathbf{p}_j}) = \frac{1}{2} \sum_j \left(\frac{|e_{ij}^*|}{|e_{ij}|} \right) (\tilde{f}|_{\mathbf{p}_i} - \tilde{f}|_{\mathbf{p}_j})$$

Solving the problem

With FEM: if the edge lengths are all the same, then the cotangent of the angle between any two edge vectors is a bilinear function. More specifically, since the cotangent produces a scalar, the cotangent of the angle between edges is a bilinear form. We learned in Chapter 4 that a bilinear form can be identified with a $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ tensor. This means we can produce a matrix which describes the action of the Laplacian at

With DDG:

10.3 Surface parameterization

10.3.1 Outline

- Let $\mathbf{F} : M \rightarrow \mathbb{R}^3$ be a smooth injection. For smooth $z : M \rightarrow \mathbb{C}$, we want to require $\hat{\mathbf{n}} \times d\mathbf{F}(v_{\mathbf{p}}) = idz(v_{\mathbf{p}})$ for all $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$, so that the notion of angle inherited from \mathbb{R}^3 by the immersed surface $\mathbf{F}(M) \subseteq \mathbb{R}^3$ is the same as the notion of angle in \mathbb{C} . Notice that there must be some $\mathcal{J} : T_{\mathbf{p}}(M) \rightarrow T_{\mathbf{p}}(M)$ for which $\hat{\mathbf{n}} \times d\mathbf{F}(v_{\mathbf{p}}) = d\mathbf{F}(\mathcal{J}(v_{\mathbf{p}}))$. Thus, we can express the condition we wanted, “ $\hat{\mathbf{n}} \times df(v_{\mathbf{p}}) = idz(v_{\mathbf{p}})$ for all $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ ”, as “ $d\mathbf{F}(\mathcal{J}(v_{\mathbf{p}})) = idz(v_{\mathbf{p}})$ for all $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ ”. This is the *Cauchy-Riemann equation*.
- Introduce Hodge dual for complex differential forms.
 - $\omega(v_{\mathbf{p}}, ?) = *\omega(v_{\mathbf{p}}, \mathcal{J}(v_{\mathbf{p}}))$, so $*$ somehow takes the place of \mathcal{J}
- Cauchy-Riemann equation is restated as $*dz = idz$.

- Now, the goal is to compute z 's that solve the Cauchy-Riemann equation. Leads to minimization problem $\min_z || * dz - idz ||$. We call $E_C(z) = || * dz - idz ||$ the *conformal energy*. We show that for a piecewise linear map z on a simplicial disc M , we have $|| * dz - idz || = \frac{1}{2} \langle \Delta z, z \rangle - A(z)$, where $A(z)$ is the signed area of the image $z(M)$.
 - The inner product of complex-valued one forms on M is $\langle \alpha, \beta \rangle = \text{Re} \int_M * \bar{\alpha} \wedge \beta$
 - * How does this inner product work for real-valued zero forms?
 - * I think I've shown $||a + bi|| = ||a||^2 + 2\text{Re}(\langle a, i \rangle) + ||b||^2$
- Somehow, we decide we need more constraints. So we solve the problem

$$\begin{aligned} \min_z E_C(z) \\ \text{s.t. } ||z|| = 1 \text{ for all } z \\ \langle z, \mathbf{1} \rangle = 0 \text{ for all } z \end{aligned}$$

where $E_C(z) = c || * dz - idz || = c(\frac{1}{2} \langle \Delta z, z \rangle - A(z))$ for some scalar $c \in \mathbb{R}$. If we ignore the third condition, this problem is an eigenvalue problem. (Not obvious that E_C is linear- is it?)

10.3.2 Hermitian manifolds

Definition 10.30. (Complex manifold).

Definition 10.31. (Complex-valued differential forms).

Definition 10.32. (Hermitian inner product).

sesquilinearity- look at lin alg project

Definition 10.33. (Hermitian manifold).

