Tensors and differential forms

Ross Grogan-Kaylor

November 2020

Acknowledgements

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

Most of all, thank you to `/u/DankKushala`, `/u/Tazerenix`, and `/u/ziggurism` for answering my questions about tensors and differential geometry on `reddit.com/r/math`. Much of my understanding of the concepts in this book is due to your excellent answers.

About this book

The goal of this book is to present generalizations of "vector calculus" concepts such as div, grad, curl, the derivative of a vector valued function of a vector variable, and vector calculus integral theorems.

Tentative prerequisites and reading advice for a student

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 take advantage of Chapter 2, the review chapter on linear algebra. My advice is to use Chapter 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 Chapter 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 be 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.

Contents

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.

Basic linear algebra

- K is used to denote a field.
- c, d, and k are used to denote elements of K.
- $K^{m \times n}$ is the set of $m \times n$ matrices with entries in K.
- n and m are used to denote nonnegative integers or positive integers.
- $\mathbf{c}$ is used to denote an element of K^n, and $\mathbf{d}$ is used to denote either another element of K^n or an element of K^m.
- V and W are used to denote vector spaces over a field K. When these spaces are finite-dimensional, we often set $\dim(V) = n$ and $\dim(W) = m$.
- $\mathbf{v}$ and $\mathbf{w}$ are used to denote elements of vector spaces.
- $E = \{\mathbf{e}_i\}_{i=1}^n$ is used to denote an arbitrary basis for V, and $F = \{\mathbf{f}_i\}_{i=1}^m$ is used to denote either another arbitrary basis for V or to denote an arbitrary basis for W.
- $U = \{\hat{\mathbf{u}}_i\}_{i=1}^n$ is used to denote orthonormal basis for some vector space, and $\widetilde{U} = \{\widetilde{\mathbf{u}}_i\}_{i=1}^n$ is used to denote another orthonormal basis for the same space.
- $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_i\}_{i=1}^n$ is the standard basis of K^n (that is, $\hat{\mathfrak{e}}_i$ is the tuple of entries from K whose jth entry is 1 when $j = i$ and 0 otherwise), and $\hat{\mathfrak{F}} = \{\hat{\mathfrak{f}}_i\}_{i=1}^m$ is the standard basis K^m (defined similarly).
- When applicable, $||\cdot||$ denotes the magnitude of a vector; for example, $||\mathbf{v}||$ is the magnitude of $\mathbf{v}$.
- Hats $\hat{}$ are used to denote unit vectors (vectors with a magnitude of 1); for example, $\hat{\mathbf{v}}$ denotes a unit vector.
- $[\mathbf{v}]_E$ is used to denote the vector that contains the coordinates of $\mathbf{v} \in V$ relative to the basis E.
- $\mathbf{f}$ is used to denote a linear function $V \to W$.
- $[\mathbf{f}(E)]_F$ denotes the $m \times n$ matrix of $\mathbf{f}$ with relative to the bases E and F of the finite-dimensional vector spaces V and W.
- $[\mathbf{f}(\hat{\mathfrak{E}})]_{\hat{\mathfrak{F}}} = \mathbf{f}(\hat{\mathfrak{E}})$ denotes the $m \times n$ matrix of the linear function $\mathbf{f} : K^n \to K^m$ relative to the standard bases $\hat{\mathfrak{E}}$ and $\hat{\mathfrak{F}}$ for K^n and K^m.
- $\cong$ is used to denote an isomorphism of vector spaces.

Multilinear algebra

- ϕ is used to denote an arbitrary element of V^*, and ψ is used to denote an arbitrary element of W^*.
- $E^* = \{\epsilon_i\}_{i=1}^n$ is an arbitrary basis for V^*, and $F^* = \{\delta_i\}_{i=1}^m$ is an arbitrary basis for W^*.
- $E^* = \{\phi^{\mathbf{e}_i}\}_{i=1}^n$ is the basis for V^* induced by the basis $E = \{\mathbf{e}_i\}_{i=1}^n$ for V, and $F^* = \{\psi^{\mathbf{f}_i}\}_{i=1}^m$ is the basis for W^* induced by the basis $F = \{\mathbf{f}_i\}_{i=1}^m$ for W.
- Φ is used to denote an element of V^{**}, and Ψ is used to denote an element of W^{**}.
- $E^{**} = \{\text{Ч}_i\}_{i=1}^n$ is an arbitrary basis of V^{**}, and $\{\Xi_i\}_{i=1}^m$ is an arbitrary basis of W^{**}. (This notation is never actually used in this book, but you might use this as a suggestion in investigations of the material).

- $\mathcal{L}(V_1 \times ... \times V_k \to W)$ is the vector space of k-linear functions $V_1 \times ... \times V_k \to W$. In particular, $\mathcal{L}(V \to W)$ is the vector space of linear functions $V \to W$.
- The set of (p, q) tensors on V is denoted $T_{p,q}(V)$, and $\mathbf{T}$ is used to denote an element of $T_{p,q}(V)$.
- B is used to denote a bilinear form on V and W.
- Given a bilinear form B on V and W, $\flat_1$ and $\flat_2$ are the natural linear maps $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ defined by $\flat_1(\mathbf{v})(\mathbf{w}) = B(\mathbf{v}, \mathbf{w})$ and $\flat_2(\mathbf{w})(\mathbf{v}) = B(\mathbf{v}, \mathbf{w})$. We define the notation $\mathbf{v}^{\flat_1} := \flat_1(\mathbf{v})$ and $\mathbf{w}^{\flat_2} := \flat_2(\mathbf{w})$. When $\flat_1$ and $\flat_2$ are isomorphisms, they are called *musical isomorphisms*, and their inverses $\flat_1^{-1}$ and $\flat_2^{-1}$ are denoted by $\sharp_1 := \flat_1^{-1}$ and $\sharp_2 := \flat_2^{-1}$.
- g is used to denote a metric tensor on V and W. (Our definition of metric tensor allows for $V \neq W$, and is also not such that there is a single metric tensor, as is sometimes the case in other conventions).
- If V and W have bases $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_m\}$, then $\mathbf{g}$ is used to denote the matrix $\mathbf{g} = (g_{ij}) := (g(\mathbf{e}_i, \mathbf{f}_j))$.
- $\widetilde{g}$ is used to denote the metric tensor on W^* and V^* induced by g, and $\widetilde{\mathbf{g}}$ are used to denote the matrix $\mathbf{g} = (g^{ij}) := (\widetilde{g}(\psi^{\mathbf{f}_i}, \phi^{\mathbf{e}_j}))$.
- The usual index conventions for *covariance and contravariance* are introduced in Definition 3.20. They are only used in Section 4.2, and not anywhere else in the book.
- The kth exterior power of V is denoted $\Lambda^k(V)$.
- Given a linear function $\mathbf{f} : V \to W$, the induced pushforward from $T_0^k(V)$ to $T_0^k(W)$ is denoted by $\otimes_0^k \mathbf{f}$, and the induced pullback from $T_k^0(W)$ to $T_k^0(V)$ is denoted by $\otimes_k^0 \mathbf{f}^*$.
- Given a linear function $\mathbf{f} : V \to W$, the induced pushforward from $\Lambda^k(V)$ to $\Lambda^k(W)$ is denoted by $\Lambda^k \mathbf{f}$, and the induced pullback from $\Lambda^k(W^*)$ to $\Lambda^k(V^*)$ is denoted by $\Lambda^k \mathbf{f}^*$.
- The standard notation for pushforwards and pullbacks is problematic in that it is ambiguous; see Remark 5.22.

Multilinear algebra: distinguishing between algebraic objects and functions

(p, q) tensors can be treated either as elements of a vector space that satisfy certain algebraic rules (this is ultimately due to Definition 3.4) or as elements of $\mathcal{L}((V^*)^{\times p} \times V^{\times q} \to K)$ (see Remark 3.41). We favor the first interpretation whenever possible, but it is sometimes necessary to use the second interpretation. Notation used for the second interpretation makes use of an overset $\sim$ to distingish it from notation used for the first interpretation.

For example, some notation favoring the first interpretation is $\otimes$, $T_{p,q}(V)$, $\Lambda^k(V)$, $\wedge$, $\Omega^k(M)$, and some notation favoring the second interpretation is $\widetilde{\otimes}$, $\widetilde{T}_{p,q}(V)$, $\widetilde{\Lambda}^k(V)$, $\widetilde{\wedge}$, $\widetilde{\Omega}^k(M)$.

Other uses for an overset $\sim$

An overset $\sim$ is not always used for the above reason. Sometimes it is used in the same manner that a "prime" ' would be used in other texts, which is to denote an object that is "similar" or "closely related" to a previous one. For example, if an other text said "Let $x, x' \in \mathbb{R}$", we would say, "Let $x, \widetilde{x} \in \mathbb{R}$".

Differential geometry

- M and N are used to denote manifolds, usually smooth manifolds.
- $\mathbf{x}$ and $\mathbf{y}$ are used to denote smooth charts on smooth manifolds.
- The acronym "WWBOC" is used as shorthand for "with or without boundary or corners".
- The set of differential k-forms on a manifold is denoted $\Omega^k(M)$, and ω is typically used to denote an element of $\Omega^k(M)$.

Misc

- As you may have noticed, vector quantities (including functions with a vector output) are bolded. Elements of vector spaces (except for "trivial" ones like $\mathbb{R}$, the field K, etc.) are considered to be vector quantities.
- $\sim$ is used to denote "rotational equivalence".
- δ^i_j is the Kronecker delta function defined by $\delta^i_j = 1$ when $i = j$ and $\delta^i_j = 0$ otherwise.
- Strikethrough notation is used to denote the omission of some element in a list. For example, $x_1, ..., \cancel{x_i}, ..., x_n$ is used to denote $x_1, ..., x_{i-1}, x_{i+1}, ..., x_n$.

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".

We can denote propositions with letters, and say things such as "let P be a proposition". When a proposition 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 propositions can be constructed from simpler propositions. Examples of some more complicated propositions are $\Big((3 > 4) \text{ and } (\text{every rectangle is a square})\Big) \cong F$ and $\Big((5 > -3) \text{ or } (2 > 100)\Big) \cong T$.

There are three fundamental operations on propositions that are used to build more complicated propositions 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 propositions P and Q as is expressed in the following "truth tables".

P	Q	P and Q
T	T	T
T	F	F
F	T	F
F	F	F

P	Q	P or Q
T	T	T
T	F	T
F	T	T
F	F	F

P	not P
T	F
F	T

So, P *and* Q evaluates to true only when both P and Q are true, P *or* Q evaluates to true whenever either of P or Q is true, and *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 propositions P and Q, we have the following two logical identities, called *DeMorgan's laws*:

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

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* is a proposition in which one of the entities is variable. (We have not formally defined functions yet). For example, $P(x) = (x > 3)$ is a predicate which is true for some values of x but false for others. We will sometimes informally refer to predicates 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 predicates:

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

The predicate $\forall x\ P(x)$ statement is true exactly when $P(x)$ is 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 true[1].

Nested quantifers

Suppose we have a predicate that has two inputs, $P(x, y)$. In the last section, applying a quantifier-variable pair to a predicate 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. For example, we can define a predicate $Q(y) \cong \forall x\ P(x, y)$.

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 P of two inputs, the four possible ways to "nest" quantifiers are as follows:

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

Always remember that the innermost pair of quantifier with proposition is a predicate.

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:

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

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

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

Negating quantifiers

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

$$\begin{gathered} \text{not}\Big(\forall x\ P(x)\Big) \cong \exists x \text{ not } P(x) \\ \text{not}\Big(\exists x\ P(x)\Big) \cong \forall x \text{ not } P(x). \end{gathered}$$

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

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 so that the above rules apply. For example, (not $(\forall x\ \exists y\ P(x,y))) \cong (\exists x\ \forall y$ 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 predicates 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, predicates, 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 unique objects. Examples of sets include $S_1 = \{\text{grass}, \text{tree}, -1, \pi\}, S_2 = \{0, 2, 4, 6, ...\}$ 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 $\{1, 2\} = \{2, 1\}$. Additionally, when we say that a set is considered to be a list of unique elements, we do not mean that a set *cannot* contain multiple copies of the same element[2], we just mean that those multiple copies "don't matter"; so, for example, we have $\{1, 1, 2, 2, 3, 3\} = \{1, 2, 3\}$.

Formally, the fact that the order of objects in a set doesn't matter is established by defining sets S and T to be equivalent exactly when S contains all of T's elements and T contains all of S's elements.

Constructing sets

$\{x \mid P(x)\}$ denotes the set of objects x which satisfy the property $P(x)$. The "$|$" symbol can be read as "such that"; we read "$\{x \mid P(x)\}$" out loud as "x such that P of x". Sometimes, authors use the notation $\{x : P(x)\}$ instead of $\{x \mid P(x)\}$.

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 an operation that accepts two predicates as input and produces a predicate as output. (Recall, a predicate is a statement that is either true or false and that makes use of variables to represent concepts). 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:

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

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, the "$\forall x$" technically indicates an inspection of *all* entities (such as "grass", the function f defined by $f(x) = 3x^2$, -1, etc.). We restrict our attention to the set S by using $\implies$, so that it is only possible for the predicate $(x \in S \implies P(x))$ to be false when $x \in S$.

[2]The meaning of the word "unique" used in the standard terminology, "unordered list of unique objects", is a bit ambiguous, so it is understandable if interpreted "unique" to mean "no copies allowed".

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 (there wasn't), then the predicate inside the "for all" would have evaluated as false for that square. (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 the proposition $\forall x\ P(x) \implies Q(x)$. Due to the reasoning "if P happened then Q must have happened", Q is said to be a *necessary condition* for P. Because of the reasoning "one of ways to make Q happen is to make P happen", P is said to be a *sufficient* condition for Q.

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 section before 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:

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

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).

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

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

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

The last above 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".

Common shorthand

- When people write something of the form "$\forall x\ P(x) \implies Q(x)$", they mean "$\Big(\forall x\ P(x) \implies Q(x)\Big) \cong T$".
- 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 extremly 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 $\Big(\forall x \; P(x) \implies Q(x)\Big) \cong \Big(\forall x \; Q(x) \implies P(x)\Big)$ 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

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

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", and is 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 $\cong$ and $\iff$ are the same operator. One way to verify this is by checking that the truth tables of $\cong$ and $\iff$ are the same.

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 contraction 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 (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 (not P) when the full context is considered. Proof by contradiction is just a more verbose way of proving the contrapositive, ((not Q) $\implies$ (not P)).

For some proofs, using the contrapositive in its raw logical 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 similarity to $\leq$ notation

Note that, for all sets S, the empty set is a subset of S, $\emptyset \subseteq S$, and S is a subset of itself, $S \subseteq S$.

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 $T \subset S$).

There is no universal set in ZFC

It seems natural that there would be a set which contains every other set, but not itself; it seems natural that there would be a so-called "universal set". However, this is not possible in ZFC. Why? Suppose there did exist such a "universal set" U. Since U contains all sets, U must contain itself: $U \in U$. But, by definition, U was assumed to not contain itself: $U \notin U$. This is a contradiction[3]. Thus, there does not exist a "set of all sets".

Common sets

We define notation for many common infinite sets.

$$\begin{aligned}
\mathbb{N} &:= \text{"the natural numbers"} = \{0, 1, 2, 3, ...\} \\
\mathbb{Z} &:= \text{"the integers"} = \{..., -3, -2, -1, 0, 1, 2, 3, ...\} \\
\mathbb{Q} &:= \text{"the rational numbers"} = \left\{ \frac{n}{m} \mid n, m \in \mathbb{Z} \right\} \\
\mathbb{R} &:= \text{"the real numbers"} = \{\text{all limits of sequences of rational numbers}\} \\
\mathbb{C} &:= \text{"the complex numbers"} = \{a + b\sqrt{-1} \mid a, b \in \mathbb{R}\}
\end{aligned}$$

We have $\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R} \subsetneq \mathbb{Q}$.

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 now define ways to build sets from a collection[4] of sets. Let I be an indexing set, and consider the collection $C = \{S_\alpha \mid \alpha \in I\}$.

The *union* of the sets in C is defined to be the result of "adding" all the sets in C together:

$$\bigcup_{\alpha \in I} S_\alpha := \{x \mid \exists \alpha \in I \; x \in S_\alpha\}.$$

The *intersection* of the sets in C is defined to be the result of considering only the elements that are common to every single set in C:

$$\bigcap_{\alpha \in I} S_\alpha := \{x \mid \forall \alpha \in I \; x \in S_\alpha\}.$$

[3]This paradox can be seen as motivation for the definition of a *class*. Essentially, classes are set-like objects that are not allowed to contain themselves. There *does* exist a class that contains all sets, but not itself

[4]The word "collection" is often used to mean "set of sets".

Set difference and set complement

Set difference

The union of two sets, introduced above, is analogous to addition of numbers. The following *set difference* operation is analogous to the subtraction of one number from another:

$$S - T := \{x \mid x \in S \text{ and } x \notin T\}.$$

Set complement

Sometimes, it is convenient to consider some set U to be the "largest possible set". For example, it is common to set U to be the unit interval, $U = [0, 1]$, when proving theorems from calculus. In this sort of situation, the *set complement (in U)* of some other set S is denoted S^C and is defined to be $S^C := U - S$.

Ordered pairs

Roughly speaking, ordered pairs are objects notated in the form "(x, y)" (an ordered pair consists of two objects, such as x and y, that are separated by a comma and are enclosed in parentheses) that are subject to the characterizing property $(x_1, y_1) = (x_2, y_2)$ iff $x_1 = y_1$ and $x_2 = y_2$.

Formally, we can *construct* the apparatus of ordered pairs by the definition

$$(x, y) := \{x, \{y\}\}.$$

The ordered pairs of this definition satisfy the characterizing property of ordered pairs mentioned above.

Cartesian product

nth Cartesian product

$((S \times S) \times S) \times ...) \times S$ vs. "$S \times ... \times S$"

1.5 Relations

inverse relations
inverse functions <=> 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).

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 $f : X \to Y$ be a function. A *left-inverse* of f is a function $\ell : Y \to X$ such that $\ell \circ f = I_X$, where I_X is the identity on X. A *right-inverse* of f is a function $r : Y \to X$ such that $f \circ r = I_Y$, where I_Y is the identity on Y.

If $f : X \to Y$ 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 $\ell = r$. □

Definition 1.12. (Proof by induction).
Defn?

For the future

- ”right hand side” -> ”right side”, same for ”left hand side”
- change definition differential to be via matrix relative to bases. within that definition, note that an arg similar to the one of the previous deriv yields the following formula.
- vector fields should be denoted as v, not $\mathbf{V}$, since a continuous map sending $\mathbf{p} \mapsto v_{\mathbf{p}}$ is a vector field
- linear alg
  - remove all uses of the explanation “$[\mathbf{f}(E)]_{\hat{\mathfrak{E}}}$ is the primitive matrix of $\mathbf{f}$ relative to $\hat{\mathfrak{E}}$”. change to only using $\mathbf{f}(E)$
  - in “(Matrix of a linear function $V \to V$ relative to one basis)”, remove the discussion that only works when $V = K^n$, and just use the expression for $[\mathbf{v}]_E$ that works in general; refer the reader to later deriv
- move four fundamental isos to right after defn of (p, q) tensor
- add remark about slanted indices after Defn 3.33, connect that to discussion before index lowering/raising theorem
- prove that (k, l) contraction is basis-independent. follows because, under a change of basis the vector (coorresp. to k) transforms via a matrix A and the dual vector (corresp. to l) transforms via a matrix B, where A and B are inverses
- fix uniqueness of adjoint remark: show that inner product condition implies that adjoint satisfies function composition condition
- rewrite Ricci law to go from E and E* to F and F?

This is a list of items that must be completed before I consider this book truly complete.

- Finish cross product section
- When Hodge-dual is used to generalize vector calculus theorems, remind reader that Hodge-dual was introduced in cross product section.

  Emphasize that Hodge-dual is basically a generalized cross product. This perspective implies the abstract definition of the Hodge-dual that involves orthogonal subspaces, which is often pulled out of thin air in other presentations. (So maybe prove that the Hodge-dual satisfies this condition involving orthogonal subspaces as a result of it satisfying cross-product like condition. Now that I think about it, this proof will be analogous to proving that $\mathbf{v} \times \mathbf{w}$ is perpendicular to $\mathbf{v}$ and $\mathbf{w}$ by using the defn of $\times$. This means that definition of Hodge dual will be in cross product section, so can remove definition near vector calculus stuff).
- Add a proof that “every orthogonal linear function on an n-dimensional inner product space with determinant 1 is a composition of ${}_nC_2$ many 2-rotations” to justify that the definition of n-rotation as a composition of 2-rotations is “correct”.
- Finish the logic and proofs chapter.
- Finish the commented-out systems of linear equations section in the linear algebra chapter.
- Add more explanatory text in-between definitions, lemmas, theorems where needed.
- Flesh out the calculus review chapter.
  - Actually define the various generalizations of derivative (and integral). See my own “Multivariable Calculus 2019” for this.

  - Fubini's theorem?
  - Add derivations of div and curl formulas; reference that guy's 2011 vector calculus notes (`http://www.supermath.info/CalculusIIIvectorcalculus2011.pdf`, p. 24 and 26).

- Figure out the unexplained step in Theorem 9.24. Somehow, $d\mathbf{x}\Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}}\Big) = \frac{\partial \mathbf{x}}{\partial \widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}}$.
- Finish the section in "Manifolds" on frames and coframes.
- Figure out why the "mnemonic" of the "Manifolds" chapter works.
- Add the divergence theorem and the less general Stokes' theorem (and Green's theorem) after the generalized Stokes' theorem.
- Revise/rewrite the expositions at the beginning of each chapter. The exposition at the beginning of "A motivated introduction to $\binom{p}{q}$" tensors is of a different style than the other chapters in that it doesn't give a preemptive outline of the entire chapter, and is more of a teaser. Make all expositions a blend between this style and an outline.
- Potentially relocate the section about differential forms in $\widetilde{\Omega}^k(M)$ that act on tangent vectors. Potentially the presentation of tensors/differential forms as actual (pointwise) multilinear functions.
- Add "Tensors in Engineering" section.
- Add section which details unconventional pedagogy in this book.
- Finish the items in the "to_do.tex" file. These are mostly very small technical fixes.
- Add the "Preview to differential forms" chapter.
- In Chapter 3, see this: "(In the first equality above, we have used Theorem 4.28, even though this theorem has not been proven yet. That theorem should be moved so that it precedes this theorem. Doing that will require checking to make sure no other things refer to that theorem as being from the later chapter)."
- Explain more explicitly how the cancellation in part 3 of the proof of 9.42 occurs. ("all the internal boundaries in the sum $\sum_{C \in D_N(\mathrm{cl}(\mathbb{H}^k))} \int_{\partial C}$ cancel, since each boundary appears twice with opposite orientations").

2

Linear algebra

Linear algebra is the study of *linear elements*- which are more commonly known as *vectors*- and of functions that “respect” the algebraic structure of linear elements.

[TO-DO]

2.1 Vectors

Definition 2.1. (Intuitive definition of vector).

Intuitively, a *vector* is a locationless directed line segment. By "locationless", we mean that two directed line segments are considered equal iff one of them can be moved- without changing the distance between its start and end point- so that it coincides with the other.

This intuitive definition is formally problematic because the definitions of *directed line segment* and *point* depend on each other; they are circular definitions.

- A *directed line segment* can be specified by listing a start and end *point*.
- A *point* is arrived at by following a path constructed of *directed line segments*.

not only that...