Hermitian metric

Hermitian manifold

Definition 10.34. (Complex Hodge dual).

$$* : \Omega^k(M) \rightarrow \Omega^{n-k}(M)$$

Do the same thing as with regular Hodge dual, just change it so that we end up with $\omega \wedge * \nu = g(\omega, \bar{\nu}) dV_g$

<http://math.uchicago.edu/~may/REU2018/REUPapers/Sheagren.pdf>, p. 13

Definition 10.35. (Hermitian inner product on complex-valued differential forms).

same as with inner product on differential forms

$$\langle \langle \omega, \nu \rangle \rangle := \int_M g(\omega, \nu) dV_g = \int_M \omega \wedge * \bar{\nu}$$

Part V

Appendix

Different definitions of tensor

$\binom{p}{q}$ tensors

$\overline{\otimes}$, $\overline{\wedge}$ and interp as multilinear maps

Exterior powers

Pushforward and pullback

(Pullback). Originally was $(\otimes_q \mathbf{f}^*)(\phi^1 \otimes \dots \otimes \phi^q) = \mathbf{f}^*(\phi^1) \otimes \dots \otimes \mathbf{f}^*(\phi^q)$, extended linearly. Now have $(\overline{\otimes}_q \mathbf{f}^*)(\phi^1 \otimes \dots \otimes \phi^q) = (\mathbf{f}^*(\phi^1) \overline{\otimes} \dots \overline{\otimes} \mathbf{f}^*(\phi^q)) = ((\phi^1 \circ \mathbf{f}) \overline{\otimes} \dots \overline{\otimes} (\phi^q \circ \mathbf{f})) = ((\phi^1 \circ \mathbf{f}) \overline{\otimes} \dots \overline{\otimes} (\phi^q \circ \mathbf{f})) = (\phi^1 \circ \overline{\otimes} \dots \overline{\otimes} \phi^q) \circ \mathbf{f}$. Check this last equality by evaluating both sides on $(\mathbf{v}_1, \dots, \mathbf{v}_q)$. Extending linearly, we see $(\overline{\otimes}_q \mathbf{f}^*)(\mathbf{T}) = \mathbf{T} \circ \mathbf{f} = \mathbf{f}^*(\mathbf{T})$. That is, $(\overline{\otimes}_q \mathbf{f}^*)(\mathbf{T})(\mathbf{v}_1, \dots, \mathbf{v}_q) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), \dots, \mathbf{f}(\mathbf{v}_q))$.

Lemma 11.1. (Pushforward on the dual).

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, \dots, \epsilon^n\}$ be the induced dual basis for V^* . For all $\phi^1, \dots, \phi^n \in V^*$, we have

$$\phi^1 \wedge \dots \wedge \phi^n = \det(\mathbf{f}^*) \epsilon^1 \wedge \dots \wedge \epsilon^n$$

from Theorem [...]. Since $\det(\mathbf{f}^*)$ is the determinant of the matrix of \mathbf{f}^* relative to E^* and E^* , which has ij entry $([\phi_i]_{E^*})_j = \phi^i(\mathbf{e}_j)$ (recall Theorem 4.26), then $\det(\mathbf{f}^*) = \det(\phi^i(\mathbf{e}_j))$. Here, $(\phi^i(\mathbf{e}_j))$ is the matrix with ij entry $\phi^i(\mathbf{e}_j)$. Thus

$$\phi^1 \wedge \dots \wedge \phi^n = \det(\phi^i(\mathbf{e}_j)) \epsilon^1 \wedge \dots \wedge \epsilon^n.$$

As a sidenote, consider the isomorphism $\mathbf{F} : V^* \rightarrow \mathbb{R}^n$ induced¹ by the isomorphism $V \rightarrow \mathbb{R}^n$ that sends $\mathbf{e}_i \mapsto \hat{\mathbf{e}}_i$, where $\hat{\mathbf{e}} = \{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n\}$ is the standard basis for \mathbb{R}^n . The determinant in the above computes the projected volume of the parallelepiped spanned by $\mathbf{F}(\phi^1), \dots, \mathbf{F}(\phi^n) \in \mathbb{R}^n$ onto the parallelepiped spanned by $\mathbf{F}(\epsilon^1), \dots, \mathbf{F}(\epsilon^n) \in \mathbb{R}^n$.

Lemma 11.2. (Action of dual n -wedge on dual basis).

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, \dots, \epsilon^n\}$ be the induced dual basis for V^* . Combining the previous two lemmas, we see

$$(\epsilon^1 \overline{\wedge} \dots \overline{\wedge} \epsilon^n)(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1.$$

Proof. Using a similar argument to the one that showed the permutation formula for the determinant on the ϵ^i , we have

¹Specifically, if $\mathbf{G} : V^* \rightarrow V$ by $\mathbf{G}(\epsilon^i) = \mathbf{e}_i$ and $\mathbf{H} : V \rightarrow \mathbb{R}^n$ by $\mathbf{H}(\mathbf{e}_i) = \hat{\mathbf{e}}_i$, then $\mathbf{F} = \mathbf{H} \circ \mathbf{G}$.

$$\epsilon^1 \bar{\wedge} \dots \bar{\wedge} \epsilon^n = \overline{\text{alt}}(\epsilon^1 \bar{\otimes} \dots \bar{\otimes} \epsilon^n) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \epsilon^{\sigma(1)} \bar{\otimes} \dots \bar{\otimes} \epsilon^{\sigma(n)}.$$

Here, $\overline{\text{alt}}$ denotes the map [...] induced by alt .
Therefore

$$(\epsilon^1 \bar{\wedge} \dots \bar{\wedge} \epsilon^n)(\mathbf{e}_1, \dots, \mathbf{e}_n) = \sum_{\sigma \in S_n} \left(\text{sgn}(\sigma) \epsilon^{\sigma(1)} \bar{\otimes} \dots \bar{\otimes} \epsilon^{\sigma(n)} \right)(\mathbf{e}_1, \dots, \mathbf{e}_n) = \sum_{\sigma \in S_n} \left(\text{sgn}(\sigma) (\epsilon^{\sigma(1)} \bar{\otimes} \dots \bar{\otimes} \epsilon^{\sigma(n)})(\mathbf{e}_1, \dots, \mathbf{e}_n) \right).$$

Now we focus on the inner term, $(\epsilon^{\sigma(1)} \bar{\otimes} \dots \bar{\otimes} \epsilon^{\sigma(n)})(\mathbf{e}_1, \dots, \mathbf{e}_n)$. By definition of $\bar{\otimes}$, we have

$$(\epsilon^{\sigma(1)} \bar{\otimes} \dots \bar{\otimes} \epsilon^{\sigma(n)})(\mathbf{e}_1, \dots, \mathbf{e}_n) = \epsilon^{\sigma(1)}(\mathbf{e}_1) \dots \epsilon^{\sigma(n)}(\mathbf{e}_n)$$

Since $\epsilon^{\sigma(i)}(\mathbf{e}_j) = \delta_j^{\sigma(i)}$, the only permutation $\sigma \in S_n$ for which the above expression is nonzero is the identity permutation i ; when $\sigma = i$, the above is 1. Thus, we have

$$(\epsilon^1 \bar{\wedge} \dots \bar{\wedge} \epsilon^n)(\mathbf{e}_1, \dots, \mathbf{e}_n) = \text{sgn}(i) \cdot 1 = 1 \cdot 1 = 1.$$

□

Theorem 11.3. (Action of dual k -wedge on vectors).

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, \dots, \epsilon^n\}$ be the induced dual basis for V^* . Consider an element $\phi^1 \bar{\wedge} \dots \bar{\wedge} \phi^n$ in the top exterior power $\Lambda^n(V)$ of V . We have

$$(\phi^1 \bar{\wedge} \dots \bar{\wedge} \phi^n)(\mathbf{e}_1, \dots, \mathbf{e}_n) = \det(\phi^i(\mathbf{e}_j)).$$

Recall from Lemma [...] that if $\mathbf{F} : V^* \rightarrow \mathbb{R}^n$ is the isomorphism induced by the isomorphism $V \rightarrow \mathbb{R}^n$ that sends $\epsilon^i \rightarrow \hat{\mathbf{e}}_i$, the determinant in the above computes the projected volume of the parallelepiped spanned by $\mathbf{F}(\phi^1), \dots, \mathbf{F}(\phi^n) \in \mathbb{R}^n$ onto the parallelepiped spanned by $\mathbf{F}(\epsilon^1), \dots, \mathbf{F}(\epsilon^n) \in \mathbb{R}^n$.

11.0.1 Differential forms

Theorem 11.4. (Pullback of a differential form).

Recall from [...] that if W and Z are vector spaces of dimension n and $\mathbf{g} : W \rightarrow Z$ is a linear map, then the pullback $\Lambda^k \mathbf{g}^* : \Lambda^k(Z^*) \rightarrow \Lambda^k(W^*)$ is defined by $\mathbf{g}^*(\mathbf{T}) = \mathbf{T} \circ \mathbf{g} = \mathbf{g}^*(\mathbf{T})$. Therefore, the above definition implies

$$\left((\bar{\Omega}^k \mathbf{f}^*)(\omega) \right)_{\mathbf{p}} = \omega_{\mathbf{f}(\mathbf{p})} \circ d\mathbf{f}_{\mathbf{p}} = (d\mathbf{f}_{\mathbf{p}})^*(\omega_{\mathbf{f}(\mathbf{p})}).$$