- defining *locationless directed line segment* requires thinking of a difference of points as being a locationless directed line segment
- defining *point* requires thinking of a sum of directed line segments

begin with element of $\mathbb{R}^n$ is locationless directed line segment

a point is an element of $\mathbb{R}^n$ a locationless directed line segment requires a notion of increment

Fortunately, we can escape this circular logic if we can give a definition of *directed line segment* or of *point* that depends on a more primitive mathematical object, as, if this is the case, then the definitions no longer both depend on each other.

To perform this "escape", we will use elements of $\mathbb{R}^n$ as the primitive mathematical objects that underlie the notion of *point*.

$\mathbb{R}^n$

Definition 2.2. ($\mathbb{R}^n$).

Let n be a positive integer. Recall that if S is any set, then S^n denotes the set of length-n tuples with entries from S:

$$S^n := \underbrace{S \times ... \times S}_{n \text{ times}} = \{(x_1, ..., x_n) \mid x_1, ..., x_n \in S\}.$$

In particular, $\mathbb{R}^n$ is the set of length-n tuples whose entries are real numbers:

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times ... \times \mathbb{R}}_{n \text{ times}} = \{(x_1, ..., x_n) \mid x_1, ..., x_n \in \mathbb{R}\}.$$

Definition 2.3. (Column notation for elements of $\mathbb{R}^n$).

To save horizontal writing space, we will use the following *column notation* to denote elements of $\mathbb{R}^n$, and write $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ rather than $(x_1, ..., x_n) \in \mathbb{R}^n$. Thus we have

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times ... \times \mathbb{R}}_{n \text{ times}} = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mid x_1, ..., x_n \in \mathbb{R} \right\}.$$

When we wish to refer to an element of $\mathbb{R}^n$ without explicitly specifying what its entries are, we use a bold letter, and say something like "let $\mathbf{x} \in \mathbb{R}^n$".

Now that we have defined $\mathbb{R}^n$ and introduced notation to represent elements of $\mathbb{R}^n$, we define an *addition* operation $+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and a *scalar multiplication* operation $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, and investigate the algebraic properties of these operations.

Definition 2.4. (Addition and scalar multiplication in $\mathbb{R}^n$).

We define *addition* $+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and *scalar multiplication* $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ operations as follows.

As a preliminary, wen note for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ we use the notation $\mathbf{x}_1 + \mathbf{x}_2 := +(\mathbf{x}_1, \mathbf{x}_2)$, and that for $c \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n$ we use the notation $c\mathbf{x} := \cdot(c, \mathbf{x})$.

We define

$$\underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} + \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{y}} := \underbrace{\begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}}_{\mathbf{x}+\mathbf{y}}$$

$$c \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} := \underbrace{\begin{pmatrix} cx_1 \\ \vdots \\ cx_n \end{pmatrix}}_{c\mathbf{x}}.$$

The above addition operation $+$ and scalar multiplication operation $\cdot$ are algebraically sensible because they obey properties one would expect of an "addition operation" and a "multiplication operation".

Theorem 2.5. (Properties of $+$ and $\cdot$ $\mathbb{R}^n$).

$+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ satisfies the following properties:

1. (Existence of additive identity).
2. (Existence of additive inverses).
3. (Associativity of $+$). For all $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^n$, $(\mathbf{x}_1 + \mathbf{x}_2) + \mathbf{x}_3 = \mathbf{x}_1 + (\mathbf{x}_2 + \mathbf{x}_3)$.
4. (Commutativity of $+$). For all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$, $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{x}_2 + \mathbf{x}_1$.

$\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ satisfies the following properties:

1. (Left-associativity of $\cdot$). For all $c_1, c_2 \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, $c_2(c_1\mathbf{x}) = c_2c_1\mathbf{x}$.
2. (Existence of multiplicative identity). For all $\mathbf{x} \in \mathbb{R}^n$, $1\mathbf{x} = \mathbf{x}$.

... The associativity and commutativity of $+$ follow from the associativity and commutativity of addition of real numbers. The left-associativity of $\cdot$ follows from the associativity of real numbers. $\square$

Vectors in $\mathbb{R}^n$

Now we define *point*, *directed line segment*, and *vector* in terms of elements of $\mathbb{R}^n$.

Definition 2.6. (Point in $\mathbb{R}^n$).

You are likely familiar with using elements of $\mathbb{R}^2$, such as $(x, y) \in \mathbb{R}^2$, to specify the two-dimensional position that has horizontal coordinate x and vertical coordinate y. We can similarly describe positions in three dimensional space by letting $(x, y, z) \in \mathbb{R}^3$ denote the position with horizontal coordinate y, vertical coordinate z, and depth coordinate x. As for elements of $\mathbb{R}^4$ and $\mathbb{R}^5$, well, humans are not really capable of visualizing these higher dimensional spaces in the same way we visualize $\mathbb{R}^3$, but we can still consider an element $(x, y, z, w) \in \mathbb{R}^4$ to be a "four dimensional position", or an element $(p_1, p_2, p_3, p_4, p_5) \in \mathbb{R}^5$ to be a "five dimensional position".

In general, an element $\mathbf{p} = (p_1, ..., p_n)$ of $\mathbb{R}^n$ can be thought of as an "n-dimensional position". For this reason, we define a *point in* $\mathbb{R}^n$ to be an element of $\mathbb{R}^n$.

Definition 2.7. (Directed line segment in $\mathbb{R}^n$).

A *directed line segment* in $\mathbb{R}^n$ is a tuple of the form $(\mathbf{p}, \mathbf{q})$, where $\mathbf{p}$ and $\mathbf{q}$ are points in $\mathbb{R}^n$. That is, a directed line segment in $\mathbb{R}^n$ is an element of $\mathbb{R}^n \times \mathbb{R}^n$, where we make sure to think of each $\mathbb{R}^n$ in the Cartesian product as being "the set of all points in $\mathbb{R}^n$".

Given a directed line segment $\boldsymbol{\ell} = (\mathbf{p}, \mathbf{q})$, we call $\mathbf{p}$ the *start point* of $\boldsymbol{\ell}$ and $\mathbf{q}$ the *end point* of $\boldsymbol{\ell}$.

Now that we have defined *directed line segment*, we only need to know of the following interpretation of vector addition in order to be able to formally define what a *locationless directed line segment* is.

Derivation 2.8. (Addition in $\mathbb{R}^n$ as point plus directed line segment).

- take a directed line segment $\boldsymbol{\ell} = (\mathbf{p}, \mathbf{q})$
- consider the difference $\mathbf{p} - \mathbf{q}$
- so far this only has algebraic meaning
- notice $(p_1 - q_1, p_2 - q_2) =$ (horizontal distance from $\mathbf{p}$ to $\mathbf{q}$, vertical distance " ") to obtain geometric meaning
- this geometric meaning involves the directed line segment that goes from $\mathbf{p}$ to $\mathbf{q}$, *not* a locationless directed line segment (so no circular logic!)

how would this go for addition?

- doesn't make sense to add points, so we'd have to add directed line segments
- would only be defined when end point of first is start point of second; $(\mathbf{p}, \mathbf{q}) + (\mathbf{q}, \mathbf{r}) := (\mathbf{p}, \mathbf{r})$

================

What is the geometric interpretation of these addition and subtraction operations? To answer this, it helps to consider the case $n = 2$ since it is hard to visualize $\mathbb{R}^n$ for arbitrary n.

Notice that the directed line segment $\boldsymbol{\ell} = \left(\begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \begin{pmatrix} p_1 + v_1 \\ p_2 + v_2 \end{pmatrix} \right)$ is the directed line segment whose start point is $\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ and whose endpoint lies in the direction of

and let $\in \mathbb{R}^2$ be an element of $\mathbb{R}^2$ (don't think of $\mathbf{v}$ as being a point!), and consider the expression $\mathbf{p} + \mathbf{v}$:

$$\underbrace{\begin{pmatrix} p_1 \\ p_2 \end{pmatrix}}_{\mathbf{p}} + \underbrace{\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}}_{\mathbf{v}} := \underbrace{\begin{pmatrix} p_1 + v_1 \\ p_2 + v_2 \end{pmatrix}}_{\mathbf{p}+\mathbf{v}}$$

Definition 2.9. (Vector in $\mathbb{R}^n$).

A *vector* in $\mathbb{R}^n$ is a locationless directed line segment in $\mathbb{R}^n$:

$(\mathbf{p}_1, \mathbf{q}_1) = (\mathbf{p}_2, \mathbf{q}_2)$ iff $\mathbf{p}_2 - \mathbf{p}_1 = \mathbf{q}_2 - \mathbf{q}_1$

Remark 2.10. There are various ways to interpret what elements of $\mathbb{R}^n$ represent. Such interpretations can often seem elaborate and perhaps confusing, so remember that this definition applies no matter the interpretation; at the end of the day, an element of $\mathbb{R}^n$ is always a length-n tuple whose entries are real numbers.

Derivation 2.11. (Geometric interpretation of addition in $\mathbb{R}^n$; vectors).

We see that $\mathbf{p} + \mathbf{v}$ is the point in $\mathbb{R}^2$ obtained by starting at $\mathbf{p}$, traveling v_1 units along the horizontal axis, and then traveling v_2 units along the vertical axis. In other words, $\mathbf{p} + \mathbf{v}$ is the end point of the directed line segment that has $\mathbf{p}$ as a start point, that has a "rise" of v_2, and that has a "run" of v_1. Notice here that the tuple $(v_1, v_2) \in \mathbb{R}^2$ *represents* the directed line segment with "rise" v_2 and "run" v_1.

We will refer to elements of $\mathbb{R}^2$ that we think of as representing directed line segments in this way as *vectors in* $\mathbb{R}^2$.

Thus, we can think of $\mathbf{p} + \mathbf{v}$ as being the point in $\mathbb{R}^2$ obtained by starting at $\mathbf{p}$ and then traveling along the directed line segment that is described by the tuple (v_1, v_2).

We have just discovered that we can

Derivation 2.12. (Vectors in $\mathbb{R}^n$).

Notice that in the above $\mathbf{w}$ represented a line segemnet

Given a point $\mathbf{p} = (p_1, p_2)$ in $\mathbb{R}^2$ and an element $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$, we can obtain the point $(p_1+v_1, p_2+v_2) \in \mathbb{R}^2$ by adding the corresponding components of $\mathbf{p}$ and $\mathbf{v}$. What

and $\mathbf{q} = (q_1, q_2)$ in $\mathbb{R}^2$, we define the *vector between* $\mathbf{p}$ *and* $\mathbf{q}$ to be the directed line segment that starts at $\mathbf{p}$ and ends at $\mathbf{q}$.

For an element $(v_1, ..., v_n) \in \mathbb{R}^n$ where $n > 3$, some vague notions of "n-dimensional position" and "traveling in the direction of a directed line segment in n-dimensional space" will have to suffice, since we can't easily visualize more than three spatial dimensions.

When an element of $\mathbb{R}^n$ is thought of as being a directed line segments, we call it a *vector in* $\mathbb{R}^n$.

The following definitions provide algebraic tools we can use to reason about points in $\mathbb{R}^n$.

Given that points in $\mathbb{R}^n$ can be added together, you might suspect that they can be subtracted as well. They can! The following two definitions explain the details.

Theorem 2.13. (The zero element of $\mathbb{R}^n$).

There is an element $\mathbf{0}$ of $\mathbb{R}^n$ that satisfies $\mathbf{v}+\mathbf{0}=\mathbf{v}=\mathbf{0}+\mathbf{v}$ for all $\mathbf{v}\in\mathbb{R}^n$. We call $\mathbf{0}$ the *zero element* of $\mathbb{R}^n$. As you may have guessed, $\mathbf{0}$ is the element of $\mathbb{R}^n$ that has a zero in every entry:

$$\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Referring to the definition of addition in $\mathbb{R}^n$, we can check that this definition of $\mathbf{0}$ does indeed satisfy $\mathbf{v}+\mathbf{0}=\mathbf{v}=\mathbf{0}+\mathbf{v}$ for all $\mathbf{v}\in\mathbb{R}^n$.

When interpreted to be a point, the zero element is often called “the origin”.

Theorem 2.14. (Additive inverses of points in $\mathbb{R}^n$).

For every point $\mathbf{v}\in\mathbb{R}^n$, there is a point $-\mathbf{v}$ in $\mathbb{R}^n$ such that $\mathbf{v}+(-\mathbf{v})=\mathbf{0}$. Specifically, we have

$$-\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} -v_1 \\ \vdots \\ -v_n \end{pmatrix}. \tag{2.1}$$

We call $-\mathbf{v}$ the *additive inverse* of $\mathbf{v}$. Of course, we colloquially refer to $-\mathbf{v}$ as “minus $\mathbf{v}$”.

As mentioned, there is indeed a notion of *subtraction in* $\mathbb{R}^n$: we use the notation $\mathbf{v}-\mathbf{w}$ to mean $\mathbf{v}+(-\mathbf{w})$. If $\mathbf{w}=\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$, one can interpret the point $\mathbf{v}-\mathbf{w}$ to be the point that is arrived at by starting at $\mathbf{v}$, traveling $-w_1$ units on the horizontal axis, and then traveling $-w_2$ units on the vertical axis.

Having introduced the notions of vector addition, additive inverse, and vector subtraction, we now present two properties that are characteristic of vector addition.

Vectors in $\mathbb{R}^n$

merge these interpretations into the above definitions; it’s best to talk about points in $\mathbb{R}^n$ and vectors in $\mathbb{R}^n$ at the same time rather than one section after another

Derivation 2.15. (Interpretation of vector addition).

Previously, we defined an addition operation on $\mathbb{R}^n$, and gave intuition for what $\mathbf{v}+\mathbf{w}$ meant when $\mathbf{v}\in\mathbb{R}^n$ and $\mathbf{w}\in\mathbb{R}^n$ were considered to be points. There is a related but slightly different interpretation of the addition operation when $\mathbf{v}$ and $\mathbf{w}$ are considered to be vectors. We can imagine $\mathbf{v}+\mathbf{w}$ as being the result of starting at the start point of $\mathbf{v}$, then traveling according to the vector $\mathbf{v}$ to reach the end point of $\mathbf{v}$, then applying “locationlessness” to place the start point of $\mathbf{w}$ on the end point of $\mathbf{v}$, and then traveling to the end point of $\mathbf{w}$. The directed line segment that reaches from “overall” beginning to the “overall” end is $\mathbf{v}+\mathbf{w}$. Thus, if we first imagine connecting the end point of $\mathbf{v}$ to the start point of $\mathbf{w}$, then $\mathbf{v}+\mathbf{w}$ is represented by the directed line segment that connects the start point of $\mathbf{v}$ to the end point of $\mathbf{w}$. This interpretation is often called the “tail-to-tip method”, since one computes $\mathbf{v}+\mathbf{w}$ by forming the vector by drawing a straight line from the “tail” of $\mathbf{v}$ to the “tip” of $\mathbf{w}$.

Derivation 2.16. (Interpretation of the zero vector).

Noticing that for any $\mathbf{v}\in\mathbb{R}^n$ we have $\mathbf{v}-\mathbf{v}=\mathbf{0}$, we can conclude that the zero vector can be interpreted to be the vector whose start and end points coincide. Thus, the zero vector is not a locationless line segment but a locationless point!

Derivation 2.17. (Interpretation of a vector’s additive inverse, interpretation of vector subtraction).

To obtain an interpretation of a vector’s additive inverse in the vector context, first recall that for any $\mathbf{v}\in\mathbb{R}^n$ the additive inverse $-\mathbf{v}$ satisfies $\mathbf{v}+(-\mathbf{v})=\mathbf{0}$. Since the start and end points of $\mathbf{0}$ coeincide, we now ask: if we are to follow the above interpretation of vector addition, how can we make the start point of the sum $\mathbf{v}+(-\mathbf{v})$ coincide with the end point of the sum, i.e., to make the start point of $\mathbf{v}$ coincide with the end point of $-\mathbf{v}$? Recall from the above tail-to-tip method for vector addition that we are indeed free to vary the start point of $\mathbf{v}$ and the end point of $-\mathbf{v}$, so we can impose that the start point of $\mathbf{v}$ is the end point of $-\mathbf{v}$. So, we now know what the end point of $-\mathbf{v}$ is. What about the start point of $-\mathbf{v}$? Recall additionally that the tail-to-tip method imposes on us that the end point of the first vector in the sum, $\mathbf{v}$, must be the start point of the second vector in the sum, $-\mathbf{v}$. Thus, the end point of $-\mathbf{v}$ is the start point of $\mathbf{v}$. In all, we have shown that the start point of $-\mathbf{v}$ is the end point of $\mathbf{v}$ and that the end point of $-\mathbf{v}$ is the start point of $\mathbf{v}$, so we can interpret $-\mathbf{v}$ to be the line segment that has the same length as $\mathbf{v}$ but that goes in the opposite direction.

To interpret $\mathbf{v} - \mathbf{w}$ when $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are vectors, just consider $\mathbf{v} - \mathbf{w}$ to be $\mathbf{v} + (-\mathbf{w})$, and make use of the interpretations of vector addition and a vector's additive inverse.

The following definition presents one operation on vectors that we haven't introduced yet.

Definition 2.18. (Scaling of vectors).

If $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ is a vector, we can "grow" and "shrink" $\mathbf{v}$ by scaling its entries. Formally, given a real number $c \in \mathbb{R}$, we define the following *vector scaling* operation $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$:

$$c \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} := \begin{pmatrix} cv_1 \\ \vdots \\ cv_n \end{pmatrix}.$$

In practice, for $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$, we write "$c\mathbf{v}$" to mean $c \cdot \mathbf{v}$".

Note, there is not really a clear geometric interpretation of this operation when $\mathbf{v} \in \mathbb{R}^n$ is considered to be a point rather than a vector.

Remark 2.19. (Scalars).

Since we can scale vectors in $\mathbb{R}^n$ by elements of $\mathbb{R}$, we often refer to elements of $\mathbb{R}$ as *scalars*.

Remark 2.20. (Vector terminology).

Since the set of points in $\mathbb{R}^n$ is equal to the set of vectors in $\mathbb{R}^n$, we will from this point forward prefer the word "vector" instead of the word "point", while keeping in mind that we can think of a vector as being a directed line segment or a location.

Properties of $+$ and $\cdot$ in $\mathbb{R}^n$

We have proved several properties of the vector addition operation $+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ that we restate in the below (1.1), (1.2), (1.3), and (1.4). In the below (2), we state a property of the scalar multiplication operation $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, and in the below (3), we state properties involving both $+$ and $\cdot$.

Theorem 2.21. (Properties of $+$ and $\cdot$ in $\mathbb{R}^n$).

1. (Properties of $+$).
   - 1.1. (Existence of additive identity). There exists $\mathbf{0} \in V$ such that for all $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
   - 1.2. (Closure under additive inverses). For all $\mathbf{v} \in \mathbb{R}^n$ there exists $-\mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
   - 1.3. (Associativity of $+$). For all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbb{R}^n$, $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$.
   - 1.4. (Commutativity of $+$). For all $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$.
2. (Normalization for $\cdot$). $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$.
3. (Compatibility of $+$ and $\cdot$).
   - 3.1. For all $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ and $c \in \mathbb{R}$, $c(\mathbf{v}_1 + \mathbf{v}_2) = c\mathbf{v}_1 + c\mathbf{v}_2$.
   - 3.2. For all $\mathbf{v} \in \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$, $(c_1 + c_2)\mathbf{v} = c_1\mathbf{v} + c_2\mathbf{v}$.
   - 3.3. For all $\mathbf{v} \in \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$, $c_2(c_1\mathbf{v}) = c_2c_1\mathbf{v}$.

Proof. To check (2) and (3), refer to the definitions of vector addition $+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and scalar multiplication $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$. $\square$

Vector-containing sets

[motivation: recap how geometric interpretations of points and vectors in $\mathbb{R}^n$ is underpinned by algebraic characterization]

At this point, we use the idea that a set like $\mathbb{R}^n$ can be spanned by another set to define an abstract notion of "vector-containing set".

Definition 2.22. (Vector-containing set over $\mathbb{R}$).

Suppose that S is a set for which there exist functions $+ : S \times S \to S$ and $\cdot : \mathbb{R} \times S \to S$. We say that S is a *vector-containing set over* $\mathbb{R}$ iff the operations $+$ and $\cdot$ satisfy the typical[1] properties one would expect of "vector addition" and "vector scaling". That is, S is a vector-containing set over $\mathbb{R}$ iff $+$ and $\cdot$ satisfy the properties listed in Theorem 2.21:

1. (Properties of $+$).
   - 1.1. (Existence of additive identity). There exists $\mathbf{0} \in S$ such that for all $\mathbf{v} \in S$, $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
   - 1.2. (Closure under additive inverses). For all $\mathbf{v} \in S$ there exists $-\mathbf{v} \in S$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
   - 1.3. (Associativity of $+$). For all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in S$, $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$.
   - 1.4. (Commutativity of $+$). For all $\mathbf{v}_1, \mathbf{v}_2 \in S$, $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$.
2. (Normalization for $\cdot$). $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in S$.
3. (Properties of $+$ and $\cdot$).
   - 3.1. (Properties of $+$).
     - 3.1.1. (Existence of additive identity). There exists $\mathbf{0} \in S$ such that for all $\mathbf{v} \in S$, $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
     - 3.1.2. (Closure under additive inverses). For all $\mathbf{v} \in S$ there exists $-\mathbf{v} \in S$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
     - 3.1.3. (Associativity of $+$). For all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in S$, $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$.
     - 3.1.4. (Commutativity of $+$). For all $\mathbf{v}_1, \mathbf{v}_2 \in S$, $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$.
   - 3.2. (Compatibility of $+$ and $\cdot$).
     - 3.2.1. For all $\mathbf{v}_1, \mathbf{v}_2 \in S$ and $c \in \mathbb{R}$, $c(\mathbf{v}_1 + \mathbf{v}_2) = c\mathbf{v}_1 + c\mathbf{v}_2$.
     - 3.2.2. For all $\mathbf{v} \in S$ and $c_1, c_2 \in \mathbb{R}$, $(c_1 + c_2)\mathbf{v} = c_1\mathbf{v} + c_2\mathbf{v}$.
     - 3.2.3. For all $\mathbf{v} \in S$ and $c_1, c_2 \in \mathbb{R}$, $c_2(c_1\mathbf{v}) = c_2c_1\mathbf{v}$.

When S is a vector-containing set over $\mathbb{R}$, we refer to elements of S as *vectors*. As before, elements of $\mathbb{R}$ are sometimes called *scalars*.

This abstract characterization of a vector-containing set is useful because many sets that at first glance don't seem to contain vectors actually do, in a sense.

Remark 2.23. (Examples of vector-containing sets over $\mathbb{R}$).

Here are some examples of vector-containing sets over $\mathbb{R}$:

- $\mathbb{R}^n$, for any positive integer n. (This one isn't intended to be surprising).
- The set of polynomials with real coefficients of degree less than n, where $+$ is function addition and $\cdot$ is the scaling of a function by a real number.
- The set of infinitely differentiable functions $\mathbb{R} \to \mathbb{R}$, where $+$ is function addition and $\cdot$ is the scaling of a function by a real number.
- The set of infinite sequences of real numbers, where $+$ and $\cdot$ are defined in the ways you would expect.

Now that we have defined what a "vector-containing set over $\mathbb{R}$" is, it is reasonable to introduce some more abstraction. In place of "vector-containing sets over $\mathbb{R}$", we will now consider "vector-containing sets over K", where K is the set-with-structure that contains the scalars. In other words, in the same way that a vector is "something that lives in a vector-containing set", we will define a scalar to be "something that lives in a 'scalar space' ". Unfortunately, the terminology "scalar space" is nonstandard; "scalar spaces" are actually called *fields*.

[1] When you try to remember these properties, there is no need to be incredibly specific. You don't need to list out "existence of additive identity", "closure under additive inverses", and so on *every* time you remind yourself of what a vector-containing set over $\mathbb{R}$ is!

Definition 2.24. (Field).

Suppose that K is a set for which there exist functions $+ : K \times K \to K$ and $\cdot : K \times K \to K$. We say that the tuple $(K, +, \cdot)$ is a *field* iff it satisfies[2] the following:

1. K is a "commutative group under addition". This means that conditions 1.1 through 1.5 must hold.

   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 $c \in K$, $0 + c = c = c + 0$.

   1.3. (Associativity of $+$). For all $c_1, c_2, c_3 \in K$, $(c_1 + c_2) + c_3 = c_1 + (c_2 + c_3)$.

   1.4. (Closure under additive inverses). For all $c \in K$ there exists $-c \in K$ such that $(-c) + c = 0 = c + (-c)$.

   1.5. (Commutativity of $+$). For all $c_1, c_2 \in K$, $c_1 + c_2 = c_2 + c_1$.

2. K is a "commutative group under multiplication". This means that conditions 2.1 through 2.5 must hold.

   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 $c \in K$, $1 \cdot c = c = c \cdot 1$.

   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 $k \in K$ with $k \neq 0$, there exists $\frac{1}{c} \in K$ such that $\frac{1}{c} \cdot c = 1 = c \cdot \frac{1}{c}$.

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

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

Colloquially, we often say "let K be a field" instead of "let $(K, +, \cdot)$ be a field".

Recall that the whole point of defining a field is to formalize the notion of what a "scalar" is. For this reason, elements of a field are called *scalars*.

Remark 2.25. (Examples of fields).

- $\mathbb{R}$ is a field.
- The complex numbers $\mathbb{C} = \{a + b\sqrt{-1} \mid a, b \in \mathbb{R}\}$ are a field.

Remark 2.26. (Don't memorize the definition of a field!).

It's not necessary to memorize all the conditions for a field. Just remember that a field is "a set that contains elements which one can add, subtract, multiply, and divide". In this book, you can imagine an arbitrary field K as being $\mathbb{R}$ almost all of the time[3].

Without further ado, we define "vector-containing set over a field".

Definition 2.27. (Vector-containing set over a field).

Suppose that K is a field, that S is a set, and that there exist functions $+ : S \times S \to S$ and $\cdot : K \times S \to S$. We say that the tuple $(S, K, +, \cdot)$ is a *vector-containing set*, or, more colloquially, that "S is a vector-containing set over K", iff:

1. (Properties of $+$).

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

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

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

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

2. (Normalization for $\cdot$). $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in S$.

3. (Properties of $+$ and $\cdot$).

   3.1. (Properties of $+$).

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

---

[2] In the terminology of abstract algebra, a field can be defined to be (1) an "integral domain that is closed under multiplicative inverses" or (2) as a "commutative division ring".

[3] It's true that there is more to it when the field is a *finite field*, but this is not a major concern in this book.

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

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

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

3.2. (Compatibility of $+$ and $\cdot$).

3.2.1. For all $\mathbf{v}_1, \mathbf{v}_2 \in S$ and $c \in K$, $c(\mathbf{v}_1 + \mathbf{v}_2) = c\mathbf{v}_1 + c\mathbf{v}_2$.

3.2.2. For all $\mathbf{v} \in S$ and $c_1, c_2 \in K$, $(c_1 + c_2)\mathbf{v} = c_1\mathbf{v} + c_2\mathbf{v}$.

3.2.3. For all $\mathbf{v} \in S$ and $c_1, c_2 \in K$, $c_2(c_1\mathbf{v}) = c_2c_1\mathbf{v}$.

Elements of vector-containing sets are called “vectors”.

In practice, we often don’t explicitly mention a field, and say “let S be a vector-containing sets” instead of “let S be a vector-containing set over a field K”.

Vector spaces

[Motivation for thinking about linear combinations...]

Definition 2.28. (Linear combination).

Let $S = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be a finite vector-containing set over a field K. We define a *linear combination of the vectors in* S, or, more colloquially, a *linear combination of* $\mathbf{v}_1, ..., \mathbf{v}_k$, to be a vector of the form

$$c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k,$$

where $c_1, ..., c_k$ are some scalars in K. So, a linear combination of $\mathbf{v}_1, ..., \mathbf{v}_k$ is a "weighted sum" involving $\mathbf{v}_1, ..., \mathbf{v}_k$.

How do we define "linear combination" when S is an arbitrary set that could be empty or contain infinitely many vectors? It is easy to extend the idea of the above definition to this more general setting. For an arbitrary vector-containing set S over a field K, we define a *linear combination of the vectors in* S to be a sum of the form

$$\sum_{c \in C, \mathbf{v} \in S} c\mathbf{v},$$

where C is some subset of K.

Definition 2.29. (Span).

Let $S = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be a finite vector-containing set over a field K. We define the *span* of S, or, more colloquially, the *span of* $\mathbf{v}_1, ..., \mathbf{v}_k$, to be the set of all linear combinations of $\mathbf{v}_1, ..., \mathbf{v}_k$:

$$\text{span}(S) = \text{span}(\{\mathbf{v}_1, ..., \mathbf{v}_k\}) := \Big\{c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k \mid c_1, ..., c_k \in K\Big\}.$$

For geometric intuition, notice that if $\mathbf{v}_1, ..., \mathbf{v}_k \in \mathbb{R}^n$, then $\text{span}(\{\mathbf{v}_1, ..., \mathbf{v}_k\})$ is the k-dimensional plane spanned by $\mathbf{v}_1, ..., \mathbf{v}_k$ that is embedded in $\mathbb{R}^n$. For example, if $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, then $\text{span}(\{\mathbf{v}, \mathbf{w}\}) = \{c\mathbf{v} + d\mathbf{w} \mid c, d \in \mathbb{R}\}$ is the 2-dimensional plane spanned by $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$.

In the more general case when S could be empty or infinite, we define the span of S to be the set of all convergent linear combinations of the vectors in S:

$$\text{span}(S) := \Big\{\mathbf{w} = \sum_{c \in C, \mathbf{v} \in S} c\mathbf{v} \mid C \subseteq K \text{ and } \mathbf{w} \text{ converges}\Big\}.$$

[More motivation...]. Without further ado, we define "vector space over a field".

Definition 2.30. (Vector space over a field).

Suppose that K is a field, that V is a set, and that there exist functions $+ : V \times V \to V$ and $\cdot : K \times V \to V$. We say that the tuple $(V, K, +, \cdot)$ is a *vector space*, or, more colloquially, that "V is a vector space over K", iff

1. $(V, K, +, \cdot)$ is a vector-containing set.
2. V is spanned by some vector-containing set, i.e., there is a vector-containing set S for which $V = \text{span}(S)$.

Elements of vector spaces are called "vectors".

In practice, we often don't explicitly mention a field, and say "let V be a vector space" instead of "let V be a vector space over a field K".

Theorem 2.31. ("Linear spaces").

[should this just be inside the definition? we can make this a remark about the span condition]

Every subset of $\mathbb{R}^n$ that is a vector space is a plane.

European mathematicians tend to call vector spaces *linear spaces*

linear elements

Remark 2.32. Vector-containing sets and subsets of vector spaces are one and the same: vector-containing sets are subsets of vector spaces, and subsets of vector spaces are vector-containing sets. From this point on, we will favor "subset of a vector space" over "vector-containing set", because the terminology "vector-containing set" is nonstandard.

All of the examples of vector-containing sets over $\mathbb{R}$ we gave before are actually vector spaces.

Remark 2.33. (Examples of vector spaces).
Here are some examples of vector spaces:

- $\mathbb{R}^n$, for any positive integer n. (This one isn't intended to be surprising).
- K^n, for any positive integer n.
- The set of polynomials with real coefficients of degree less than n, where $+$ is function addition and $\cdot$ is the scaling of a function by a real number.
- The set of polynomials with rational coefficients of degree less than n, where $+$ is function addition and $\cdot$ is the scaling of a function by a real number.
- The set of infinitely differentiable functions $\mathbb{R} \to \mathbb{R}$, where $+$ is function addition and $\cdot$ is the scaling of a function by a real number.
- The set of infinite sequences of real numbers, where $+$ and $\cdot$ are defined in the ways you would expect.

Here are some examples of vector-containing sets that are *not* vector spaces:

- $\left\{\mathbf{0}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\} \subseteq \mathbb{R}^2$
- $\{z \geq ax + by \mid a, b \text{ not both zero}\} \subseteq \mathbb{R}^3$
- $\{x_n \geq c_1 x_1 + ... + c_{n-1} x_{n-1} \mid c_1, ..., c_{n-1} \text{ not all zero}\} \subseteq \mathbb{R}^n$

Remark 2.34. (Why is the generality of "vector space" useful?)
In this book, we will often consider arbitrary vector space and use it as a "base" from which to construct more vector spaces, while secretly imagining the "base" vector space as being $\mathbb{R}^n$. Dealing in the abstraction of vector spaces makes it easier to see precisely how vector spaces produced from $\mathbb{R}^n$ are related to $\mathbb{R}^n$ by removing any doubt as to whether coordinates, which are easy to run into when one deals with $\mathbb{R}^n$ explicitly, are fundamentally at play.

Now that we have defined what a vector space is in full generality, we present an alternative characterization of vector spaces that is easier in practice to check than the above definition.

Derivation 2.35. (Vector spaces are closed under addition and scalar multiplication).
Suppose V is a vector space over a field K. Working heuristically, we can discover that the spanning condition satisfied by vector spaces is equivalent to another condition:

$$\begin{gathered}
\text{There is a vector-containing set } S \text{ for which } V = \text{span}(S). \\
\iff \\
\text{For all } \mathbf{v}_1, \mathbf{v}_2 \in V \text{ we have } \mathbf{v}_1 + \mathbf{v}_2 \in V \text{ (“}V\text{ is closed under vector addition”)} \\
\text{and} \\
\text{for all } c \in K \text{ and } \mathbf{v} \in V \text{ we have } c\mathbf{v} \in V \text{ (“}V\text{ is closed under vector scaling”).}
\end{gathered}$$

The idea for showing that the hypothesis implies the first part of the conclusion (the part before the "and") is this: if we assume that V is spanned by some vector-containing set S, then we can compute $\mathbf{v}_1 + \mathbf{v}_2$ by writing $\mathbf{v}_1$ and $\mathbf{v}_2$ as "weighted sums" of elements from S, and the result, being a (longer) weighted sum of elements from S, is thus in V. Similar reasoning shows the second part of the conclusion. The reverse implication is simple: there is always a vector-containing set spanning V, since every vector-containing set spans itself.

When determining whether a set V is a vector space or not, it is almost always easier to check the above closure conditions than to check if there is a set that spans V. Thus, the following characterization of vector spaces is almost always what one should use when testing if a set is a vector space.

Theorem 2.36. (Alternate characterization of vector spaces).
Suppose that K is a field, that V is a set, and that there exist functions $+ : V \times V \to V$ and $\cdot : K \times V \to V$. The tuple $(V, K, +, \cdot)$ is a vector space iff

1. $(V, K, +, \cdot)$ is a vector-containing set.
2. "V is closed under vector addition and vector scaling".
   - 2.1. (Closure under $+$). For all $\mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{v}_1 + \mathbf{v}_2 \in V$.
   - 2.2. (Closure under $\cdot$). For all $c \in \mathbb{R}$ and $\mathbf{v} \in V$, $c\mathbf{v} \in V$.

Miscellaneous facts about vector spaces

Theorem 2.37. ("Obvious" facts about the additive identity and additive inverses).
If V is a vector space, we have $0\mathbf{v} = \mathbf{0}$ and $-\mathbf{v} = (-1)\mathbf{v}$ for all $\mathbf{v} \in V$.

Proof. Both facts follow from the fact that if V is a vector space then $(c_1 + c_2)\mathbf{v} = c_1\mathbf{v} + c_2\mathbf{v}$ for all $\mathbf{v} \in V$ and $c_1, c_2 \in K$.

For the first fact, we note that $0\mathbf{v} + 0\mathbf{v} = (0 + 0)\mathbf{v} = 0\mathbf{v}$ and subtract $0\mathbf{v}$ from both sides to obtain $0\mathbf{v} = \mathbf{0}$. To prove the second fact, we must verify that $\mathbf{v} + (-1)\mathbf{v} = \mathbf{0}$. This is the case because $\mathbf{v} + (-1)\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = (1 - 1)\mathbf{v} = 0\mathbf{v}$; we have $0\mathbf{v} = \mathbf{0}$ via the first fact. $\square$

Remark 2.38. ($\emptyset$ is not a vector space).
The empty set $\emptyset$ is not a vector space over any field because it contains no additive identity.

Vector subspaces

Definition 2.39. (Vector subspace).
If V and W are vector spaces over K and $W \subseteq V$, then W is a *vector subspace* of V.

Remark 2.40. (Examples of vector subspaces).

- $\mathbb{R}^2$ is a vector subspace of $\mathbb{R}^3$.
- $\mathbb{R}^n$ is a vector subspace of $\mathbb{R}^m$ whenever $n < m$.
- Any line in the plane is a vector subspace of $\mathbb{R}^2$.
- A line in three-dimensional space is not a vector subspace of $\mathbb{R}^2$ but is a vector subspace of $\mathbb{R}^3$.
- Two-dimensional planes of the form $\text{span}(\mathbf{v}, \mathbf{w})$, for nonzero $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, are not vector subspaces of $\mathbb{R}^2$ but are vector subspaces of $\mathbb{R}^3$.
- Two-dimensional planes of the form $\text{span}(\mathbf{v}, \mathbf{w})$, for nonzero $\mathbf{v}, \mathbf{w} \in \mathbb{R}^4$, are not vector subspaces of $\mathbb{R}^3$ but are vector subspaces of $\mathbb{R}^4$.
- "k-dimensional planes" of the form $\text{span}(\{\mathbf{v}_1, ..., \mathbf{v}_k\})$, for nonzero $\mathbf{v}_1, ..., \mathbf{v}_k \in \mathbb{R}^n$, are not vector subspaces of $\mathbb{R}^k$ for $k < n$, but are vector subspaces of $\mathbb{R}^n$.

Linear independence

In this subsection, we introduce the concept of *linear independence*, which will serve as formal footing that can be used to justify future geometric intuitions.

Definition 2.41. (Linear independence).

Let V be a vector space. If $S = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is a finite subset of V, then we say S is *linearly independent* iff for all i we have $\mathbf{v}_i \notin \text{span}(\{\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \mathbf{v}_k\})$.

Intuitively, S is linearly independent if every vector inside it is required to produce span(S). Indeed, we prove in Lemma 2.45 that removing a vector from a linearly independent set produces a set whose span is a proper subset of what it was before.

To generalize the notion of linear independence to arbitrary subsets of V, we say that a subset $S \subseteq V$ is linearly independent iff for all $\mathbf{v} \in S$ we have $\mathbf{v} \notin \text{span}(S - \{\mathbf{v}\})$.

A subset of vector space is called *linearly dependent* if it is not linearly independent.

Theorem 2.42. Any subset of a vector space that contains $\mathbf{0}$ is linearly dependent.

Proof. If a subset S of a vector space contains $\mathbf{0}$, then the condition $(\forall \mathbf{v}\ \mathbf{v} \notin \text{span}(S - \{\mathbf{v}\}))$ is violated by $\mathbf{v} = \mathbf{0}$, since $\mathbf{0}$ is in the span of every set. □

Theorem 2.43. A single-element subset $\{\mathbf{v}\}$ of a vector space is linearly independent iff $\mathbf{v} \neq \mathbf{0}$.

Proof. $\{\mathbf{v}\}$ is linearly independent iff $\mathbf{v} \notin \text{span}(\{\mathbf{v}\} - \{\mathbf{v}\}) = \text{span}(\emptyset) = \{\mathbf{0}\}$, i.e., iff $\mathbf{v} \neq \mathbf{0}$. □

Theorem 2.44. (Condition for linear independence).

Let V be a vector space over a field K. The set $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is linearly dependent iff there are scalars $c_1, ..., c_k \in K$ not all 0 such that

$$c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k = \mathbf{0}.$$

More generally, a subset $S \subseteq V$ is linearly dependent iff there is a subset $C \subseteq K$ with $C \neq \{0\}$ for which

$$\sum_{c \in C, \mathbf{v} \in S} c\mathbf{v} = \mathbf{0}.$$

Proof. **TO-DO: account for $S = \emptyset$**

We show the more general statement (S is linearly dependent) $\iff$ $(\exists C \neq \{0\}\ \sum_{c \in C, \mathbf{v} \in S} c\mathbf{v} = \mathbf{0})$.

($\Longrightarrow$). Assume S is linearly dependent. Then S is not linearly independent, so there is some $\mathbf{v} \in S$ for which $\mathbf{v} \in \text{span}(S - \{\mathbf{v}\})$. This makes $\mathbf{v}$ a linear combination of vectors from $S - \{\mathbf{v}\}$, $\mathbf{v} = \sum_{c \in C, \mathbf{w} \in (S - \{\mathbf{v}\})} c\mathbf{w}$, where $C \subseteq K$. Subtracting $\mathbf{v}$ from both sides of the previous equation, we have $-\mathbf{v} + \sum_{c \in C, \mathbf{w} \in (S - \{\mathbf{v}\})} c\mathbf{w} = \mathbf{0}$. That is, we have $\sum_{d \in D, \mathbf{w} \in S} d\mathbf{w} = \mathbf{0}$, where $D \subseteq K$ is such that the coefficients on elements of $S - \{\mathbf{v}\}$ are from C and where the coefficient on $\mathbf{v}$ is -1. Since $-1 \in D$, we have shown, as desired, that there is a subset $D \neq \{0\}$ of K such that $\sum_{c \in D, \mathbf{w} \in S} d\mathbf{w} = \mathbf{0}$.

($\Longleftarrow$). Assume there is a subset $C \subseteq K$ with $C \neq \{0\}$ for which $\sum_{c \in C, \mathbf{v} \in S} c\mathbf{v} = \mathbf{0}$. Since $C \neq \{0\}$, then some $d \in C$ is nonzero, and we can therefore divide by d. Supposing that d is the coefficient of $\mathbf{u} \in S$, we subtract the part of the sum not involving $d\mathbf{u}$ from both sides of the previous equation and then divide by d to obtain $\mathbf{u} = -\frac{1}{d}\sum_{c \in C - \{d\}, \mathbf{v} \in S - \{\mathbf{u}\}} c\mathbf{v} = \sum_{c \in C - \{d\}, \mathbf{v} \in S - \{\mathbf{u}\}} -\frac{c\mathbf{v}}{d}$. Thus $\mathbf{u} \in \text{span}(S - \{\mathbf{u}\})$; $\mathbf{u}$ violates the linear independence condition, so S is linearly dependent. □

Lemma 2.45. (Adding and removing vectors from linearly independent and dependent sets).

1. Appending a vector not in the span of a linearly independent set to that linearly independent set produces another linearly independent set.
2. Any subset of a linearly independent set is also linearly independent.
3. Appending a vector to a linearly dependent set produces another linearly dependent set.
4. Removing a vector from a linearly dependent set does not change the span of that set.

Proof. Left as exercise. □

Basis and dimension

This subsection builds on the concept of linear independence to define the concepts of *basis* and *dimension*. A *basis* for a vector space will be defined to be a spanning set of that vector space with as few vectors in it as possible; a basis is a "minimal spanning set". The *dimension* of a *finite-dimensional* vector space will be defined to be the number of vectors in any basis for it.

Definition 2.46. (Finite- and infinite- dimensionality).
A vector space is said to be *finite-dimensional* iff it is spanned by a finite set of vectors and is said to be *infinite-dimensional* iff this is not the case.

Definition 2.47. (Basis for a finite-dimensional vector space).
Let V be a finite-dimensional vector space. We say that a subset $E \subseteq V$ is a *basis* for V iff

1. E spans V.
2. The number of vectors in E is minimal. (I.e. if F is another set of vectors spanning V, then F contains at least as many vectors as E, $|F| \geq |E|$).

Definition 2.48. (Dimension of a finite-dimensional vector space).
Let V be a finite-dimensional vector space. The *dimension* $\dim(V)$ of V is the number of basis vectors in a basis for V.

Remark 2.49. (0-dimensional vector spaces).
The empty set $\emptyset$ is a basis for $\{\mathbf{0}\}$, and so $\{\mathbf{0}\}$ is a zero-dimensional vector space. Since $\emptyset$ is the only set with cardinality zero, it is the only basis with cardinality zero; thus, $\{\mathbf{0}\}$ is the only 0-dimensional vector space.

The following lemma enables us to obtain a more useful characterization of bases.

Lemma 2.50. (Linearly independent spanning set lemma).
For any n-dimensional vector space V with $n \neq 0$, we have the following equivalent facts:

1. A set of more than n vectors is linearly dependent.
2. A set of less than n vectors does not span V.

Proof. First we show the logical equivalence.
((1) $\implies$ (2)). Suppose for contradiction that a set F of vectors with $|F| < n$ spans V. Remove vectors from F until a linearly independent set G is obtained. Since removing vectors from linearly dependent sets does not change their span, we know G spans V. Thus, G is a linearly independent spanning set with $|G| < n$. Since $n > |G|$, it follows from (1) that any set of n vectors is linearly dependent; contradiction.
((2) $\implies$ (1)). Suppose for contradiction that a set F of vectors with $|F| > n$ is linearly independent. Add vectors to F until a spanning set G of V is obtained. Since adding vectors to linearly independent sets that are not in the span of those sets preserves linear independence, we know G is linearly independent. Thus, G is a linearly independent spanning set with $|G| > n$. Since $n < |G|$, it follows from (2), that any set of n vectors does not span V; contradiction.
Now, we prove that (1) is true.
TO-DO. See p. 30 of *Linear and Geometric Algebra* by Alan MacDonald. □

Theorem 2.51. (Bases for finite-dimensional vector spaces are linearly independent spanning sets).
Let V be a finite-dimensional vector space. A set E of vectors is a basis for V iff

1. E spans V.
2. E is linearly independent.

Proof.
We show (E is a minimal spanning set of V) $\iff$ (E is linearly independent).
(Case $E \neq \emptyset$).
($\implies$). Suppose that E is a minimal spanning set of V. We need to show that E is linearly independent, so suppose for contradiction that E is linearly dependent. If E is linearly dependent, we can remove a vector from E without changing $\text{span}(E) = V$, so, E is not minimal; contradiction.
($\impliedby$). Suppose that E spans V and that E is linearly independent. We need to show that $|E|$ is a minimal spanning set of V, i.e., that a set of less than $|E|$ vectors does not span V. This is precisely what fact (2) of Lemma 2.50 tells us.

(Case $E = \emptyset$). Since there is only one E satisfying $E = \emptyset$, the if-and-only-if we need to show simplifies to an *and* statement: ($\emptyset$ is a minimal spanning set of V) and ($\emptyset$ is linearly independent). The first statement here is clearly true: $\emptyset$ is a minimal spanning set because there aren't any sets that contain fewer than zero vectors. The second statement, which is equivalent to $(\forall \mathbf{v} \in \emptyset \ \mathbf{v} \notin \text{span}(\emptyset - \{\mathbf{v}\}))$, is true by false hypothesis, since[4] there is no $\mathbf{v}$ for which "$\mathbf{v} \in \emptyset$" is a true statement.