Therefore

$$(\bar{\Omega}^k \mathbf{f}^*)(\omega) = \omega_{\mathbf{f}(\mathbf{p})} \circ d\mathbf{f}_{\mathbf{p}}.$$

More explicitly, $(\bar{\Omega}^k \mathbf{f}^*)(\omega)$ is defined at each point $\mathbf{p} \in M$ by

$$\left((\bar{\Omega}^k \mathbf{f}^*)(\omega) \right)_{\mathbf{p}}(\mathbf{v}_1, \dots, \mathbf{v}_k) = \omega_{\mathbf{f}(\mathbf{p})}(d\mathbf{f}_{\mathbf{p}}(\mathbf{v}_1), \dots, d\mathbf{f}_{\mathbf{p}}(\mathbf{v}_k)).$$

12

Tensors in physics and engineering

Choice of basis

Cauchy stress tensor

Outer product

Theorem 12.1. If we have identified $V \cong V^*$, then every linear transformation $V \rightarrow W$ has a matrix of the form $\sum_k \mathbf{w}_k \mathbf{v}_k^\top = \mathbf{W}^\top \mathbf{V}$.

Proof. By the discussion on rank 1 tensors above, every element of $\text{Hom}(V, W)$ has a matrix of the form $\sum_k [\mathbf{w}_k]_F [\mathbf{v}_k]_E^\top$, where $[\mathbf{v}_k]_F = (v_{k1}, \dots, v_{kn})^\top$, $[\mathbf{w}_k]_E = (w_{k1}, \dots, w_{kn})^\top$. The matrix $\sum_k [\mathbf{w}_k]_F [\mathbf{v}_k]_E^\top$ has ij entry $\sum_k w_{ki} v_{kj}$. Let \mathbf{W} be the matrix whose i th column is $[\mathbf{w}_i]_F$ and \mathbf{V} be the matrix whose i th column is $[\mathbf{v}_i]_E$. Then the ij entry of $\sum_k [\mathbf{w}_k]_F [\mathbf{v}_k]_E^\top$ is $\sum_k w_{ki} v_{kj} = \mathbf{W}_{:i} \cdot \mathbf{V}_{:j} = \mathbf{W}_{i:}^\top \cdot \mathbf{V}_{:j} = ij$ entry of $\mathbf{W}^\top \mathbf{V}$. Therefore $\sum_k [\mathbf{v}_k]_E [\mathbf{w}_k]_F^\top = \mathbf{W}^\top \mathbf{V}$. \square

Levi-Civita symbol

Bilinear forms, metric tensors, and coordinates of tensors

Definition 12.2. (The physicist's index notation, the physicist's slanted index notation).

This definition contains a series of definitions used by physicists. The content of this definition will not be used anywhere else in this book; we present it for completeness only.

Let V be an n -dimensional vector space, let E be a basis for V , and let E^* be the induced dual basis for V^* . We define $v^i := ([\mathbf{v}]_E)^i$ and $\phi_i := ([\phi]_{E^*})_i$.

Now let V and W be finite-dimensional vector spaces with bases E and F , and let $E^* = \{\phi^{e_1}, \dots, \phi^{e_n}\}$ and $F^* = \{\psi^{f_1}, \dots, \psi^{f_m}\}$ be corresponding induced dual bases of V^* and W^* . Suppose that we also have a metric tensor g on V and W , so $V \cong W^*$ naturally via the musical isomorphism $b_1 : V \rightarrow W^*$ that is induced by g . In this situation, we define $v_i := (\mathbf{v}^{b_1})_i$.

Now that these definitions are out of the way, we consider the situation most often seen in physics: suppose $V = W$, and fix a basis E for V (fixing a basis allows us to use the notation v_i).

Theorem 4.30 is then restated as

$$v^i = \sum_{j=1}^n g^{ij} v_j$$

$$v_i = \sum_{j=1}^n g_{ij} v^j.$$

Due to the above two equations, one converts between the coordinates of a vector and the coordinates of its corresponding dual vector by “multiplying by the metric”. Though, a physicist would really say that “one converts between upper indices and lower indices by multiplying by the metric”.

(Physicists also use Einstein summation notation, so this operation looks more like straightforward multiplication because, in Einstein summation notation, the sum is “implied”, and not written).