□

Because the minimality condition in the definition we gave for "basis" boils down to a comparison of cardinalities of spanning sets, it is only applicable to finite-dimensional vector spaces. The above theorem suggests a more general definition that applies to all vector spaces, however.

Definition 2.52. (Basis).

Let V be a vector space. We say that a set E of vectors is a *basis* for V iff

1. E spans V.
2. E is linearly independent.

Theorem 2.53. Every finite-dimensional vector space has a basis.

Proof. Let V be a finite-dimensional vector space, and take a finite set $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ of vectors that span V. Remove vectors from E one by one. At some point, we must obtain a linearly independent set[5]. That is, there must be some i for which $E_i := E - \{\mathbf{e}_1, ..., \mathbf{e}_i\}$ is linearly independent and for which all E_j, $j < i$, are linearly dependent. In particular, E_{i-1} is a linearly dependent set that spans V. Applying the fact that removing a vector from a linearly dependent set does not change the span of that set, we can conclude that since E_i is obtained by removing a vector from E_{i-1}, we have $\text{span}(E_i) = \text{span}(E_{i-1}) = V$. Thus, E_i is a linearly independent set that spans V; it is a basis for V. □

Definition 2.54. (Standard basis for K^n).

Let K be a field, and consider K^n as a vector space over K. We define the *standard basis* of K^n to be the basis $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$, where the jth entry of $\hat{\mathfrak{e}}_i$ is 1 when $j = i$ and 0 otherwise.

Proof. We should check that $\hat{\mathfrak{E}}$ is indeed a basis. Checking that $\hat{\mathfrak{E}}$ spans K^n is easy and is left as an exercise.

For linear independence, consider the equation $c_1\hat{\mathfrak{e}}_1 + ... + c_n\hat{\mathfrak{e}}_n = \mathbf{0}$, where $c_1, ..., c_n \in K$. The equation can be rewritten as $\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \mathbf{0}$, which is true only when $c_i = 0$ for all i. In all we have that $c_1\hat{\mathfrak{e}}_1 + ... + c_n\hat{\mathfrak{e}}_n = \mathbf{0}$ only when $c_1, ..., c_n = 0$, so $\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n$ are linearly independent. □

Remark 2.55. (Why not define dimensionality in terms of bases?).

It is tempting to define a finite-dimensional vector spaces to be those that have finite bases. 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. Thus, if we defined "infinite-dimensional" to mean "has an infinite basis", 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"! For this reason, it is best to define an infinite-dimensional vector space to be one spanned by an infinite number of vectors rather than one that has as an infinite basis.

Remark 2.56. (Vector spaces that *don't* have bases?).

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

---

[4]$(\forall \mathbf{v} \in \emptyset \ \mathbf{v} \notin \text{span}(\emptyset - \{\mathbf{v}\}))$ is equivalent to $(\forall \mathbf{v} \ \mathbf{v} \in \emptyset \implies \mathbf{v} \notin \text{span}(\emptyset - \{\mathbf{v}\}))$. Because $\mathbf{v} \in \emptyset$ is false for every $\mathbf{v}$, the overall implication $\mathbf{v} \in \emptyset \implies \mathbf{v} \notin \text{span}(\emptyset - \{\mathbf{v}\})$ is true for every $\mathbf{v}$.

[5]We can always just remove all vectors to obtain $\emptyset$, which is linearly independent

2.2 Linear functions

At the beginning of the previous section, it was said that the two fundamental ideas underlying linear algebra are those of "objects that behave like vectors" and of "functions that preserve the decomposition of their input vectors". We learned that "objects that behave like vectors" are elements of vector spaces. Now, we will investigate "functions that preserve the decomposition of their input vectors", which are more formally referred to as *linear functions*.

Definition 2.57. (Linear function).

Let V and W be vector spaces over a field K. A function $\mathbf{f} : V \to W$ is said to be *linear* iff we have $\mathbf{f}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\mathbf{f}(\mathbf{v}_1) + c_2\mathbf{f}(\mathbf{v}_2)$ for all $c_1, c_2 \in K$ and $\mathbf{v}_1, \mathbf{v}_2 \in V$.

What does this definition really mean, though? It may help understanding to introduce the notation $\mathbf{w}_1 := \mathbf{f}(\mathbf{v}_1)$ and $\mathbf{w}_2 := \mathbf{f}(\mathbf{v}_2)$. With this notation, we see that $\mathbf{f}$ is linear iff we have the following for every $c_1, c_2 \in K$ and $\mathbf{v}_1, \mathbf{v}_2 \in V$:

$$\mathbf{v}_1 \overset{\mathbf{f}}{\mapsto} \mathbf{w}_1 \text{ and } \mathbf{v}_2 \overset{\mathbf{f}}{\mapsto} \mathbf{w}_2$$
$$\Longrightarrow$$
$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \overset{\mathbf{f}}{\mapsto} c_1\mathbf{w}_1 + c_2\mathbf{w}_2$$

So, roughly speaking, $\mathbf{f}$ is linear iff elements in V "interact" in the same way as do their corresponding elements in the image $\mathbf{f}(V)$.

Linear functions are also referred to as *linear transformations*, *linear operators*, or *linear maps*. We will stick with the terminology "linear function".

Remark 2.58. ("Linear" in "linear function" means "vector").

Initially, one might be confused as to what is actually "linear" about a linear function, and ask something like, "What do linear functions have to do with lines?" The answer is that the "linear" in "linear function" is meant to connote "linear element". (Recall that elements of vector spaces are also sometimes called "linear elements"). Linear functions are called such because they are the functions that play nicely with linear elements. A better name for "linear function" would be "vector-respecting function".

We now quickly present an slightly alternative characterization of linear functions. This is the characterization that is most often used for the definition of a linear function in other texts, and is often easier to check than the above definition in practice.

Theorem 2.59. (The most common characterization of linear functions).

Let V and W be vector spaces over a field K. A function $\mathbf{f} : V \to W$ is linear iff

1. $\mathbf{f}(\mathbf{v} + \mathbf{w}) = \mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$.
2. $\mathbf{f}(c\mathbf{v}) = c\mathbf{f}(\mathbf{v})$ for all $c \in K$ and $\mathbf{v} \in V$.

Proof. Left as exercise. □

This characterization of linear functions quickly generalizes to the following fact.

Theorem 2.60. (Generalization of the most common characterization of linear functions).

Let V and W be vector spaces over a field K. A function $\mathbf{f} : V \to W$ is linear iff

$$\mathbf{f}\Big(\sum_{i=1}^{n} c_i\mathbf{v}_i\Big) = \sum_{i=1}^{n} c_i\mathbf{f}(\mathbf{v}_i) \text{ for all } c_1, ..., c_n \in K \text{ and } \mathbf{v}_1, ..., \mathbf{v}_n \in V.$$

Proof. Left as exercise. □

Theorem 2.61. (Every linear function includes $\mathbf{0}$ in its image).

Every linear function includes $\mathbf{0}$ in its image because $\mathbf{f}(0 \cdot \mathbf{v}) = 0 \cdot \mathbf{f}(\mathbf{v}) = \mathbf{0}$ for linear function $\mathbf{f}$ and any vector $\mathbf{v}$.

Somewhat counterintuitively, this means that only lines that pass through the origin $\mathbf{0}$ can be images of linear functions.

Theorem 2.62. (Any basis can get sent to any list of vectors by some linear function).

Let V and W be finite-dimensional vector spaces and let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be a basis of V. For all $\mathbf{w}_1, ..., \mathbf{w}_n \in W$, there exists a linear function sending $\mathbf{e}_i \mapsto \mathbf{w}_i$.

Proof. For any linear function $\mathbf{f}$, we have $\mathbf{f}(\mathbf{v}) = \sum_{i=1}^{n}([\mathbf{v}]_E)_i \mathbf{f}(\mathbf{e}_i)$. We want to construct a linear function $\mathbf{g}$ with $\mathbf{g}(\mathbf{e}_i) = \mathbf{w}_i$. Thinking about replacing the $\mathbf{f}(\mathbf{e}_i)$ in the previous sum with $\mathbf{w}_i$ gives us the idea to define $\mathbf{g}(\mathbf{v}) := \sum_{i=1}^{n}([\mathbf{v}]_E)_i \mathbf{w}_i$. It is straightforward to check that we indeed have $\mathbf{g}(\mathbf{e}_i) = \mathbf{w}_i$ and that $\mathbf{g}$ is linear. $\square$

Now that we are familiar with theoretical characterizations of linear functions, we will investigate linear functions from a geometric perspective.

Remark 2.63. (Examples of linear functions $\mathbb{R}^2 \to \mathbb{R}^2$).
The following are examples of linear functions from $\mathbb{R}^2$ to $\mathbb{R}^2$:

- rotations about the origin
- reflection across a line through the origin
- projection onto a line through the origin

In order to convince yourself that the above functions are linear, check that each above function $\mathbf{f}$ satisfies $\mathbf{f}(\mathbf{v} + \mathbf{w}) = \mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w})$ and $\mathbf{f}(c\mathbf{v}) = c\mathbf{v}$.

The following theorem gives further geometric intuition for how linear functions behave.

Theorem 2.64. Linear functions $\mathbb{R}^n \to \mathbb{R}^n$ fix the origin and keep parallel lines parallel.

Proof. Let $\mathbf{f}$ be a linear function $\mathbb{R}^n \to \mathbb{R}^n$. Theorem 2.61 showed that $\mathbf{f}(\mathbf{0}) = \mathbf{0}$, so it remains to show that $\mathbf{f}$ sends parallel lines to parallel lines.

Consider two parallel lines in $\mathbb{R}^n$ described by $\boldsymbol{\ell}_1(t) = \mathbf{v}_0 + t\mathbf{v}$ and $\boldsymbol{\ell}_2(t) = \mathbf{w}_0 + t\mathbf{v}$. We have $\mathbf{f}(\boldsymbol{\ell}_1(t)) = \mathbf{f}(\mathbf{v}_0) + t\mathbf{f}(\mathbf{v})$ and $\mathbf{f}(\boldsymbol{\ell}_2(t)) = \mathbf{f}(\mathbf{w}_0) + t\mathbf{f}(\mathbf{v})$. These transformed lines are parallel because they have the same direction vector, $\mathbf{f}(\mathbf{v})$. $\square$

Kernel and image of linear functions

Definition 2.65. (Kernel, image of a linear function).

Let V and W be vector spaces, and let $\mathbf{f}: V \to W$ be a linear function. The *kernel* of $\mathbf{f}$ is the set of all vectors that get sent to $\mathbf{0}$ by $\mathbf{f}$:

$$\ker(\mathbf{f}) := \mathbf{f}^{-1}(\{\mathbf{0}\}) = \{\mathbf{v} \in V \mid \mathbf{f}(\mathbf{v}) = \mathbf{0}\}.$$

The *image* of $\mathbf{f}$ is the set of all vectors that are mapped to by $\mathbf{f}$:

$$\operatorname{im}(\mathbf{f}) := \mathbf{f}(V) = \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V \ \mathbf{w} = \mathbf{f}(\mathbf{v})\}.$$

Definition 2.66. (Rank of a linear function).

Let V and W be vector spaces, and let $\mathbf{f}: V \to W$ be 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 2.67. (Kernel and image are vector subspaces).

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

Definition 2.68. (Trivial kernel).

Since $\{\mathbf{0}\}$ is the smallest (in the sense of set-containment) kernel possible for a linear function, we say that the kernel of a linear function is *trivial* iff it is equal to $\{\mathbf{0}\}$.

Theorem 2.69. (One-to-one linear functions have trivial kernels).

Let V and W be vector spaces. A linear function $\mathbf{f}: V \to W$ is one-to-one iff $\mathbf{f}^{-1}(\{\mathbf{0}\}) = \{\mathbf{0}\}$. That is, a linear function is one-to-one iff it has a trivial kernel.

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

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

($\Longleftarrow$). $\mathbf{f}$ is not one-to-one $\iff$ for some $\mathbf{v}_1, \mathbf{v}_2 \in V$ with $\mathbf{v}_1 \neq \mathbf{v}_2$ we have $\mathbf{f}(\mathbf{v}_1) = \mathbf{f}(\mathbf{v}_2) \implies \mathbf{f}(\mathbf{v}_1) - \mathbf{f}(\mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$. Since $\mathbf{v}_1 \neq \mathbf{v}_2$, we know $\mathbf{v}_1 - \mathbf{v}_2 \neq \mathbf{0}$, and so $\mathbf{f}$ has a nontrivial kernel. □

Remark 2.70. The idea in the above proof is that vectors in the preimage of every $\mathbf{w} \in W$ "differ by an element of the kernel". You could prove the following fact to formalize this further: $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{f}^{-1}(\mathbf{w})$ for some $\mathbf{w} \in \mathbf{f}(V)$ if and only if $\mathbf{v}_2 = \mathbf{v}_1 + \mathbf{v}$, where $\mathbf{v}_1 \in V$ and $\mathbf{v} \in \mathbf{f}^{-1}(\{\mathbf{0}\})$.

Theorem 2.71. (Only one-to-one linear functions preserve linear independence).

Let V and W be finite-dimensional vector spaces. A linear function $\mathbf{f}: V \to W$ preserves the linear independence of vectors iff it is one-to-one. That is,

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

Proof.

($\Longleftarrow$). Suppose that $\mathbf{f}$ is one-to-one and that $\mathbf{v}_1, ..., \mathbf{v}_k$ are linearly independent. Since $\mathbf{f}$ is one-to-one, it has a trivial kernel, and thus for any $c_1, ..., c_k \in K$ we have that $\mathbf{f}(c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k) = \mathbf{0}$ implies $c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k = \mathbf{0}$. Since $\mathbf{v}_1, ..., \mathbf{v}_k$ are linearly independent, $c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k = \mathbf{0}$ implies that the c_i's are all 0. In all, we have that "$\mathbf{f}(c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k) = \mathbf{0}$ implies the c_i's are all 0". Since $\mathbf{f}$ is linear, this statement becomes "$c_1\mathbf{f}(\mathbf{v}_1) + ... + c_k\mathbf{f}(v_k) = \mathbf{0}$ implies the c_i's are all 0". Thus $\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_n)$ are linearly independent, as claimed.

($\Longrightarrow$). Suppose that if $\mathbf{v}_1, ..., \mathbf{v}_k$ are linearly independent, then $\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_n)$ 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 $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ for V. Expressing $\mathbf{v}$ relative to E, we have $\mathbf{f}(\mathbf{v}) = \mathbf{f}\Big(\sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{e}_i\Big) = \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{f}(\mathbf{e}_i) = \mathbf{0}$. The hypothesis implies that $\mathbf{f}(\mathbf{e}_1), ..., \mathbf{f}(\mathbf{e}_n)$ are linearly independent, so $([\mathbf{v}]_E)_i = 0$ for all i is the only solution to $\sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{f}(\mathbf{e}_i) = \mathbf{0}$. Thus $([\mathbf{v}]_E)_i = 0$ for all i, i.e., $\mathbf{v} = \mathbf{0}$. □

Theorem 2.72. (Main dimension theorem).

Let V and W be vector spaces, and let $\mathbf{f} : V \to W$ be a linear function. If V is finite-dimensional, then $\mathbf{f}^{-1}(\{\mathbf{0}\})$ and $\mathbf{f}(V)$ are also finite-dimensional, and we have

$$\dim(\mathbf{f}(V)) = \dim(V) - \dim(\mathbf{f}^{-1}(\{\mathbf{0}\})).$$

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 $\mathbf{f}^{-1}(\{\mathbf{0}\}) \subseteq V$, so we can choose a basis $\{\mathbf{e}_1, ..., \mathbf{e}_k\}$ for $\mathbf{f}^{-1}(\{\mathbf{0}\})$. Using the uniqueness of dimension and Lemma 2.50, one can show **add footnote** that it is possible to add vectors $\mathbf{e}_{k+1}, ..., \mathbf{e}_n$ to this basis so that it becomes $\{\mathbf{e}_1, ..., \mathbf{e}_k, \mathbf{e}_{k+1}, ..., \mathbf{e}_n\}$, a basis for V. Since $\dim(\mathbf{f}^{-1}(\{\mathbf{0}\})) = k$ and $\dim(V) = n$, we want to show $\dim(\mathbf{f}(V)) = \dim(V) - \dim(\mathbf{f}^{-1}(\{\mathbf{0}\})) = n - k$; we want to show $\dim(\mathbf{f}(V)) = n - k$.

Since $\mathbf{f} : V \to W$ is linear, we have $\mathbf{f}(\mathbf{v}) = ([\mathbf{v}]_E)_1\mathbf{f}(\mathbf{e}_1) + ... + ([\mathbf{v}]_E)_k\mathbf{f}(\mathbf{e}_k) + ([\mathbf{v}]_E)_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + ... + ([\mathbf{v}]_E)_n\mathbf{f}(\mathbf{e}_n)$. Because $\mathbf{e}_1, ..., \mathbf{e}_k \in \mathbf{f}^{-1}(\{\mathbf{0}\})$, this simplifies to $\mathbf{f}(\mathbf{v}) = ([\mathbf{v}]_E)_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + ... + ([\mathbf{v}]_E)_n\mathbf{f}(\mathbf{e}_n)$.

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

It remains to show $\{\mathbf{e}_{k+1}, ..., \mathbf{e}_n\}$ is a linearly independent set. Suppose for the sake of contradiction it's linearly dependent, i.e., that there are $c_{k+1}, ..., c_n$ not all zero such that $c_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + ... + c_n\mathbf{f}(\mathbf{e}_n) = \mathbf{0}$. By the linearity of $\mathbf{f}$, this is equivalent with $\mathbf{f}(c_{k+1}\mathbf{e}_{k+1} + ... + c_n\mathbf{e}_n) = \mathbf{0}$ for some c_i's not all zero. Thus $c_{k+1}\mathbf{e}_{k+1} + ... + c_n\mathbf{e}_n \in \mathbf{f}^{-1}(\{\mathbf{0}\}) = \text{span}(\{\mathbf{e}_1, ..., \mathbf{e}_k\})$, which means $c_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + ... + c_n\mathbf{f}(\mathbf{e}_n) = d_1\mathbf{e}_1 + ... + d_k\mathbf{e}_k$ for some c_i's and d_i's not all zero. Then $-(d_1\mathbf{e}_1 + ... + d_k\mathbf{e}_k) + c_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + ... + c_n\mathbf{f}(\mathbf{e}_n) = 0$ for some c_i's and d_i's not all zero. But $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is a basis for V, so this cannot happen. Thus $\mathbf{f}(\mathbf{e}_{k+1}), ..., \mathbf{f}(\mathbf{e}_n)$ are linearly independent. □

Linear isomorphisms

Definition 2.73. (Linear isomorphism).

Let V and W be vector spaces over a field K. Iff $\mathbf{f} : V \to W$ is an invertible linear function, i.e. iff it is a bijection[6], then it is called a *linear isomorphism*, or an *isomorphism (of vector spaces)*.

Let's quickly explain this terminology. Recall that when we have a linear function $\mathbf{f} : V \to W$, then elements in V "interact" in the same way as do their corresponding elements in $\mathbf{f}(V)$. When $\mathbf{f}$ is also invertible, then $\mathbf{f}(V)$ is *all* of W, so, not only are all interactions in V "mirrored" in W, but all interactions in W are also mirrored in V! Thus, when $\mathbf{f}$ is linear and invertible, V and W are in some sense the "same" vector space. For this reason, when V and W are isomorphic, we often say that an element $\mathbf{v} \in V$ can be *identified* with an element $\mathbf{w} \in W$.

We write $V \cong W$ to denote that the vector spaces V and W are isomorphic. Note that $\cong$ is an equivalence relation.

The following theorem tells us that the inverse of a linear isomorphism is also a linear isomorphism.

Theorem 2.74. (The inverse of a linear function is also a linear function).

If $\mathbf{f} : V \to W$ is an invertible linear function, then the inverse $\mathbf{f}^{-1}$ is also a linear function.

Proof. Left as exercise. □

Theorem 2.75. (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 $\mathbf{f} : V \to W$ be a linear function. Then $\mathbf{f}$ is one-to-one iff $\mathbf{f}$ is onto. In other words, $\mathbf{f}$ is a linear isomorphism iff it is either one-to-one or onto.

Proof. We use the contrapositive to show that, assuming the hypotheses, $\mathbf{f}$ is not one-to-one iff it is not onto.

$\mathbf{f}$ is not one-to-one if and only if it has a nontrivial kernel, i.e., iff $\dim(\mathbf{f}^{-1}(\{\mathbf{0}\})) > 0$. By the main dimension theorem, this is equivalent to $\dim(V) > \dim(\mathbf{f}(V))$. We assumed $\dim(V) = \dim(W)$, so this is the same as $\dim(W) > \dim(\mathbf{f}(V))$. One can check that if Y and Z are finite-dimensional vector spaces, then $\dim(Y) < \dim(Z)$ iff $Y \subsetneq Z$. So, $\dim(W) > \dim(\mathbf{f}(V))$ is equivalent to equivalent to $W \supsetneq \mathbf{f}(V)$, i.e., to $\mathbf{f}$ not being onto. □

[6]Recall from Theorem 1.10 that any (not necessarily linear) function is invertible iff it is a bijection.

Theorem 2.76. (Finite-dimensional vector spaces have the same dimension iff they are isomorphic).

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

Proof.

($\Longrightarrow$). If $\mathbf{f} : V \to W$ is a linear isomorphism, then $\mathbf{f}$ preserves linear independence of vectors. Therefore, if $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is a basis of n vectors for V then $\{\mathbf{f}(\mathbf{e}_1), ..., \mathbf{f}(\mathbf{e}_n)\}$ is a basis of n vectors for W. Thus $\dim(V) = \dim(W)$.

($\Longleftarrow$). It suffices to show that every n-dimensional vector space is isomorphic to K^n. So, let V be an n-dimensional vector space, and let E be a basis for V. In Definition 2.78 we will define the linear function $[\cdot]_E : V \to K^n$; Theorem 2.79 will show that $[\cdot]_E$ is a linear isomorphism. $\square$

Definition 2.77. (Natural linear isomorphism).

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 "natural isomorphism category theory" online.

Coordinatization of vectors

Definition 2.78. (Coordinates of a finite-dimensional vector relative to a basis).

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

$$\mathbf{v} = ([\mathbf{v}]_E)_1\mathbf{e}_1 + ... + ([\mathbf{v}]_E)_n\mathbf{e}_n.$$

We are guaranteed that such scalars exist because E, being a basis for V, spans V.

Theorem 2.79. ($[\cdot]_E$ is an invertible linear function).

If V is a finite-dimensional vector space with basis E, then $[\cdot]_E : V \to K^{\dim(V)}$ 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 $[c\mathbf{v}]_E = c[\mathbf{v}]_E$.

We have

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

Now we show $[c\mathbf{v}]_E = c[\mathbf{v}]_E$. Since $[\mathbf{v}]_E = \begin{pmatrix} ([\mathbf{v}]_E)_1 \\ \vdots \\ ([\mathbf{v}]_E)_n \end{pmatrix}$, we have $\mathbf{v} = \sum_{i=1}^n ([\mathbf{v}]_E)_i\mathbf{e}_i$, and $c\mathbf{v} = c\sum_{i=1}^n ([\mathbf{v}]_E)_i\mathbf{e}_i = \sum_{i=1}^n c([\mathbf{v}]_E)_i\mathbf{e}_i$. By the definition of $[\cdot]_E$, we see $([c\mathbf{v}]_E)_i = c([\mathbf{v}]_E)_i$, so $[c\mathbf{v}]_E = c[\mathbf{v}]_E$. Therefore $[\cdot]_E$ is linear. $[\cdot]_E$ is invertible because it sends basis vectors to basis vectors, and therefore preserves linear independence. $\square$

2.3 Coordinatization of linear functions with matrices

Standard matrices

In the previous section, we saw that when we have a finite-dimensional vector space V with a basis E, we can represent any vector $\mathbf{v} \in V$ by taking its coordinates $[\mathbf{v}]_E$ relative to E. In this section, we discover that when we have finite-dimensional vector spaces V, W with respective bases E, F, we can also coordinatize a linear function $\mathbf{f} : V \to W$ by making use of the bases E and F. The first step in doing so is to identify said $\mathbf{f}$ with a linear function $K^{\dim(V)} \to K^{\dim(W)}$. The next definition shows us how to produce this linear function.

Definition 2.80. (Induced linear function from $K^{\dim(V)}$ to $K^{\dim(W)}$).

Let V and W be finite-dimensional vector spaces over a field K, and let E and F be the respective bases of V and W. Whenever we have a linear function $\mathbf{f} : V \to W$, there is also an *induced* linear function $\mathbf{f}_{E,F} : K^{\dim(V)} \to K^{\dim(W)}$ for which this diagram commutes:

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

A diagram like the one above is said to "commute" iff the compositions of functions corresponding to different paths through the diagram are the same whenever the paths have the same start and end nodes. So, to say that the above diagram commutes is to say that $[\cdot]_F \circ \mathbf{f} = \mathbf{f}_{E,F} \circ [\cdot]_E$. That is,

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

Concretely, the commutative diagram tells us that we can think of $\mathbf{f}_{E,F}$ as accepting an input that is expressed relative to the basis E for V and producing an output that is expressed relative to the basis F for W.

As we continue our investigations into the coordinatization of a linear function $\mathbf{f} : V \to W$, we can restrict ourselves to the case where $V = K^n$ and $W = K^m$, where K is the field and n and m are positive integers. This is because any linear function $\mathbf{f}$ *not* of this form can immediately be identified[7] with the linear function $\mathbf{f}_{E,F}$ that *is* of this form: just send $\mathbf{f} \mapsto \mathbf{f}_{E,F}$!

The following derivation shows us how to coordinatize a linear function $K^n \to K^m$.

Derivation 2.81. (Standard matrix of a linear function $K^n \to K^m$).

Let K be a field, and consider a linear function $\mathbf{f} : K^n \to K^m$. For any vector $\mathbf{v} \in K^n$, we have

$$\mathbf{v} = \begin{pmatrix} ([\mathbf{v}]_{\hat{\mathfrak{e}}})_1 \\ \vdots \\ 0 \end{pmatrix} + ... + \begin{pmatrix} 0 \\ \vdots \\ ([\mathbf{v}]_{\hat{\mathfrak{e}}})_n \end{pmatrix},$$

and so using the linearity of $\mathbf{f}$, we find that

$$\mathbf{f}(\mathbf{v}) = \mathbf{f}\left(\begin{pmatrix} ([\mathbf{v}]_{\hat{\mathfrak{e}}})_1 \\ \vdots \\ 0 \end{pmatrix} + ... + \begin{pmatrix} 0 \\ \vdots \\ ([\mathbf{v}]_{\hat{\mathfrak{e}}})_n \end{pmatrix}\right) = \mathbf{f}\left(\begin{pmatrix} ([\mathbf{v}]_{\hat{\mathfrak{e}}})_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right) + ... + \mathbf{f}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ ([\mathbf{v}]_{\hat{\mathfrak{e}}})_n \end{pmatrix}\right)$$

$$= ([\mathbf{v}]_{\hat{\mathfrak{e}}})_1 \mathbf{f}\left(\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right) + ... + ([\mathbf{v}]_{\hat{\mathfrak{e}}})_n \mathbf{f}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}\right).$$

The above equation can be written more compactly as

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

[7] The map $\mathbf{f} \mapsto \mathbf{f}_{E,F}$ is a linear isomorphism.

(Recall that the jth component of $\hat{\mathfrak{e}}_i$ is 1 when $i = j$ and 0 otherwise, and that $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ is the standard basis of K^n).

Thinking more about the above, we see that the action of $\mathbf{f}$ on an arbitrary $\mathbf{v} \in K^n$ is determined by $\mathbf{f}(\hat{\mathfrak{e}}_1), ..., \mathbf{f}(\hat{\mathfrak{e}}_n)$. That is, if we know $\mathbf{f}(\hat{\mathfrak{e}}_1), ..., \mathbf{f}(\hat{\mathfrak{e}}_n)$, then we can figure out what $\mathbf{f}$ is!

Formally, we have discovered a function $\mathbf{p}$ that takes as input the ordered list $\begin{pmatrix}\mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n)\end{pmatrix}$, the vector $\mathbf{v} \in K^n$, and produces $\mathbf{f}(\mathbf{v}) \in K^m$ as output:

$$\mathbf{p}\Big(\begin{pmatrix}\mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n)\end{pmatrix}, \mathbf{v}\Big) := ([\mathbf{v}]_E)_1 \mathbf{f}(\hat{\mathfrak{e}}_1) + ... + ([\mathbf{v}]_E)_n \mathbf{f}(\hat{\mathfrak{e}}_n) = \mathbf{f}(\mathbf{v}).$$

We turn our attention to the ordered list of column vectors that is an input to $\mathbf{p}$:

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

This ordered list of n many column vectors from K^m can be interpreted to be a grid of scalars with m rows and n columns. In general, an m by n grid of scalars is called a $m \times n$ *matrix*. ($m \times n$ is read as "m by n").

Of course, the above matrix isn't just "any old matrix": this matrix represents[8]the linear function $\mathbf{f}$! For this reason, the above matrix is called the *standard matrix of* $\mathbf{f} : K^n \to K^m$.

Derivation 2.82. (Matrix-vector product).

The previous derivation showed that a linear function $\mathbf{f} : K^n \to K^m$ is represented by its standard matrix, $\begin{pmatrix}\mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n)\end{pmatrix}$. In this derivation, we formalize the notion of using $\mathbf{f}$'s standard matrix to determine $\mathbf{f}$ itself.

The previous derivation showed that there is a function $\mathbf{p}$ that returns $\mathbf{f}(\mathbf{v})$ when given the standard matrix of $\mathbf{f}$ and a vector $\mathbf{v}$; specifically

$$\mathbf{p}\Big(\begin{pmatrix}\mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n)\end{pmatrix}, \mathbf{v}\Big) := ([\mathbf{v}]_E)_1 \mathbf{f}(\hat{\mathfrak{e}}_1) + ... + ([\mathbf{v}]_E)_n \mathbf{f}(\hat{\mathfrak{e}}_n) = \mathbf{f}(\mathbf{v}).$$

Notice that since the columns of the standard matrix of $\mathbf{f}$ vary over K^m, and since $([\mathbf{v}]_{\hat{\mathfrak{E}}})_1, ..., ([\mathbf{v}]_{\hat{\mathfrak{E}}})_n$ vary[9]over K, we can now restate the action of $\mathbf{p}$ in terms of an arbitrary $m \times n$ matrix $\mathbf{A}$ having ith column $\mathbf{a}_i$ and an arbitrary column vector $\mathbf{v} \in K^n$:

$$\mathbf{p}\left(\underbrace{\begin{pmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_n\end{pmatrix}}_{\mathbf{A}}, \underbrace{\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}}_{\mathbf{v}} \right) = v_1 \mathbf{a}_1 + ... + v_n \mathbf{a}_n.$$

So, $\mathbf{p}$ is really a function that accepts an $m \times n$ matrix and n-dimensional column vector as input and that produces an m-dimensional column vector as output. For this reason, we will call $\mathbf{p}$ the *matrix-vector product*.

No one actually uses the letter $\mathbf{p}$ when notating matrix-vector products. Instead, we simply establish the convention that writing a column vector to the right of a matrix indicates the evaluation of the corresponding matrix-vector product. That is, given an $m \times n$ matrix $\mathbf{A} = \begin{pmatrix}\mathbf{a}_1 \dots & \mathbf{a}_n\end{pmatrix}$ and a column vector $\mathbf{v} \in K^m$, we define

$$\boxed{\mathbf{A}\mathbf{v} = \underbrace{\begin{pmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_n\end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}}_{\mathbf{v}} := v_1 \mathbf{a}_1 + ... + v_n \mathbf{a}_n}$$

Expanding out the columns of $\mathbf{A}$, here is what the above definition looks like when written out more explicitly:

[8] When we say that the standard matrix of $\mathbf{f} : K^n \to K^m$ "represents" $\mathbf{f}$, we mean that the function $\mathbf{F}$ that associates a linear function with its standard matrix is a bijection between the set of linear functions $K^n \to K^m$ and the set of $m \times n$ matrices. Theorem 2.62 implies that $\mathbf{F}$ is onto. Showing that $\mathbf{F}$ is one-to-one is a simple exercise.

[9]The ith column of $\mathbf{f}(\hat{\mathfrak{E}})$ is $\mathbf{f}(\mathbf{e}_i)$. Each $\mathbf{f}(\mathbf{e}_i)$ can vary over K^m because of Theorem 2.62. The $([\mathbf{v}]_E)_i$ vary over K because $\mathbf{v}$ varies over K^n.

$$\underbrace{\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}}_{\mathbf{v}} := v_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \dots + v_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Now that we have the notation of the matrix-vector product, we can restate and expand upon the fact "$\mathbf{p}(\begin{pmatrix} \mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n) \end{pmatrix}, \mathbf{v}) = \mathbf{f}(\mathbf{v})$".

Theorem 2.83. (Characterizing property of standard matrices, verbosely stated).

Let K be a field and let $\mathbf{f} : K^n \to K^m$ be a linear function. The standard matrix $\begin{pmatrix} \mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n) \end{pmatrix}$ of $\mathbf{f}$ is the unique matrix satisfying

$$\begin{pmatrix} \mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n) \end{pmatrix} \mathbf{v} = \mathbf{f}(\mathbf{v}) \text{ for all } \mathbf{v} \in K^n.$$

(The left side of the above equation is a matrix-vector product).

Proof. Derivation 2.81 showed that $\begin{pmatrix} \mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n) \end{pmatrix}$ satisfies the equation. What we have not yet shown is that the standard matrix is the *only* matrix satisfying this equation. To prove this, suppose that some matrix $\mathbf{A}$ satisfies $\mathbf{A}\mathbf{v} = \mathbf{f}(\mathbf{v})$ for all $\mathbf{v} \in K^n$. We need to show $\mathbf{A}$ is in fact equal to the standard matrix of $\mathbf{f}$.

Theorem 2.62 guarantees that there is a linear function $\mathbf{g}$ such that $\begin{pmatrix} \mathbf{g}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{g}(\hat{\mathfrak{e}}_n) \end{pmatrix} = \mathbf{A}$. Thus $\begin{pmatrix} \mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n) \end{pmatrix} \mathbf{v} = \begin{pmatrix} \mathbf{g}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{g}(\hat{\mathfrak{e}}_n) \end{pmatrix} \mathbf{v}$ for all $\mathbf{v} \in K^n$. That is, $\mathbf{f}(\mathbf{v}) = \mathbf{g}(\mathbf{v})$ for all $\mathbf{v} \in K^n$, so $\mathbf{f} = \mathbf{g}$, and $\mathbf{A} = \begin{pmatrix} \mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{f}(\hat{\mathfrak{e}}_n) \end{pmatrix}$, which is the standard matrix of $\mathbf{f}$. □

To make notating standard matrices more compact, we make the following definition.

Definition 2.84. (Function acting on a list).

Let X and Y by sets, let $f : X \to Y$ be a function, and let $L = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}$ be a finite list of elements of X. We define the notation

$$f(L) := \begin{pmatrix} f(x_1) & \dots & f(x_n) \end{pmatrix}.$$

The following theorem illustrates the succinctness of this new notation.

Theorem 2.85. (Compact notation for standard matrix).

Let K be a field and let $\mathbf{f} : K^n \to K^m$ be a linear function. The standard matrix of $\mathbf{f}$ is

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

where, in a slight abuse of notation, we use $\hat{\mathfrak{E}}$ here to denote the *list* $\begin{pmatrix} \hat{\mathfrak{e}}_1 & \dots & \hat{\mathfrak{e}}_n \end{pmatrix}$ whose ith element is $\hat{\mathfrak{e}}_i$ rather than the *set* $\{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ containing $\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n$. (It is necessary to distinguish between lists and sets because sets have no ordering).

Now, we can restate the characterizing property of standard matrices concisely.

Theorem 2.86. (Characterizing property of standard matrices).

Let K be a field and let $\mathbf{f} : K^n \to K^m$ be a linear function. The standard matrix $\mathbf{f}(\hat{\mathfrak{E}})$ of $\mathbf{f}$ is the unique matrix satisfying the following characterizing property:

$$\boxed{\mathbf{f}(\hat{\mathfrak{E}})[\mathbf{v}]_E = \mathbf{f}(\mathbf{v}) \text{ for all } \mathbf{v} \in K^n}$$

(The left side of the above equation is a matrix-vector product).

[segway]

Theorem 2.87. (A composition of linear functions is also linear).

Let V, W, and Y be vector spaces over the same field. If $\mathbf{f} : V \to W$ and $\mathbf{g} : W \to Y$ are linear functions, then the composition $\mathbf{g} \circ \mathbf{f}$ is also a linear function.

Proof. Left as an exercise. □

Since we know that a composition of linear functions is another linear function and that every linear function is represented by the matrix, we naturally ask: "What is the matrix of a composition of linear functions"?

Definition 2.88. (Standard matrix of a composition of linear functions, matrix-matrix product).

Let K be a field, and consider linear functions $\mathbf{f} : K^n \to K^m$ and $\mathbf{g} : K^m \to K^p$. Additionally, let $\hat{\mathfrak{E}} = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be the standard basis for K^n, $\hat{\mathfrak{F}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_m\}$ be the standard basis for K^m, and $\hat{\mathfrak{G}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_p\}$ be the standard basis for K^p.

Since $\mathbf{g} \circ \mathbf{f}$ is a linear function $K^n \to K^p$, it has a standard matrix, $(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}})$:

$$\begin{aligned}
(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}}) &= \begin{pmatrix} (\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{e}}_1) & \dots & (\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{e}}_n) \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{g}(\mathbf{f}(\hat{\mathfrak{e}}_1)) & \dots & \mathbf{g}(\mathbf{f}(\hat{\mathfrak{e}}_n)) \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{g}(\hat{\mathfrak{F}})\mathbf{f}(\hat{\mathfrak{e}}_1) & \dots & \mathbf{g}(\hat{\mathfrak{F}})\mathbf{f}(\hat{\mathfrak{e}}_n) \end{pmatrix}.
\end{aligned}$$

Recalling that $\mathbf{f}(\hat{\mathfrak{e}}_i)$ is the ith column of $\mathbf{f}(\hat{\mathfrak{E}})$, the standard matrix of $\mathbf{f}$, we see that

$$(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}}) = \begin{pmatrix} \mathbf{g}(\hat{\mathfrak{F}})(\mathbf{f}(\hat{\mathfrak{E}}))_1 & \dots & \mathbf{g}(\hat{\mathfrak{F}})(\mathbf{f}(\hat{\mathfrak{E}}))_n \end{pmatrix}.$$

Thus, the standard matrix $(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}})$ depends on the standard matrices $\mathbf{f}(\hat{\mathfrak{E}})$ and $\mathbf{g}(\hat{\mathfrak{E}})$. In other words, we have discovered that there is a function $\mathbf{P}$ that takes the standard matrices $\mathbf{f}(\hat{\mathfrak{E}})$ and $\mathbf{g}(\hat{\mathfrak{F}})$ as input and returns the standard matrix $(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}})$ as output:

$$\mathbf{P}(\mathbf{g}(\hat{\mathfrak{F}}), \mathbf{f}(\hat{\mathfrak{E}})) := \begin{pmatrix} \mathbf{g}(\hat{\mathfrak{F}})(\mathbf{f}(\hat{\mathfrak{E}}))_1 & \dots & \mathbf{g}(\hat{\mathfrak{F}})(\mathbf{f}(\hat{\mathfrak{E}}))_n \end{pmatrix}.$$

Since $\mathbf{f}(\hat{\mathfrak{E}})$ varies over $K^{m \times n}$ and $\mathbf{g}(\hat{\mathfrak{F}})$ varies over $K^{m \times p}$, we can restate the action of $\mathbf{P}$ in terms of arbitrary matrices $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{pmatrix} \in K^{m \times n}$ and $\mathbf{B} \in K^{m \times p}$:

$$\mathbf{P}\left(\mathbf{B}, \underbrace{\begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{pmatrix}}_{\mathbf{A}}\right) := \begin{pmatrix} \mathbf{B}\mathbf{a}_1 & \dots & \mathbf{B}\mathbf{a}_n \end{pmatrix}.$$

Just as was the case with the matrix-vector product $\mathbf{p}$, no one actually uses the letter $\mathbf{P}$ when notating matrix-matrix products. Instead, we establish the convention that writing two matrices of compatible sizes next to each other indicates the evaluation of the corresponding matrix-matrix product. That is, given an $m \times n$ matrix $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{pmatrix}$ and a $p \times m$ matrix $\mathbf{B}$, we define

$$\boxed{\mathbf{B}\mathbf{A} = \mathbf{B} \underbrace{\begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{pmatrix}}_{\mathbf{A}} := \begin{pmatrix} \mathbf{B}\mathbf{a}_1 & \dots & \mathbf{B}\mathbf{a}_n \end{pmatrix}}$$

With the above definition, the standard matrix $(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}})$ of $\mathbf{g} \circ \mathbf{f}$ relative to $\hat{\mathfrak{E}}$ is expressed as the matrix-matrix product $\mathbf{g}(\hat{\mathfrak{F}})\ \mathbf{f}(\hat{\mathfrak{E}})$:

$$(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}}) = \mathbf{g}(\hat{\mathfrak{F}})\ \mathbf{f}(\hat{\mathfrak{E}}).$$

Remark 2.89. (Compatibility of matrices for matrix-matrix products).

We now expand on what was meant when we said the matrix-matrix product $\mathbf{BA}$ is only defined when the sizes of $\mathbf{B}$ and $\mathbf{A}$ are "compatible".

Consider that the composition $\mathbf{g} \circ \mathbf{f}$ of linear functions $\mathbf{f}$ and $\mathbf{g}$ is only defined when the output space of $\mathbf{f}$ is the entire input space of $\mathbf{g}$, i.e., when the dimension of $\mathbf{f}$'s output is the same as the dimension of $\mathbf{g}$'s input. Because of this, the matrix-matrix product $\mathbf{BA}$ of an $m \times n$ matrix with an $r \times s$ matrix $\mathbf{B}$ is only defined when $r = n$, i.e., when $\mathbf{B}$ has as many columns as $\mathbf{A}$ has rows.

Remark 2.90. Expanding out the columns of $\mathbf{BA}$, here is what the matrix-matrix product $\mathbf{BA}$ looks like when written out more explicitly:

$$\underbrace{\begin{pmatrix} b_{11} & \dots & b_{1m} \\ \vdots & \dots & \vdots \\ b_{p1} & \dots & b_{pm} \end{pmatrix}}_{\mathbf{B}} \underbrace{\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}}_{\mathbf{A}} := \left(\underbrace{\begin{pmatrix} b_{11} & \dots & b_{1m} \\ \vdots & \dots & \vdots \\ b_{p1} & \dots & b_{pm} \end{pmatrix}}_{\mathbf{B}} \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} \quad \dots \quad \underbrace{\begin{pmatrix} b_{11} & \dots & b_{1m} \\ \vdots & \dots & \vdots \\ b_{p1} & \dots & b_{pm} \end{pmatrix}}_{\mathbf{B}} \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right)$$

$$= \begin{pmatrix} b_{11}a_{11}...b_{1m}a_{m1} & \dots & b_{11}a_{1n} + ... + b_{1m}a_{mn} \\ \vdots & & \vdots \\ b_{p1}a_{11}...b_{pm}a_{m1} & \dots & b_{p1}a_{1n} + ... + b_{pm}a_{mn} \end{pmatrix}.$$

Don't try to make too much sense of this now. Theorem 2.102 makes thinking about entries of matrix-matrix products much more tractable.

Matrices

The previous section showed us that matrices are important because they can be used to represent linear functions. We now provide explanations for how ideas about linear functions translate over to results ideas about matrices.

First, we restate the earlier hasty definition of "matrix"; in this definition we also define some new notation for matrix entries and for the set of all $m \times n$ matrices over a field.

Definition 2.91. (Matrix).

Let K be a field. An $m \times n$ *matrix (with entries in K)* is a "grid" of elements of K with m rows and n columns. For example, the following is a matrix with three rows and two columns that has entries in $\mathbb{R}$:

$$\begin{pmatrix} -1 & \frac{3}{4} \\ \pi & 0 \\ 5 & -11 \end{pmatrix}.$$

The entry in the ith row and jth column of a matrix is called the ij *entry* of that matrix. (For example, the above matrix has a 21 entry of π). Specifying matrices by describing their ijth entry is relatively common. We write "$\mathbf{A} = (a_{ij})$" iff $\mathbf{A}$ is the matrix with ij entry a_{ij}.

We also define $K^{m \times n}$ to be the set of $m \times n$ matrices with entries in K.

Definition 2.92. (Identity matrix).

Let K be a field, and consider the identity function on K^n, which is the function $\mathbf{I}_{K_n} : K^n \to K^n$ defined by $\mathbf{I}_{K^n}(\mathbf{v}) = \mathbf{v}$. The standard matrix of $\mathbf{I}_{K^n}$ relative to the standard basis $\hat{\mathfrak{E}}$ is called the *$(n \times n)$ identity matrix.* Notice that since the ith column of the identity matrix is $\mathbf{I}_{K_n}(\hat{\mathfrak{e}}_i) = \hat{\mathfrak{e}}_i$, it follows that the identity matrix has a diagonal of 1's, with 0's everywhere else; its ij entry is 1 if $i = j$ and 0 if $i \neq j$.

The 3×3 identity matrix, for example, is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We denote the identity matrix by $\mathbf{I}$ rather than by the more verbose notation $\mathbf{I}_{K^n}(\hat{\mathfrak{E}})$, and infer n from context.

This following topic of "matrix transposes" does not become significant until Chapter 4. We give the definition anyhow because matrix transposes do appear occasionally before Chapter 4, even if not in a theoretically significant way.

Definition 2.93. (Transpose of a matrix).

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix with entries in a field K. The *transpose* $\mathbf{A}^\top$ *of* $\mathbf{A}$ *is the matrix whose ij entry is the ji entry of* $\mathbf{A}$: $\mathbf{A}^\top := (a_{ji})$.

For any matrix $\mathbf{A}$, the columns of $\mathbf{A}^\top$ are the rows of $\mathbf{A}$, and the rows of $\mathbf{A}^\top$ are the columns of $\mathbf{A}$.

Now we return to the relationship between linear functions and matrices.

Theorem 2.94. (Linear functions from matrices).

If $\mathbf{A}$ is an $m \times n$ matrix with entries in a field K, then the function $\mathbf{v} \mapsto \mathbf{A}\mathbf{v}$ is linear, and $\mathbf{A}$ is the standard matrix of this function.

Proof. Define $\mathbf{f}$ by $\mathbf{f}(\mathbf{v}) = \mathbf{A}\mathbf{v}$. We will show that $\mathbf{f}$ is linear and that $\mathbf{A} = \mathbf{f}(\hat{\mathfrak{E}})$.