As was shown out in the previous theorem, we can use the above “multiplication by the metric” (i.e. we can use one of the previous two equations) to convert the coordinates of a $\binom{p}{q}$ tensor to the coordinates of a $\binom{p-1}{q+1}$ or $\binom{p+1}{q-1}$ tensor. Since we have already set up the index notation v^i and v_i , and shown how to “raise and lower” indices, we define a similar notation that involves the raising and lowering of individual indices for coordinates of $\binom{p}{q}$ tensors. For simplicity, we will introduce this index notation for tensors for the case of $\binom{0}{2}$ tensors. The notation easily generalizes to $\binom{p}{q}$ tensors for any p, q .

Consider the $\binom{0}{2}$ tensor with coordinates (T_{ij}) . We define

$$T^i{}_j := \sum_{k=1}^n g^{ik} T_{kj}$$

$$T_j{}^i := \sum_{k=1}^n g^{ik} T_{jk}$$

Recall from the equation $v^i = \sum_{j=1}^n g^{ij} v_j$ that the index which gets summed over is the index that gets raised. Also notice that if one index appears to the left of another index before either of the indices is raised, then that same index must appear to the left of the other index after either of the indices is raised. So, for instance, in the first equation, i appears before j before i is raised because the coordinates of the tensor are (T_{ij}) . Correspondingly, i appears before j after i is raised in the coordinates $(T^i{}_j)$.

The above definition implies that we can lower indices as follows:

$$T_{ij} = \sum_{k=1}^n g_{jk} T^k{}_i$$

$$T_{ji} = \sum_{k=1}^n g_{jk} T^k{}_i.$$

Similarly as to before, the equation $v_i = \sum_{j=1}^n g_{ij} v^j$ dictates that the index which gets summed over is the index that gets lowered.

We have not explicitly presented how this notation works for $\binom{2}{0}$ tensors or for arbitrary $\binom{p}{q}$ tensors. It would be awkward to fully write out a general description of this notation, due to the need for a notation for an arbitrary arrangement of upper and lower indices. The idea generalizes straightforwardly, though, so this is no issue.

Suppose \mathbf{T} is a $\binom{p}{q}$ tensor with an almost-arbitrary arrangement of upper and lower indices (think of a $\binom{p}{q}$ tensor with a “random” arrangement of upper and lower indices, such as $(T^{i_1 i_2}{}_{j_1}{}^{i_3}{}_{j_2}{}^{i_4 i_5}{}_{j_3})$); we only require that there be at least two lower indices and at least two upper indices. A lower index j_k is raised by taking the coordinates of \mathbf{T} , replacing j_k with the dummy index r , and then computing $\sum_{r=1}^n g^{ur} \cdot$ (the coordinates of \mathbf{T}), where u_ℓ is some other lower index. An upper index i_k is raised by taking the coordinates of \mathbf{T} , replacing i_k with the dummy index r , and then computing $\sum_{r=1}^n g_{v_\ell r} \cdot$ (the coordinates of \mathbf{T}), where v_ℓ is some other upper index.

Double contraction

The next definition and theorem addresses the *double contraction* : of $T_2^2(V)$ tensors. The double contraction is often used in physics contexts. We present the double contraction only for completeness; we do not use it in this book.

Definition 12.3. (Double contraction).

Consider $T_2^2(V) \otimes T_2^2(V)$, and note that $T_2^2(V) \otimes T_2^2(V) \cong T_4^4(V)$ naturally. The *double contraction* : is the map $T_2^2(V) \otimes T_2^2(V) \xrightarrow{\cdot} T_0^4(V)$ induced by the composition of a $\binom{3}{1}$ contraction and a $\binom{4}{2}$ contraction on $T_4^4(V)$. The double contraction is also called the *double dot product*.

Example 12.4. (Hooke's law).

13

Review of pedagogy

Test citation 1 [Lee], test citation 2 [HH], test citation 3 [Cra20], test citation 4 [War]

Bibliography

- [BW97] Javier Bonet and Richard D. Wood. Nonlinear Continuum Mechanics for Finite Element Analysis. Cambridge Univeristy Press, The address, 1 edition, 1997.
- [Cra20] Keenan Crane. Discrete Differential Geometry: An Applied Introduction. 1 edition, 4 2020. <https://www.cs.cmu.edu/~kmcrane/Projects/DDG/>.
- [GP74] Victor Guillemin and Alan Pollack. Differential Topology. Prentice-Hall, 1 edition, 1974.
- [HH] Barbara Burke Hubbard and John Hamal Hubbard. Vector Calculus, Linear Algebra, And Differential Forms. Prentice Hall, 1 edition.
- [Lee] John M. Lee. Introduction to Smooth Manifolds. Springer, 2 edition.
- [War] Frank W. Warner. Foundations of Differentiable Manifolds and Lie Groups. 1 edition.