Because of Theorem 2.62, we can interpret the ith column $\mathbf{a}_i$ of $\mathbf{A}$ to be $\mathbf{g}(\hat{\mathfrak{e}}_i)$ for some linear function $\mathbf{g}$. Thus $\mathbf{A} = \mathbf{g}(\hat{\mathfrak{E}})$, and so $\mathbf{f}(\mathbf{v}) = \mathbf{A}\mathbf{v} = \mathbf{g}(\hat{\mathfrak{E}})\mathbf{v} = \mathbf{g}(\hat{\mathfrak{E}})[\mathbf{v}]_{\hat{\mathfrak{E}}}$. The characterizing property of standard matrices says that $\mathbf{g}(\hat{\mathfrak{E}})[\mathbf{v}]_{\hat{\mathfrak{E}}} = \mathbf{g}(\mathbf{v})$, so we have $\mathbf{f}(\mathbf{v}) = \mathbf{g}(\mathbf{v})$ for all $\mathbf{v} \in K^n$, i.e. $\mathbf{f} = \mathbf{g}$. Since $\mathbf{g}$ is linear, $\mathbf{f}$ is also linear.

The standard matrix $\mathbf{f}(\hat{\mathfrak{E}})$ of $\mathbf{f}$ has ith column $\mathbf{f}(\hat{\mathfrak{e}}_i) = \mathbf{A}\hat{\mathfrak{e}}_i$. Compute the matrix-vector product $\mathbf{A}\hat{\mathfrak{e}}_i$ to verify that $\mathbf{A}\hat{\mathfrak{e}}_i = \mathbf{a}_i$, where $\mathbf{a}_i$ is the ith column of $\mathbf{A}$. This tells us that $\mathbf{f}(\hat{\mathfrak{e}}_i) = \mathbf{a}_i$, which means $\mathbf{f}(\hat{\mathfrak{E}}) = \mathbf{A}$. $\square$

Theorem 2.95. (Properties of the matrix-vector product).

Practically speaking, the linearity of the function $\mathbf{v} \mapsto \mathbf{A}\mathbf{v}$ translates into properties of the matrix-vector product: we have $\mathbf{A}(\mathbf{v} + \mathbf{w}) = \mathbf{A}\mathbf{v} + \mathbf{A}\mathbf{w}$ and $\mathbf{A}(c\mathbf{v}) = c(\mathbf{A}\mathbf{v})$ for any matrix $\mathbf{A}$, column vectors $\mathbf{v}$ and $\mathbf{w}$, and scalar c.

Theorem 2.96. (Bijection between linear functions and matrices).

Let V and W be finite-dimensional vector spaces over a field K. The set of linear functions from V to W and the set of $\dim(W) \times \dim(V)$ matrices with entries in K are in bijection.

Proof. The set of linear functions $V \to W$ is in bijection with the set of linear functions $K^{\dim(V)} \to K^{\dim(W)}$, since the functions $\mathbf{f} \mapsto \mathbf{f}_{E,F} = [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E$ and $\mathbf{g} \mapsto [\cdot]_F^{-1} \circ \mathbf{g} \circ [\cdot]_E^{-1}$ are inverses. Thus, it suffices to show the set of linear functions $K^{\dim(V)} \to K^{\dim(W)}$ is in bijection with $K^{\dim(V) \times \dim(W)}$. We claim that the function $\mathbf{F}$ sending a linear function $\mathbf{f} : K^{\dim(V)} \to K^{\dim(W)}$ to its standard matrix $\mathbf{f}(\hat{\mathfrak{E}}) \in K^{\dim(W) \times \dim(V)}$ is a bijection. Derivation 2.81 hints[10] at a proof that $\mathbf{F}$ is one-to-one, and Theorem 2.62 shows that $\mathbf{F}$ is onto. $\square$

Restate definition of matrix-vector product for convenience:

Definition 2.97. (Matrix-vector product).

Let K be a field. Given $\mathbf{A} \in K^{m \times n}$ with ith column $\mathbf{a}_i$ and $\mathbf{v} \in K^n$ with ith entry v_i, we define the *matrix-vector product* $\mathbf{Av}$ to be the following:

$$\mathbf{Av} = \underbrace{\begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}}_{\mathbf{v}} := v_1 \mathbf{a}_1 + ... + v_n \mathbf{a}_n.$$

Theorem 2.98. (ith entry of matrix-vector product).

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix with entries in a field K and let $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in K^n$ be a column vector.

Referring to the definition of matrix-vector product in Derivation 2.81, we see the matrix-vector product $\mathbf{Av}$ is equal to the following:

$$\mathbf{Av} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ \vdots \\ \vdots \\ v_n \end{pmatrix} = v_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{m1} \end{pmatrix} + ... + v_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{in} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} v_1 a_{11} + ... + v_n a_{1n} \\ \vdots \\ v_1 a_{i1} + ... + v_n a_{in} \\ \vdots \\ v_1 a_{m1} + ... + v_n a_{mn} \end{pmatrix}.$$

Therefore, the ith entry $(\mathbf{Av})_i$ of $\mathbf{Av}$ is $v_i a_{i1} + ... + v_n a_{in}$, which is

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

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

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = v_1 w_1 + ... + v_n w_n.$$

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

The last section of this chapter discusses the dot product in depth.

[segway to compositions of linear functions]

We now state some facts about matrix-matrix products. First, we restate the definition of the matrix-matrix product for convenience.

[10] If $\mathbf{f}$ and $\mathbf{g}$ are distinct linear functions, then, because linear functions are determined by what they do to the standard basis, there must be some $\hat{\mathfrak{e}}_i$ for which $\mathbf{f}(\hat{\mathfrak{e}}_i) \neq \mathbf{g}(\hat{\mathfrak{e}}_i)$. This causes the standard matrices of $\mathbf{f}$ and $\mathbf{g}$ to differ (in the ith column).

Definition 2.99. (Matrix-matrix product).

Let K be a field. Given $\mathbf{A} \in K^{m \times n}$ with ith column $\mathbf{a}_i$ and $\mathbf{B} \in K^{p \times m}$, we define the *matrix-matrix product* $\mathbf{BA}$ to be the following:

$$\mathbf{BA} = \mathbf{B}\underbrace{\begin{pmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_n\end{pmatrix}}_{\mathbf{A}} := \begin{pmatrix}\mathbf{Ba}_1 & \dots & \mathbf{Ba}_n\end{pmatrix}.$$

Remark 2.100. (Matrix-matrix products are associative).

One would expect that $(\mathbf{BA})\mathbf{v} = \mathbf{B}(\mathbf{Av})$ for all column vectors $\mathbf{v}$ when $\mathbf{A}$ and $\mathbf{B}$ are "compatible" matrices. This is indeed true because the corresponding linear functions $\mathbf{f}$ and $\mathbf{g}$ satisfy $(\mathbf{g} \circ \mathbf{f})(\mathbf{v}) = \mathbf{g}(\mathbf{f}(\mathbf{v}))$.

Definition 2.101. (Invertibility of matrices).

Let $\mathbf{A}$ be an $m \times n$ matrix with entries in a field K. We say that $\mathbf{A}$ is *invertible* iff the function $\mathbf{v} \mapsto \mathbf{Av}$ is invertible. If $\mathbf{A}$ is invertible, then we use $\mathbf{A}^{-1}$ to denote the standard matrix of the inverse of $\mathbf{v} \mapsto \mathbf{Av}$. We have $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{AA}^{-1}$ for all invertible matrices $\mathbf{A}$.

Since only same-dimension vector spaces can be linearly isomorphic, we must have $n = m$ for an $m \times n$ matrix $\mathbf{A}$ to be invertible. That is, $\mathbf{A}$ must be a *square matrix* for it to be invertible.

Theorem 2.102. (ij entry of matrix-matrix product).

Let K be a field, let $\mathbf{A} = (a_{ij}) \in K^{m \times n}$, and let $\mathbf{B} = (b_{ij}) \in K^{m \times p}$. The ij entry of the matrix-matrix product $\mathbf{BA}$ can be computed by using the definition of the matrix-matrix product (Theorem 2.88) together with the fact that the ith entry of the matrix-vector product $\mathbf{Av}$ is (ith row of $\mathbf{A}$)$\cdot\mathbf{v}$, where $\cdot$ is the dot product. We have

$$\mathbf{BA} = \mathbf{B}\begin{pmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_n\end{pmatrix} = \begin{pmatrix}\mathbf{Ba}_1 & \dots & \mathbf{Ba}_n\end{pmatrix} = \begin{pmatrix}\mathbf{b}_1 \cdot \mathbf{a}_1 & \dots & \mathbf{b}_1 \cdot \mathbf{a}_n \\ \vdots & & \vdots \\ \mathbf{b}_m \cdot \mathbf{a}_1 & \dots & \mathbf{b}_m \cdot \mathbf{a}_n\end{pmatrix},$$

where $\mathbf{a}_i$ is the ith column of $\mathbf{A}$ and $\mathbf{b}_i$ is the ith row of $\mathbf{B}$. So the ij entry $(\mathbf{BA})_{ij}$ of $\mathbf{BA}$ is $\mathbf{b}_i \cdot \mathbf{a}_j$, which is

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

As was the case in Theorem 2.98, what we mean by "ith row of $\mathbf{B}$" in the boxed equation is "column vector that contains entries of ith row of $\mathbf{A}$".

Remark 2.103. (Matrix pedagogy).

Most linear algebra texts present the relationship between linear functions and matrices in the following way: first define matrices in the context of systems of linear equations (we have not seen how matrices are related to systems of linear equations), then define a linear function to be one for which $\mathbf{f}(\mathbf{v}+\mathbf{w}) = \mathbf{f}(\mathbf{v})+\mathbf{f}(\mathbf{w})$ and $\mathbf{f}(c\mathbf{v}) = c\mathbf{f}(\mathbf{v})$ for all vectors $\mathbf{v}, \mathbf{w}$ and scalars c, and then prove that each linear function has a standard 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 2.81. (Furthermore, while systems of linear equations are an important application of linear algebra, and while their study does enhance our knowledge about the kernels of linear functions, they should not be the starting point).

Oftentimes, linear algebra texts present the formula for the ith entry of a matrix-vector product and the formula for the ij entry of a matrix-matrix product as facts that should be memorized rather than understood. Be wary of this! You *should not* memorize these formulas. If you can't quite remember them, try to derive them by starting with the fact that linear functions on finite-dimensional vector spaces are determined by what they do to bases, and by following the derivations given in this book!

Matrices relative to bases

We started this section- "Coordinatization of linear functions with matrices"- by noting that if V and W are finite-dimensional vector spaces, we can effectively study linear functions $V \to W$ by studying linear functions $K^{\dim(V)} \to K^{\dim(W)}$. Then, we showed that because a linear function $K^n \to K^m$ is determined by how it acts on a basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ of V, any linear function $\mathbf{f} : K^n \to K^m$ is represented by the matrix $\mathbf{f}(\hat{\mathfrak{E}})$.

In Theorem 2.96, we saw that this result about representations of linear functions $K^n \to K^m$ "ripples up" to a result about representations of linear functions $V \to W$: that is, every linear function $V \to W$ is represented by a $\dim(V) \times \dim(W)$ matrix. The theorem does not tell us how to compute the matrix representation of an arbitrary linear function, however. We investigate this now.

Derivation 2.104. (Matrix relative to bases).

Let V and W be finite-dimensional vector spaces over a field K with bases E and F, and consider a linear function $\mathbf{f} : V \to W$.

The standard matrix $\mathbf{f}_{E,F}(\hat{\mathfrak{E}})$ of $\mathbf{f}_{E,F} : K^{\dim(V)} \to K^{\dim(W)}$ represents $\mathbf{f}_{E,F}$, and $\mathbf{f}_{E,F}$ represents $\mathbf{f}$, so[11] we conclude by transitivity that $\mathbf{f}_{E,F}(\hat{\mathfrak{E}})$ represents $\mathbf{f}$. For this reason, we define the *matrix of* $\mathbf{f}$ *relative to* E *and* F to be the standard matrix $\mathbf{f}_{E,F}(\hat{\mathfrak{E}})$ of $\mathbf{f}_{E,F}$.

Now, let's investigate $\mathbf{f}_{E,F}(\hat{\mathfrak{E}})$ itself. We have $\mathbf{f}_{E,F}(\hat{\mathfrak{E}}) = ([\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1})(\hat{\mathfrak{E}}) = \Big[\mathbf{f}\Big([\cdot]_E^{-1}(\hat{\mathfrak{E}})\Big)\Big]_F$. Since $[\mathbf{e}_i]_E = \hat{\mathfrak{e}}_i$, we have $[\cdot]_E^{-1}(\hat{\mathfrak{e}}_i) = \mathbf{e}_i$ and thus $[\cdot]_E^{-1}(\hat{\mathfrak{E}}) = E$. So:

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

Thus, the matrix of a linear function $\mathbf{f} : V \to W$ relative to the bases E and F for V and W is $[\mathbf{f}(E)]_F$. Explicitly, $[\mathbf{f}(E)]_F$ looks like this:

$$[\mathbf{f}(E)]_F = \begin{pmatrix} [\mathbf{f}(\mathbf{e}_1)]_F & \dots & [\mathbf{f}(\mathbf{e}_n)]_F \end{pmatrix}$$

The characterizing property of standard matrices for linear functions $K^n \to K^m$ gives us the following equivalent statements:

$$\begin{aligned}
\mathbf{f}_{E,F}(\mathbf{v}) &= \mathbf{f}_{E,F}(\hat{\mathfrak{E}})\ \mathbf{v} \text{ for all } \mathbf{v} \in K^n \\
([\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1})(\mathbf{v}) &= [\mathbf{f}(E)]_F\ \mathbf{v} \text{ for all } \mathbf{v} \in K^n \\
([\cdot]_F \circ \mathbf{f})(\mathbf{v}_1) &= [\mathbf{f}(E)]_F[\mathbf{v}_1]_E \text{ for all } \mathbf{v}_1 \in V \\
[\mathbf{f}(\mathbf{v}_1)]_F &= [\mathbf{f}(E)]_F[\mathbf{v}_1]_E \text{ for all } \mathbf{v}_1 \in V
\end{aligned}$$

So, we have the following characterizing property for matrices relative to bases:

$$\boxed{[\mathbf{f}(\mathbf{v})]_F = [\mathbf{f}(E)]_F[\mathbf{v}]_E \text{ for all } \mathbf{v} \in V}$$

This characterizing property tells a similar story as does the commutative diagram that describes $\mathbf{f}_{E,F}$: it says that we can think of the function $\mathbf{u} \mapsto [\mathbf{f}(E)]_F\mathbf{u}$ as accepting an input that is expressed relative to the basis E for V and as producing an output that is expressed relative to the basis F for W.

Remark 2.105. (Standard matrix as special case of matrix relative to bases).

The standard matrix $\mathbf{f}(\hat{\mathfrak{E}})$ of a linear function $\mathbf{f} : K^n \to K^m$ is the matrix $\mathbf{f}(\hat{\mathfrak{E}}) = [\mathbf{f}(\hat{\mathfrak{E}})]_{\hat{\mathfrak{E}}}$ of $\mathbf{f} : K^n \to K^m$ relative to the bases $\hat{\mathfrak{E}}$ and $\hat{\mathfrak{E}}$.

Not all linear algebra resources use the above notion of matrices relative to bases and instead investigate matrices of the form $\mathbf{E}^{-1}\mathbf{A}\mathbf{E}$. The following theorem explains that such matrices arise as a special case of matrices relative to bases.

Theorem 2.106. (Matrices of linear functions $K^n \to K^n$ relative to E and E).

Let K be a field, let E be a basis for K^n, and consider a linear function $\mathbf{f} : K^n \to K^n$. The matrix $[\mathbf{f}(E)]_E$ of $\mathbf{f}$ relative to E and E is

$$\mathbf{f}(E)_E = \mathbf{E}^{-1}\mathbf{f}(\hat{\mathfrak{E}})\mathbf{E},$$

where $\mathbf{E}$ is the matrix whose ith column is $\mathbf{e}_i$.

[11] Again, when we say that $\mathbf{h}$ *represents* $\mathbf{g}$, we mean that the map sending $\mathbf{g} \mapsto \mathbf{h}$ is a bijection.

Proof. The matrix $[\mathbf{f}(E)]_E$ of $\mathbf{f}$ relative to E and E is the standard matrix $\mathbf{f}_{E,E}(\hat{\mathfrak{E}})$ of $\mathbf{f}_{E,E} = [\cdot]_E \circ \mathbf{f} \circ [\cdot]_E^{-1}$, which is $\mathbf{f}_{E,E}(\hat{\mathfrak{E}}) = [\cdot]_E(\hat{\mathfrak{E}})\ \mathbf{f}(\hat{\mathfrak{E}})\ [\cdot]_E^{-1}(\hat{\mathfrak{E}})$. To complete the proof, we will determine the standard matrices $[\cdot]_E(\hat{\mathfrak{E}})$ and $[\cdot]_E^{-1}(\hat{\mathfrak{E}})$.

Since $[\mathbf{e}_i]_E = \hat{\mathfrak{e}}_i$, we have $[\cdot]_E^{-1}(\hat{\mathfrak{e}}_i) = \mathbf{e}_i$ and[12] thus $[\cdot]_E^{-1}(\hat{\mathfrak{E}}) = \mathbf{E}$, where $\mathbf{E}$ is the matrix with ith column $\mathbf{e}_i$. Also, since $[\cdot]_E$ is the inverse of $[\cdot]_E^{-1}$, we have $[\cdot]_E(\hat{\mathfrak{E}}) = ([\cdot]_E^{-1}(\hat{\mathfrak{E}}))^{-1} = \mathbf{E}^{-1}$. Using the two facts $[\cdot]_E(\hat{\mathfrak{E}}) = \mathbf{E}^{-1}$ and $[\cdot]_E^{-1}(\hat{\mathfrak{E}}) = \mathbf{E}$, we conclude that $([\cdot]_E \circ \mathbf{f} \circ [\cdot]_E^{-1})(\hat{\mathfrak{E}}) = [\cdot]_E(\hat{\mathfrak{E}})\ \mathbf{f}(\hat{\mathfrak{E}})\ [\cdot]_E^{-1}(\hat{\mathfrak{E}}) = \mathbf{E}^{-1}\mathbf{f}(\hat{\mathfrak{E}})\mathbf{E}$. □

Change of basis

We will now discover how when we have a finite-dimensional vector space with bases E and F we can relate $[\mathbf{v}]_E$ to $[\mathbf{v}]_F$ for any vector $\mathbf{v}$.

Theorem 2.107. (Change of basis).

Let V be a finite-dimensional vector space with bases $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$.

The characterizing property of matrices relative to E and F says that for any linear function $\mathbf{f} : V \to V$, we have $[\mathbf{f}(\mathbf{v})]_F = [\mathbf{f}(E)]_F[\mathbf{v}]_E$. In particular, when $\mathbf{f}$ is the identity $\mathbf{I}_V$ on V, we obtain

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

where $\mathbf{E}$ is the list of vectors whose ith entry is $\mathbf{e}_i$, so that $[\mathbf{E}]_F$ is the matrix whose ith column is $[\mathbf{e}_i]_F$.

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.

Remark 2.108. (Alternate proof of change of basis).

Here is an alternate proof to the above. The matrix of $[\cdot]_F$ relative to E and $\hat{\mathfrak{E}}$ is $[[\cdot]_F(E)]_{\hat{\mathfrak{E}}} = [\cdot_F(E)] = [\mathbf{E}]_F$. Using the characterizing property of matrices relative to bases, we have $[\mathbf{v}]_F = [[\cdot]_F(E)]_{\hat{\mathfrak{E}}}[\mathbf{v}]_E = [\mathbf{E}]_F[\mathbf{v}]_E$.

Theorem 2.109. $((\mathbf{I}_V)_{E,F}^{-1} = (\mathbf{I}_V)_{F,E})$.

Let V be a finite-dimensional vector space with bases E and F. The identity function $\mathbf{I}_V : V \to V$ 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. Since $\mathbf{f}_{E,F} = [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}$, then $(\mathbf{I}_V)_{E,F} = [\cdot]_F \circ [\cdot]_E^{-1}$ and $(\mathbf{I}_V)_{F,E} = [\cdot]_E \circ [\cdot]_F^{-1}$. We clearly have $(\mathbf{I}_V)_{E,F}^{-1} = (\mathbf{I}_V)_{F,E}$. □

Theorem 2.110. (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 $[\cdot]_E$, we have

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

where $\mathbf{F}$ is the matrix whose ith column is $\mathbf{f}_i$.

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

Remark 2.111. (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 2.107. This approach for proving Theorem 2.107 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.

[12] In Derivation 2.104 we used the facts from the sentence this footnote appears in to conclude $[\cdot]_E^{-1}(\hat{\mathfrak{E}}) = E$ rather than $[\cdot]_E^{-1}(\hat{\mathfrak{E}}) = \mathbf{E}$. We are able to use the latter matrix notation $\mathbf{E}$ now because we are in the special situation where E is a basis of K^n, so the list E of vectors is a list of *column* vectors, i.e. a matrix.

2.4 Systems of linear equations with matrices

This section presents how systems of linear equations can be solved by using concepts from linear algebra.

Reading this section is entirely unnecessary for the remaining content in this book. This treatment of systems of linear equations has only been provided to serve as a superior alternative to the traditional approach of teaching linear algebra (which is: start with systems of linear equations, define the matrix-vector product, and then haphazardly discover the other linear algebra material we have covered). The approach we take emphasizes that linear algebra should not be introduced as a field of study that is discovered by starting with systems of linear equations[13], and that systems of linear equations should be viewed as an application of linear algebra.

Before we study systems of linear equations, we will discuss the equation of a plane.

Derivation 2.112. (Equation of an n-dimensional plane).

[In all, we have seen that an n-dimensional plane can be represented as a set of points $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ which satisfy an equation of the form

$$a_1x_1 + ... + a_nx_n = b.$$

]

An equation whose solution set is a plane is often called a *equation for a plane*. Less commonly, such an equation is called a *linear equation*[14].

Definition 2.113. (System of linear equations).

A system of linear equations is simply a set of linear equations such as the following:

$$\begin{aligned} a_{1,1}x_1 + ... + a_{1,n}x_n &= b_1, \\ a_{2,1}x_1 + ... + a_{2,n}x_n &= b_2, \\ &\vdots \\ a_{m,1}x_1 + ... + a_{m,n}x_n &= b_m. \end{aligned}$$

Just as was the case with linear equations, when one considers a system of linear equations, one is typically interested in the set of points $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ that satisfy all equations in the system.

Derivation 2.114. (Number of solutions to systems of linear equations).

no solutions, infinitely many solutions, or exactly one solution.

Proof. three different proofs: (1) reasoning about n-dimensional planes and induction; (2) reasoning about $\mathbf{f}(\mathbf{x}) = \mathbf{b}$ where $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is the linear function representing the system and considering cases $n < m$, $n = m$, $n > m$; (3) reasoning about $\mathbf{A}\mathbf{x} = \mathbf{b}$, RREFing, and considering analogous cases to ones in (2)

(1)

the intersection of two n-dimensional planes is a plane whose dimension is either n or $n-1$.

consider a system of linear equations. the solution set to the system is the intersection of the m many n-dimensional planes $P_{11}, ..., P_{1m}$ represented by the m equations in n unknowns. Set $P_{21} := P_{11} \cap P_{12}, P_{22} := P_{13} \cap P_{14}$, etc., with $P_{2\ \text{floor}(m/2)}$ being $P_{1\ m-1} \cap P_{1m}$ when m is even and P_{1m} when m is odd. We claim that it is possible to choose the P_{1i} so that the P_{2j} are whatever we want them to be. (Proof of claim: easy when m is even. When m is odd, use the case when m is even and then simply choose P_{1n} to be whatever we want). □

Derivation 2.115. (Systems of linear equations correspond to vector equations).

Consider the system of linear equations

[13] Linear algebra *did* historically grow out of the study of systems of linear equations. Just remember that the historical approach is not always the most enlightening one!

[14] Although linear equations are related to linear functions (since the dot product is a bilinear function), the word "linear“ in "linear function" has a different meaning than it does in "linear equation"

$$
\begin{aligned}
a_{1,1}x_1 + \ldots + a_{1,n}x_n &= b_1, \\
a_{2,1}x_1 + \ldots + a_{2,n}x_n &= b_2, \\
&\vdots \\
a_{m,1}x_1 + \ldots + a_{m,n}x_n &= b_m.
\end{aligned}
$$

We can represent this system of equations in an alternate form to the above by reorganizing information. If we collect all the coefficents $a_{i,j}$ into a matrix $\mathbf{A}$,

$$
\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ a_{2,1} & \cdots & a_{2,n} \\ & \vdots & \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix},
$$

and also collect the x_i and b_i into a vectors $\mathbf{x}$ and $\mathbf{b}$,

$$
\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},
$$

then the above system of equations corresponds to the tuple $(\mathbf{A}, \mathbf{x}, \mathbf{b})$. We can go a step further, however.

Notice that the system of linear equations that arises from the vector equation $\mathbf{Ax} = \mathbf{b}$ is exactly the system of linear equations we started with.

The vector equation

$$
\underbrace{\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ a_{2,1} & \cdots & a_{2,n} \\ & \vdots & \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}}_{\mathbf{b}}
$$

is equivalent to

$$
\begin{pmatrix} \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \end{pmatrix} \cdot \mathbf{x} \\ \begin{pmatrix} a_{2,1} & \cdots & a_{2,n} \end{pmatrix} \cdot \mathbf{x} \\ \vdots \\ \begin{pmatrix} a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \cdot \mathbf{x} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},
$$

which is equivalent to

$$
\begin{pmatrix} a_{1,1}x_1 + \ldots + a_{1,n}x_n \\ a_{2,1}x_1 + \ldots + a_{2,n}x_n \\ \vdots \\ a_{m,1}x_1 + \ldots + a_{m,n}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.
$$

The set of equations obtained by equating the components of the vectors on either sides of the equality is the original system of equations we began with.

Theorem 2.116. (Row operations are linear transformations).

Remark 2.117. (Column operations).

Lemma 2.118. (A linear function is invertible iff it is a composition of elementary linear functions).

Equivalently, a matrix is invertible iff it is a product of elementary matrices.

Proof. □

Theorem 2.119. (A matrix is invertible iff its RREF is the identity matrix).

Proof.
($\Longrightarrow$). Consider an arbitrary invertible matrix $\mathbf{A}$. Since $\mathbf{A}$ is invertible, it is a product of elementary matrices, $\mathbf{A} = \mathbf{E}_i ... \mathbf{E}_1$.
($\Longleftarrow$). Suppose $\text{rref}(\mathbf{A}) = \mathbf{I}$. Since RREF is obtained by performing a sequence of row operations, and since performing a row operation is equivalent to multiplying on the left by an elementary matrix, we have $\text{rref}(\mathbf{A}) = \mathbf{E}_i ... \mathbf{E}_1$ for some elementary matrices $\mathbf{E}_1, ..., \mathbf{E}_i$. Thus $\mathbf{E}_i ... \mathbf{E}_1 \mathbf{A} = \mathbf{I}$. Elementary matrices are invertible, so we see that
[...]
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. □

2.5 Eigenvectors and eigenvalues

[This section not required for rest of book]

Definition 2.120. (Eigenvectors and eigenvalues of a linear function).
Let V and W be vector spaces over a field K, and let $\mathbf{f} : V \to W$ be a linear function. A vector $\mathbf{v} \in V$ is said to be an *eigenvector (of* $\mathbf{f}$*)* iff there is a scalar $c \in K$ such that $\mathbf{f}(\mathbf{v}) = c\mathbf{v}$. Iff $\mathbf{v}$ is indeed an eigenvector of $\mathbf{f}$, then the $c \in K$ for which $\mathbf{f}(\mathbf{v}) = c\mathbf{v}$ is said to be the *eigenvalue (corresponding to* $\mathbf{v}$*)*.
In other words, the eigenvectors of a linear function are the vectors that get sent to scalar multiples of themselves by the function, and the eigenvalues corresponding to those eigenvectors are the scalars involved in said scalar multiples.

Remark 2.121. ("Characteristic vectors").
The word "eigen" within "eigenvector" is German. Back when eigenvectors were first defined, English-speaking mathematicians called them "characteristic vectors". This terminology is not used today, but is helpful to keep it in mind, as we will see that, in some cases, a linear function is completely specified if its eigenvectors and corresponding eigenvalues are known.

Theorem 2.122. Let V be a vector space. A linear function $\mathbf{f} : V \to V$ be a linear function is invertible the only eigenvalue with 0 as its eigenvector is the zero vector $\mathbf{0}$.

Proof. Consider the contrapositive. Some nonzero eigenvector $\mathbf{v}$ has 0 as its eigenvalue iff $\mathbf{f}$ has a nontrivial kernel, which is equivalent to $\mathbf{f}$ not being invertible. □

Remark 2.123. (Eigenvectors, eigenvalues, and zero).
According to the above definition, $\mathbf{0}$ is an eigenvector of every linear function, with its corresponding eigenvalue being 0.
Some authors explicitly disallow $\mathbf{0}$ from being an eigenvector of any linear function (and thus disallow 0 from being an eigenvalue of any linear function) in their definition of "eigenvector" so that the condition "the only eigenvector with 0 as an eigenvalue is $\mathbf{0}$" is equivalent to the condition "0 is not an eigenvalue". The later is easier to say than the former.

Theorem 2.124. (Intuition for an invertibility condition).
[intuition on what a determinant is]
$\mathbf{f}$ is not invertible iff for any basis E of V, the set $\mathbf{f}(E)$ is linearly dependent
[n-dimensional volume spanned by a linearly dependent set is 0]

Derivation 2.125. (Characterization of eigenvectors and eigenvalues).
Let V and W be vector spaces over a field K, and let $\mathbf{f} : V \to W$ be linear function. In this derivation, we will discover a characterization of the eigenvectors and eigenvalues of $\mathbf{f}$.
Suppose $\mathbf{v} \in V$ is an eigenvector of $\mathbf{f}$. Then

$$\begin{aligned}
&\exists c \in K \text{ s.t. } \mathbf{f}(\mathbf{v}) = c\mathbf{v} \\
\Longleftrightarrow\ &\exists c \in K \text{ s.t. } \mathbf{f}(\mathbf{v}) - c\mathbf{v} = \mathbf{0} \\
\Longleftrightarrow\ &\exists c \in K \text{ s.t. } \mathbf{f}(\mathbf{v}) - c\mathbf{I}(\mathbf{v}) = \mathbf{0}, \text{ where } \mathbf{I} \text{ is the identity on } V \\
\Longleftrightarrow\ &\exists c \in K \text{ s.t. } (\mathbf{f} - c\mathbf{I})(\mathbf{v}) = \mathbf{0} \\
\Longleftrightarrow\ &\exists c \in K \text{ s.t. } \mathbf{v} \in \ker(\mathbf{f} - c\mathbf{I}).
\end{aligned}$$

[use $\mathbf{I}_V$?]

(In the last line of the above, $\mathbf{f} - c\mathbf{I}$ is the linear function defined by $(\mathbf{f} - c\mathbf{I})(\mathbf{v}) = \mathbf{f}(\mathbf{v}) - c\mathbf{I}(\mathbf{v})$.)

In all, we see that $\mathbf{v}$ is an eigenvector of $\mathbf{f}$ iff $\exists c \in K$ s.t. $\mathbf{v} \in \ker(\mathbf{f} - c\mathbf{I})$. Thus the set of eigenvectors of $\mathbf{f}$ is equal to $\bigcup_{c \in K} \ker(\mathbf{f} - c\mathbf{I})$.

An argument analogous to the one above shows that c is a nonzero eigenvalue of $\mathbf{f}$ iff $\exists \mathbf{v} \neq \mathbf{0}$ $\mathbf{v} \in \ker(\mathbf{f} - c\mathbf{I})$, i.e., iff $\ker(\mathbf{f} - c\mathbf{I}) \neq \{\mathbf{0}\}$. Since $\ker(\mathbf{f} - c\mathbf{I}) \neq \{\mathbf{0}\}$ iff $\det(\mathbf{f} - c\mathbf{I}) = 0$ (see the previous theorem), we have in all that c is a nonzero eigenvalue of $\mathbf{f}$ iff $\det(\mathbf{f} - c\mathbf{I}) = 0$.

So:

$$\text{eigenvectors of } \mathbf{f} = \bigcup_{c \in K} \ker(\mathbf{f} - c\mathbf{I})$$
$$\text{nonzero eigenvalues of } \mathbf{f} = \{c \in K \mid \det(\mathbf{f} - c\mathbf{I}) = 0\}$$

We can get even more specific with our characterization of the eigenvectors of $\mathbf{f}$. Notice that, in the above union, we need not include kernels that are just $\{\mathbf{0}\}$, since every kernel contains $\mathbf{0}$. An equivalent union consists of nontrivial kernels, i.e., of kernels of noninvertible linear functions, i.e., of kernels of linear functions with determinant 0. This gives us the following:

$$\text{eigenvectors of } \mathbf{f} = \bigcup_{\substack{c \in K \\ c \,:\, \det(\mathbf{f} - c\mathbf{I}) = 0}} \ker(\mathbf{f} - c\mathbf{I}) = \bigcup_{\substack{c \in K \\ c \,:\, c \text{ is a nonzero eigenvalue of } \mathbf{f}}} \ker(\mathbf{f} - c\mathbf{I})$$

$$\text{nonzero eigenvalues of } \mathbf{f} = \{c \in K \mid \det(\mathbf{f} - c\mathbf{I}) = 0\}$$

2.6 The dot product

Most explanations of the dot product engage in at least one of two pedagogically problematic approaches.

The most common of these approaches is: first, define the dot product as $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i$; second, show 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 $\mathbf{v}_1 \cdot \mathbf{v}_2 = \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i$, rather than starting with $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i$. Secondly, the law of cosines gives no intuition[15]. There is a much better way to prove that $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i$ implies $\mathbf{v}_1 \cdot \mathbf{v}_2 = ||\mathbf{v}_1|| \, ||\mathbf{v}_2|| \cos(\theta)$, which we present.

The second problematic approach arises when an author *does* decide to start with the geometrically intuitive definition, $\mathbf{v}_1 \cdot \mathbf{v}_2 := ||\mathbf{v}_1|| \, ||\mathbf{v}_2|| \cos(\theta)$. Authors will prove that $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i$ follows as a result of the law of cosines. Again, using the law of cosines gives no intuition. Instead, $\mathbf{v}_1 \cdot \mathbf{v}_2 := \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i$ 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 last section of Chapter 5 describes these problems and presents a satisfying explanation of the cross product. (The cross product is addressed in Chapter 5 because understanding cross products requires understanding determinants).

Magnitude and angle in $\mathbb{R}^n$

Definition 2.126. (Length of a vector in $\mathbb{R}^n$).

Let $\hat{\mathfrak{E}}$ 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 $||\mathbf{v}|| := \sqrt{\sum_{i=1}^n ([\mathbf{v}]_{\hat{\mathfrak{E}}})_i^2}$.

Definition 2.127. (Unit vector hat notation).

For $\mathbf{v} \in \mathbb{R}^n$, we define the notation $\hat{\mathbf{v}} := \frac{\mathbf{v}}{||\mathbf{v}||}$. We have $||\hat{\mathbf{v}}|| = 1$ for all $\mathbf{v} \in \mathbb{R}^n$.

Definition 2.128. (Unsigned angle between vectors in $\mathbb{R}^n$).

Let vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ have the same length $r = ||\mathbf{v}_1|| = ||\mathbf{v}_2||$, and consider the n-sphere that of radius r results when the initial points of $\mathbf{v}_1, \mathbf{v}_2$ coincide. (The initial points of $\mathbf{v}_1, \mathbf{v}_2$ are the center of the sphere). Let s be the length of the shortest path on the n-sphere from $\mathbf{v}_1$ to $\mathbf{v}_2$. We define the *unsigned angle θ between $\mathbf{v}_1$ and $\mathbf{v}_2$* to be the ratio $\theta := \frac{s}{r}$. When two vectors don't have the same length, we define the unsigned angle between them to be the unsigned angle between two vectors that point in the same directions but that have the same length.

Note, we have used the descriptor "unsigned" because, since $s, r \geq 0$, we know $\theta \geq 0$.

Definition 2.129. (Perpendicular vectors in $\mathbb{R}^n$). We say that two vectors in $\mathbb{R}^n$ are *perpendicular*, or *orthogonal*, iff the unsigned angle between them is $\frac{\pi}{2}$.

Moreso than prematurely introducing the dot product, the purpose of the following theorem is to engage with the intuitive definition of angle as $\theta := \frac{s}{r}$, which is so often abandoned in higher mathematics.

Theorem 2.130. (Unsigned angle formula in $\mathbb{R}^n$).

The unsigned angle θ between vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ with the same length $r = ||\mathbf{v}_1|| = ||\mathbf{v}_2||$ is

$$\theta = \arccos\Big(\sum_{i=1}^n [\hat{\mathbf{v}}_1]_{\hat{\mathfrak{E}}})_i ([\hat{\mathbf{v}}_2]_{\hat{\mathfrak{E}}})_i\Big).$$

Proof. We prove the theorem for the case $n = 2$. For the general case, one would make use of the following description of the n-sphere of radius r, which uses generalized spherical coordinates[16]:

[15] When used as a starting point, the law of cosines is unintuitive because the validity of the law of cosines is established via Euclidean geometry, which is unintuitive (to me, at least).

[16] See `https://en.wikipedia.org/wiki/N-sphere#spherical%20coordinates` for more.

$$\mathbf{x}\begin{pmatrix} t_1 \\ ... \\ t_{n-1} \end{pmatrix} = r\sum_{i=1}^{n}\Big(\sin(t_1)...\sin(t_{i-1})\cos(t_i)\hat{\mathfrak{e}}_i\Big) = r\begin{pmatrix} \cos(t_1) \\ \sin(t_1)\cos(t_2) \\ \sin(t_1)\cos(t_2)\cos(t_3) \\ \vdots \\ \sin(t_1)...\sin(t_{i-1})\cos(t_i) \\ \vdots \\ \sin(t_1)...\sin(t_{n-1})\cos(t_n) \end{pmatrix},$$

where $t_1, ..., t_{n-2} \in [0, \pi]$, $t_{n-1} \in [0, 2\pi)$.
Since we are considering the case $n = 2$, we have

$$\mathbf{x}(t) = r\begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}.$$

We compute θ as

$$\theta = \frac{s}{r} = \frac{1}{r}\int_{t_1}^{t_2}\Big|\Big|\frac{d\mathbf{x}(t)}{dt}\Big|\Big|dt = \frac{1}{r}\int_{t_1}^{t_2} r\ dt = \int_{t_1}^{t_2} dt = t_2 - t_1.$$

Since t_i, $i \in \{1, 2\}$ are such that $\mathbf{x}(t_i) = \mathbf{v}_i$, we have $t_i = \arccos(\frac{([\mathbf{v}_i]_{\hat{\mathfrak{E}}})_1}{r})$. So

$$\theta = t_2 - t_1 = \arccos\Big(\frac{([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1}{r}\Big) - \arccos\Big(\frac{[\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1}{r}\Big).$$

The cosine angle addition identity $\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$ implies[17] the identity $\arccos(\alpha) + \arccos(\beta) = \arccos(\alpha\beta - \sqrt{1-\alpha^2}\sqrt{1-\beta^2})$. Using this identity with the fact that $([\mathbf{v}_i]_{\hat{\mathfrak{E}}})_1^2 + ([\mathbf{v}_i]_{\hat{\mathfrak{E}}})_2^2 = 1$, $i \in \{1, 2\}$, the above becomes

$$\begin{aligned}
\theta &= \arccos\Bigg(\frac{([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1)([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1)}{r^2} + \sqrt{1 - \Big(\frac{([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1}{r}\Big)^2}\sqrt{1 - \Big(\frac{([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1}{r}\Big)^2}\Bigg) \\
&= \arccos\Bigg(\frac{([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1)([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1)}{r^2} + \sqrt{\frac{r^2 - ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1^2}{r^2}}\sqrt{\frac{r^2 - ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1^2}{r^2}}\Bigg) \\
&= \arccos\Bigg(\frac{([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1)([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1)}{r^2} + \sqrt{\frac{([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_2^2}{r^2}}\sqrt{\frac{([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_2^2}{r^2}}\Bigg) \\
&= \arccos\Big(\frac{([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_2)([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1) + ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_2([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_2}{r^2}\Big) \\
&= \arccos\Big(\frac{([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1}{r}\frac{([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1}{r} + \frac{([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_2}{r}\frac{([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_2}{r}\Big) \\
&= \arccos\Big(([\hat{\mathbf{v}}_1]_{\hat{\mathfrak{E}}})_1([\hat{\mathbf{v}}_2]_{\hat{\mathfrak{E}}})_1 + ([\hat{\mathbf{v}}_2]_{\hat{\mathfrak{E}}})_1([\hat{\mathbf{v}}_2]_{\hat{\mathfrak{E}}})_2\Big) \\
&= \arccos\Big(\sum_{i=1}^{2}[\hat{\mathbf{v}}_1]_{\hat{\mathfrak{E}}})_i([\hat{\mathbf{v}}_2]_{\hat{\mathfrak{E}}})_i\Big).
\end{aligned}$$

The second to last step follows because $r = ||\mathbf{v}_1|| = ||\mathbf{v}_2||$. □

[17]Solve the first identity for $\gamma = \cos(\alpha)$ and $\delta = \cos(\beta)$ to obtain the second identity in terms of γ and δ.

The dot product

Now that we have a notion of perpendicularity between vectors in $\mathbb{R}^n$, we can define the notion of vector projection, which is the main idea involved in the dot product.

Definition 2.131. (Vector projection in $\mathbb{R}^n$).
Consider vectors $\mathbf{v}_1, \mathbf{v}_2, (\mathbf{v}_2)_\perp \in \mathbb{R}^n$, where $(\mathbf{v}_2)_\perp$ is perpendicular to $\mathbf{v}_2$.
The *vector projection* of $\mathbf{v}_1$ onto $\mathbf{v}_2$ is the unique vector $\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2) := (v_1)_{||}\hat{\mathbf{v}}_2$ such that $\mathbf{v}_1 = (v_1)_{||}\hat{\mathbf{v}}_2 + (v_1)_\perp(\hat{\mathbf{v}}_2)_\perp$, where $(v_1)_{||}, (v_1)_\perp \in K$.

Remark 2.132. Note that, for $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, we have $\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2) = \text{proj}(\mathbf{v}_1 \to \hat{\mathbf{v}}_2)$ because $\hat{\hat{\mathbf{v}}}_2 = \hat{\mathbf{v}}_2$.

Definition 2.133. (Dot product).
The *dot product on* $\mathbb{R}^n$ is the function $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

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

Why do we care about the dot product? The primary reason is that $\mathbf{v}_1 \cdot \mathbf{v}_2 = \text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$ when $||\mathbf{v}_2|| = 1$, so investigating the dot product can tell us more about vector projections[18]. Another reason is that in physics, the *work* done by a force $\mathbf{F}$ along a displacement $\Delta\mathbf{x}$ is defined to be the product of the magnitude of the force that "aligns" with the displacement with the magnitude of the displacement[19]; i.e., work is defined to be $\mathbf{F} \cdot \Delta\mathbf{x}$.

Before moving on to more important business, we quickly note the following fact relating the dot product and the length of a vector.

Theorem 2.134. (Length in $\mathbb{R}^n$ in terms of dot product).
For any $\mathbf{v} \in \mathbb{R}^n$, we have $||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

Proof. $\mathbf{v} \cdot \mathbf{v} = ||\text{proj}(\mathbf{v} \to \mathbf{v})||\ ||\mathbf{v}|| = ||\mathbf{v}||^2$, so $\sqrt{\mathbf{v} \cdot \mathbf{v}} = ||\mathbf{v}||$. □

Now, we set out to show that the dot product is a bilinear function. We do so by appealing to the fact that the dot product involves vector projection.

Lemma 2.135. (Projection onto a vector is a linear function).
Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$. The map $\mathbf{v}_1 \mapsto \text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$ is linear.

Proof. Define $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}$ by $\mathbf{f}(\mathbf{v}_1) = \text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$. We show $\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2)$ and $\mathbf{f}(c\mathbf{v}_1) = c\mathbf{f}(\mathbf{v}_1)$.

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

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

□

Theorem 2.136. (The dot product is a bilinear function).
The dot product 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 dot product is symmetric, so it suffices to show that it is a linear function in either argument; it suffices to show that $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}$ defined by $\mathbf{f}(\mathbf{v}_1) = \mathbf{v}_1 \cdot \mathbf{v}_2$ is a linear function. Well, $\mathbf{f}(\mathbf{v}_1) = ||\mathbf{v}_2||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$, and we know that $\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$ is linear in $\mathbf{v}_2$. Since $\mathbf{f}$ is the result of scaling a linear function by $||\mathbf{v}_2||$, it too is a linear function. □

[18] We will indeed find that the dot product tells us something about projections in Theorem 2.139!
[19] A negative sign occurs when the force opposes the displacement.

Knowing that the dot product is bilinear allows us to derive an alternate "algebraic" (as opposed to "geometric") expression for the dot product of two vectors in $\mathbb{R}^n$. Before we derive this alternate expression, though, we will need to note the following basic fact.

Lemma 2.137. Let $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ be the standard basis for $\mathbb{R}^n$. We have

$$\hat{\mathfrak{e}}_i \cdot \hat{\mathfrak{e}}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} = \delta_{ij}.$$

Proof. The lemma is equivalent to the statement $||\text{proj}(\hat{\mathfrak{e}}_i \to \hat{\mathfrak{e}}_j)|| = \delta_{ij}$, which is easily proved. $\square$

Derivation 2.138. (Algebraic dot product on $\mathbb{R}^n$).

We can now derive an "algebraic" formula for the dot product, using its bilinearity (Theorem 2.136) together with the previous lemma.

We will make use of the following general fact about bilinear functions. 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 $B : V \times V \to K$ satisfies

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

In the above, we've written $\mathbf{v}_1$ and $\mathbf{v}_2$ as sums of basis vectors and then used the linearity of B in each argument.

The dot product is a bilinear function $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, so the above fact can be applied to the dot product. We know from the previous lemma that $B(\hat{\mathfrak{e}}_i, \hat{\mathfrak{e}}_j) = \delta_{ij}$, so we have

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

Therefore

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

We could have proved the following theorem immediately after defining $\mathbf{v}_1 \cdot \mathbf{v}_2 := ||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)||\ ||\mathbf{v}_2||$. The usefulness of this next theorem, however, wouldn't have been apparent until knowing that $\mathbf{v}_1 \cdot \mathbf{v}_2 = \sum_{i=1}^n ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i$.

Theorem 2.139. (Vector projection in terms of dot product).

If $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, then

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

If one didn't know the algebraic formula for the dot product, this theorem would be a bit of a tautology. However, knowing the algebraic dot product allows for easy computation of the expressions $\mathbf{v}_1 \cdot \mathbf{v}_2$ and $\mathbf{v}_2 \cdot \mathbf{v}_2$ whenever we know the coordinates of $\mathbf{v}_1$ and $\mathbf{v}_2$ relative to the standard basis[20] for $\mathbb{R}^n$.

Proof. Since $\mathbf{v}_1 \cdot \mathbf{v}_2 = ||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)||\ ||\mathbf{v}_2||$, then $||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_2||}$. We have $\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2) = ||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| \widehat{\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)} = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_2||} \hat{\mathbf{v}}_2$. $\square$

[20] Actually, we only need to know the coordinates of $\mathbf{v}_1$ and $\mathbf{v}_2$ relative to some *orthonormal* basis! Orthonormal bases, also known as *self-dual bases*, are discussed in Section **??**.

Theorem 2.140. (Unsigned angle between vectors and dot product).

Before we discovered the algebraic dot product, Theorem 2.130 showed

$$\theta = \arccos\Big(\sum_{i=1}^{n}([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i\Big),$$

as a consequence of the definition $\theta := \frac{s}{r}$. Now that we know about the algebraic dot product, we can say that the signed angle θ between $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ is

$$\theta = \arccos(\hat{\mathbf{v}}_1 \cdot \hat{\mathbf{v}}_2)$$

Notably, since $\hat{\mathbf{v}}_i = \frac{\mathbf{v}_i}{||\mathbf{v}_i||}$, the above implies that

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

Remark 2.141. (Dot product as a function of unsigned angle in $\mathbb{R}^2$).

One can easily show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = ||\mathbf{v}_1||\ ||\mathbf{v}_2|| \cos(\theta)$ in $\mathbb{R}^2$ by using trigonometry. Trigonometry shows $||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = ||\mathbf{v}_1|| \cos(\theta)$, so we have $\mathbf{v}_1 \cdot \mathbf{v}_2 = ||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)||\ ||\mathbf{v}_2|| = ||\mathbf{v}_1|| \cos(\theta)||\mathbf{v}_2|| = ||\mathbf{v}_1||\ ||\mathbf{v}_2|| \cos(\theta)$.

Now that we have used the geometric definition of the dot product to derive an "algebraic" formula for the dot product, we will show that one can start with the "algebraic" formula as a definition and discover the geometric formula. We need the following definition and two lemmas before we can do this.

Definition 2.142. (Orthogonal linear function on $\mathbb{R}^n$).

We say that a linear function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is *orthogonal* iff it preserves length, i.e., iff $||\mathbf{v}|| = ||\mathbf{f}(\mathbf{v})||$ for all $\mathbf{v} \in \mathbb{R}^n$.

("Orthogonal linear function" is a somewhat misleading name. It's true that certain facts about orthogonal linear functions involve orthogonal *vectors*, but orthogonal linear functions aren't "perpendicular" to other orthogonal linear functions in some sense).

Lemma 2.143. (Orthogonal linear functions on $\mathbb{R}^n$ preserve algebraic dot product).

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

Proof. By definition, orthogonal linear functions on $\mathbb{R}^n$ preserve length. Thus, if we showed that $\mathbf{v}_1 \cdot \mathbf{v}_2$ depends on $||\mathbf{v}_1||$ and $||\mathbf{v}_2||$ (on lengths), it would follow that the algebraic dot product is preserved by orthogonal linear functions. This is what we will do.

We need to show that $\mathbf{v}_1 \cdot \mathbf{v}_2$ depends on $||\mathbf{v}_1||$ and $||\mathbf{v}_2||$. Note that $||\mathbf{v}||^2 = \mathbf{v} \cdot \mathbf{v}$ for $\mathbf{v} \in \mathbb{R}^n$. 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 see that $\mathbf{v}_1 \cdot \mathbf{v}_2$ does indeed depend on $||\mathbf{v}_1||$ and $||\mathbf{v}_2||$:

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

□

Lemma 2.144. (Rotated projection is projection of rotated vectors).

Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$. If $\mathbf{f}$ is a rotation[21], then $||\mathbf{f}(\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = ||\text{proj}(\mathbf{f}(\mathbf{v}_1) \to \mathbf{f}(\mathbf{v}_2))||$.

Proof. We have $\mathbf{v}_1 = (v_1)_{||}\hat{\mathbf{v}}_1 + (v_1)_\perp(\hat{\mathbf{v}}_1)_\perp$, so $\mathbf{f}(\mathbf{v}_1) = (v_1)_{||}\mathbf{f}(\hat{\mathbf{v}}_2) + (v_1)_\perp\mathbf{f}((\hat{\mathbf{v}}_2)_\perp)$. 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 there exists a rotation that sends $\mathbf{v}_2$ to $(\mathbf{v}_2)_\perp$, because rotations commute with each other, and because rotations are length-preserving. □

Theorem 2.145. (Algebraic dot product formula implies geometric dot product formula).

We've used the bilinearity of the geometrically defined dot product to derive the algebraic dot product formula; we've showed that defining $\cdot : \mathbb{R}^n \to \mathbb{R}^n$ by $\mathbf{v}_1 \cdot \mathbf{v}_2 := ||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)||\ ||\mathbf{v}_2||$ implies $\mathbf{v}_1 \cdot \mathbf{v}_2 = \sum_{i=1}^n([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i$.

We now show that we can do things the other way around: we can derive the geometric dot product formula from the algebraic dot product formula. More specifically, we will show that defining $\cdot : \mathbb{R}^n \to \mathbb{R}^n$ by $\mathbf{v}_1 \cdot \mathbf{v}_2 = \sum_{i=1}^n([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i$ implies that $\mathbf{v}_1 \cdot \mathbf{v}_2 = ||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)||\ ||\mathbf{v}_2||$.

[21] See Definition 5.47 in Chapter 5 to see what it means for a function to be an *n-rotation*. All we need to know for the purposes of the dot product is that all n-rotations are length-preserving.

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

Lemma 2.143 says that orthogonal linear functions on $\mathbb{R}^n$ preserve the algebraic dot product, so

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{f}(\mathbf{v}_1) \cdot \mathbf{f}(\mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) \cdot ||\mathbf{v}_2||\hat{\mathfrak{e}}_1 = \begin{pmatrix} ([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathfrak{E}}})_1 \\ \vdots \\ ([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathfrak{E}}})_n \end{pmatrix} \cdot \begin{pmatrix} ||\mathbf{v}_2|| \\ 0 \\ \vdots \\ 0 \end{pmatrix} = ([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathfrak{E}}})_1 ||\mathbf{v}_2||.$$

We have

$$\begin{aligned} ([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathfrak{E}}})_1 &= \operatorname{proj}(\mathbf{f}(\mathbf{v}_1) \to \hat{\mathfrak{e}}_1) = \operatorname{proj}(\mathbf{f}(\mathbf{v}_1) \to \mathbf{f}(\hat{\mathbf{v}}_2)) = \operatorname{proj}(\mathbf{f}(\mathbf{v}_1) \to \widehat{\mathbf{f}(\mathbf{v}_2)}) \\ &= \operatorname{proj}(\mathbf{f}(\mathbf{v}_1) \to \mathbf{f}(\mathbf{v}_2)) = \operatorname{proj}(\mathbf{v}_1 \to \mathbf{v}_2), \end{aligned}$$

where the last equality is by Lemma 2.144.

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

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 the next theorem.

Theorem 2.146. (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 $\mathbf{v}_1 - \mathbf{v}_2$.

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.

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 2.147. 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 $||\mathbf{v}_i||^2 = \sum_{j=1}^2 (([\mathbf{v}_i]_{\hat{\mathfrak{E}}})_j)^2$ to discover that we must have $([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1 + ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_2([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_2 = 0$.

Part I

Multilinear algebra and tensors

3

A motivated introduction to tensors

In this chapter, we introduce the idea of a *tensor*, since tensors underpin differential forms. 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". Thus, we can say "linear functions respect the decomposition of linear elements". After defining the notion of *multilinear function*, we will define the notion of "multilinear elements". Multilinear elements will be objects that (in a sense) are respected by multilinear functions. Formally, mutlilinear elements will be elements of *tensor product spaces*.

There are two main contributions 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 identified with 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 second key idea in the theory of tensors is to think of linear functions as vectors- that is, as elements of vector spaces. We achieve this 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 discover *dual spaces*. Soon after, we utilize the two key ideas, (1) tensor product spaces and (2) dual spaces, to prove the theorem which underlies the definition of a (p, q) *tensor*.

3.1 Multilinear functions and tensor product spaces

Definition 3.1. (Multilinear function).

Let $V_1, ..., V_k, W$ be vector spaces over a field K. We say a function $\mathbf{f} : V_1 \times ... \times V_k \to W$ is *k-linear* iff for all $\mathbf{v}_1 \in V_1, ..., \mathbf{v}_i \in V_i, ..., \mathbf{v}_k \in V_k$, the function $\mathbf{f}_i : V_i \to W$ defined by $\mathbf{f}_i(\mathbf{v}_1, ..., \mathbf{v}_i, ..., \mathbf{v}_k) = \mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_i, ..., \mathbf{v}_k)$ 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*. Note that a 1-linear function is simply a linear function. 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 $V_1, ..., V_k, W$ are vector spaces over a field K, then we use $\mathcal{L}(V_1 \times ... \times V_k \to W)$ to denote the vector space over K formed by the set of k-linear functions $V_1 \times ... \times V_k \to W$ under the operations of function addition and function scaling. In particular, the case $k = 1$ implies that $\mathcal{L}(V \to W)$ denotes the set of linear functions $V \to W$.

Proof. The proof that $\mathcal{L}(V_1 \times ... \times V_k \to W)$ 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 **??**). 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 $V_1, ..., V_k$ be finite-dimensional vector spaces over a field K. The *tensor product space* $V_1 \otimes ... \otimes V_k$ is defined to be the vector space over K whose elements are from the set $V_1 \times ... \times V_k$, where the elements are also subject to an equivalence relation $=$, which will be specified soon. We also denote a typical element of $V_1 \otimes ... \otimes V_k$ as $\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k$, rather than as $(\mathbf{v}_1, ..., \mathbf{v}_k)$.

The equivalence relation is defined by the condition that, for all $\mathbf{v}_1, ..., \mathbf{v}_k \in V$ and $c \in K$, we have:

$$\begin{gathered} \mathbf{v}_1 \otimes ... \otimes \mathbf{v}_{i-1} \otimes \mathbf{v}_{i1} \otimes \mathbf{v}_{i+1} ... \otimes \mathbf{v}_k \\ + \\ \mathbf{v}_1 \otimes ... \otimes \mathbf{v}_{i-1} \otimes \mathbf{v}_{i2} \otimes \mathbf{v}_{i+1} ... \otimes \mathbf{v}_k \\ = \\ \mathbf{v}_1 \otimes ... \otimes \mathbf{v}_{i-1} \otimes (\mathbf{v}_{i1} + \mathbf{v}_{i2}) \otimes \mathbf{v}_{i+1} ... \otimes \mathbf{v}_k \end{gathered}$$

and

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

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 "(p, q) tensor". (We have not defined (p, q) tensors yet, but we will in Definition 3.37). We will use the word "tensor" to either mean an element of a tensor product space or a (p, q) tensor, but we only do this when the meaning is clear from context.

Definition 3.6. (Elementary tensor).

Let $V_1, ..., V_k$ be vector spaces, and consider the tensor product space $V_1 \otimes ... \otimes V_k$. An element of $V_1 \otimes ... \otimes V_k$ that is of the form $\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k$ 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 $V_1 \otimes ... \otimes V_k$ 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 2.73 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 2.77).

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

$$\{\mathbf{e}_{1i_1} \otimes ... \otimes \mathbf{e}_{ki_k} \mid i_k \in \{1, ..., n_k\}\}.$$

Proof. It suffices to show that if V and W are finite-dimensional vector spaces with bases $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_m\}$, then

$$\{\mathbf{e}_i \otimes \mathbf{f}_j \mid i \in \{1, ..., n\}, j \in \{1, ..., m\}\}$$

is a basis of $V \otimes W$.

To show that this set spans $V \otimes W$, it suffices to show it spans the set of elementary tensors in $V \otimes W$, since any tensor in $V \otimes W$ is a linear combination of elementary tensors. For an elementary tensor $\mathbf{v} \otimes \mathbf{w} \in V \otimes W$, we have

$$\mathbf{v} \otimes \mathbf{w} = \sum_{i=1}^{n} ([\mathbf{v}]_E)^i ([\mathbf{w}]_F)^j \mathbf{e}_i \otimes \mathbf{f}_j,$$

by the seeming-bilinearity of $\otimes$.

To show linear independence, assume that $\sum_{i,j} T_{i,j} \mathbf{e}_i \otimes \mathbf{f}_j$ is the zero tensor. We must show that all of the $T_{i,j}$'s are 0. Note that a tensor in $V \otimes W$ is the zero tensor iff[1] it is an elementary tensor of the form $\mathbf{v} \otimes \mathbf{0}$ for $\mathbf{v} \in V$, or $\mathbf{0} \otimes \mathbf{w}$ for $\mathbf{w} \in W$. Due to the linear independence of the bases E and F, it is impossible to obtain a $\mathbf{0}$ in either position unless all $T_{i,j}$ are 0. ☐

The following 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.

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

Let V_1, V_2, W be vector spaces, and let $\mathbf{f} : V_1 \times V_2 \to W$ be a bilinear function. Then there exists a linear function $\mathbf{h_f} : V_1 \otimes V_2 \to W$ with $\mathbf{f} = \mathbf{h_f} \circ \mathbf{g}$ that uniquely depends on $\mathbf{f}$, where $\mathbf{g} : V_1 \times V_2 \to V_1 \otimes V_2$ is invertible.

Proof. First we send $(\mathbf{v}_1, \mathbf{v}_2) \overset{\mathbf{g}}{\mapsto} \mathbf{v}_1 \otimes \mathbf{v}_2$ and then $\mathbf{v}_1 \otimes \mathbf{v}_2 \overset{\mathbf{h_f}}{\mapsto} \mathbf{f}(\mathbf{v}_1, \mathbf{v}_2)$, where we impose that $\mathbf{h_f}$ be linear. (Note, requiring that $\mathbf{h_f}$ is linear implies that $\mathbf{h_f}(\mathbf{T})$ is indeed defined when $\mathbf{T}$ is a nonelementary tensor, since defining how $\mathbf{h_f}$ acts on elementary tensors is enough to determine how $\mathbf{h_f}$ acts on any tensor). We have $\mathbf{f} = \mathbf{h_f} \circ \mathbf{g}$ 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 $\mathbf{f} = \mathbf{h_f} \circ \mathbf{g}$. The composition map $\circ$ is well-defined, so $\mathbf{h_f} = \mathbf{f} \circ \mathbf{g}$ is uniquely determined.

We have shown everything except for the fact that $\mathbf{g}$ is invertible. Show this by proving that $\mathbf{g}$ has a trivial kernel. ☐

Using the previous theorem to define a natural isomorphism of vector spaces yields the following more general result.

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

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

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

Proof. We prove the theorem for the case $k = 2$, and show $\mathcal{L}(V_1 \times V_2 \to W) \cong \mathcal{L}(V_1 \otimes V_2 \to W)$. 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.

---

[1]$(\implies)$. We have $\mathbf{v} \otimes \mathbf{0} = 0 \cdot (\mathbf{v} \otimes \mathbf{0}) = \mathbf{0}$. $(\impliedby)$. If $\sum_{i,j} T_{i,j} \mathbf{e}_i \otimes \mathbf{f}_j = \mathbf{0}$, then $\sum_{i,j} T_{i,j} \mathbf{e}_i \otimes \mathbf{f}_j = \left(\sum_{i,j} T_{i,j} \mathbf{e}_i \otimes \mathbf{f}_j\right) \cdot 0 = \mathbf{e}_i \otimes \mathbf{0}$ for some i.

To construct a linear isomorphism $\mathcal{L}(V_1 \times V_2 \to W) \mapsto \mathcal{L}(V_1 \otimes V_2 \to W)$, we send $\mathbf{f} \in \mathcal{L}(V_1 \times V_2 \to W) \mapsto \mathbf{h_f} = \mathbf{f} \circ \mathbf{g}^{-1}$, where $\mathbf{g}$ and $\mathbf{h_f}$ are from the proof of Theorem 3.9.

This map is linear because, given vector spaces Y, Z, W, the map $\circ$ which composes linear functions, $\circ : \mathcal{L}(Y \to Z) \times \mathcal{L}(Z \to W) \to \mathcal{L}(Y \to W)$, is a bilinear map. (Check this fact for yourself. The consequences of this are explored in Derivation 4.38). It is an injection because we know from Theorem 3.9 that $\mathbf{h_f}$ is uniquely determined by $\mathbf{f}$. Since the map is a linear injection between two vector spaces of the same dimension, it is a linear isomorphism. $\square$

Remark 3.11. (The sense in which multilinear functions respect multilinear elements).

We can use the previous theorem to formalize precisely how multilinear functions respect the decomposition of multilinear elements.

We know linear functions respect the decomposition of their input, so the decomposition of $\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k$ is respected by any linear function that acts on it. The theorem tells us that the action of a linear function on the tensor $\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k$ can be identified with the action of a multilinear function on the tuple of vectors $(\mathbf{v}_1, ..., \mathbf{v}_k)$. Thus, the equivalence class $[\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k]_=$ given by the equivalence relation $=$ used to define tensor products space in Definition 3.4 is respected by that multilinear function as well. Every multilinear function can be identified with a linear function, so this reasoning applies to all multilinear functions.

3.2 A motivated introduction to (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 $\mathcal{L}(V \to W) \cong W \otimes V^*$, where V^* is the *dual vector space* to W. 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 a dual vector 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 $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_m\}$, respectively, and consider a linear function $\mathbf{f} : V \to W$. 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 $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ be the standard basis of $K^n = K^{\dim(V)}$ and let $\hat{\mathfrak{F}} = \{\hat{\mathfrak{f}}_1, ..., \hat{\mathfrak{f}}_m\}$ be the standard basis of $K^m = K^{\dim(W)}$. So, in the example, $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ and $\hat{\mathfrak{F}} = \{\hat{\mathfrak{f}}_1, \hat{\mathfrak{f}}_2, \hat{\mathfrak{f}}_3\}$. The first "big leap" is to notice that the $m \times n$ matrix with ij entry 1 and all other entries 0 is $\hat{\mathfrak{f}}_i \hat{\mathfrak{e}}_j^\top$, where $\hat{\mathfrak{f}}_i \hat{\mathfrak{e}}_j^\top$ is the product of a $m \times 1$ matrix with a $1 \times n$ matrix (see Theorem 2.102). This means that the above 3×2 matrix can be expressed as

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

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

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

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 $\{\hat{\mathfrak{e}}_i\hat{\mathfrak{f}}_j^\top\}$ 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 $\hat{\mathfrak{E}}$ and $\hat{\mathfrak{F}}$ for K^n and K^m. Nothing is stopping us from using different bases, however. Suppose $G = \{\mathbf{g}_1, ..., \mathbf{g}_n\}$ is a basis for K^n and $H = \{\mathbf{h}_1, ..., \mathbf{h}_m\}$ 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

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

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 **??** and possibly Theorem **??**). 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 $K^n \to K$ 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 Definition 2.66).

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 $K \to K^m$ 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 a linear map and on the right hand side to denote a vector. Note, the image of the map $\mathbf{g}_i : K \to K^m$ 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 $\mathbf{g}_i \circ \phi_{\mathbf{h}_j} : K^n \to K$, where $\mathbf{g}_i$ again denotes the linear map $K \to K^m$ defined by $\mathbf{g}_i(c) = c\mathbf{g}_i$. Note that $\mathbf{g}_i \circ \phi_{\mathbf{h}_j}$ is also a rank-1 linear map.

Overall, we have shown that the matrix with respect to bases of a linear function $\mathbf{f} : V \to W$ can be expressed as a linear combination of the (primitive (see Derivation **??**)) matrices that represent the linear maps $\mathbf{g}_i \circ \phi_{\mathbf{h}_j}$. Therefore, the linear function $\mathbf{f}_{E,F} : K^n \to K^m$ that is induced by the choice of bases E and F (see Theorem **??**) is a linear combination of the linear functions $\mathbf{g}_i \circ \phi_{\mathbf{h}_j}$:

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

where the b_{ij} are the same as in the above sum. So, $\mathbf{f} : V \to W$ is very similar:

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

where $\mathbf{g}_i' : K \to W$ and $\phi_{\mathbf{h}_j}' : V \to K$ are such that $(\mathbf{g}_i')_{E,F} = \mathbf{g}_i$ and $(\phi_{\mathbf{h}_j}')_{E,F} = \phi_{\mathbf{h}_j}$. Note that $\mathbf{g}_i' \circ \phi_{\mathbf{h}_j}' : V \to W$.

At the beginning of this discussion, we chose bases E and F for V and W. At this point, we have seen enough motivation to justify getting rid of bases[2]. The following theorem is a basis-independent reinterpretation of the above statement.

Theorem 3.12. Any linear function $V \to W$ is a linear combination of rank-1 linear functions $V \to W$.

To recover the particular format of the above basis-dependent result, we use the previous theorem in conjunction with the following theorem.

Theorem 3.13. Let V and W be vector spaces. Any rank-1 linear function $V \to W$ can be expressed as $\mathbf{w} \circ \phi$, for some $\mathbf{w} \in W$ and some linear function $\phi : V \to K$, where $\mathbf{w} : K \to W$ is the linear map defined by $\mathbf{w}(c) = c\mathbf{w}$.

Proof. Let $\mathbf{f}$ be a rank-1 linear function $V \to W$. 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 d_{12} is the unique scalar for which $\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) = d_{12}\mathbf{w}$. Consider the condition "$\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) = d_{12}\mathbf{w}$" for a moment. Since $\mathbf{f}$ is linear, the condition becomes $\mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2) = d_{12}\mathbf{w}$, 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 implies $d_1 + d_2 = d_{12}$.

Thus, we have $\phi(\mathbf{v}_1 + \mathbf{v}_2) = d_{12}$, where $d_{12} = d_1 + d_2$; we have $\phi(\mathbf{v}_1 + \mathbf{v}_2) = d_{12} = d_1 + d_2 = \phi(\mathbf{v}_1) + \phi(\mathbf{v}_2)$. A similar argument shows that $\phi(c\mathbf{v}) = c\phi(\mathbf{v})$, so ϕ is linear. □

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

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

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

This is the basis-independent generalization of the above equation we were after.

Since we have seen that linear functions $V \to K$ play a fundamental role in this decomposition, we make the following definition.

[2] We can get rid of bases to some extent by avoiding coordinates relative to bases. However, we can't truly escape the notion of bases because we need to speak of rank-1 linear functions; this implies a notion of dimension, and thus of bases.

Definition 3.14. (Dual vector 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 $V \to K$ under the operations of function addition and function scaling:

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

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

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

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

where $\mathbf{w}_i \in \mathcal{L}(K \to W)$ 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 $\mathbf{F} : \mathcal{L}(V \to W) \to W \otimes V^*$ that sends the rank-1 element $(\mathbf{w}_i \circ \phi_j) \in \mathcal{L}(V \to W)$ to the elementary tensor $\mathbf{w}_i \otimes \phi_j \in W \otimes V^*$:

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

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 $(\mathbf{w} \circ \phi) \in \mathcal{L}(V \to W)$. (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

$$\mathbf{F}(\mathbf{f}_1 + \mathbf{f}_2) = \mathbf{F}(\mathbf{f}_1) + \mathbf{F}(\mathbf{f}_2)$$
$$\mathbf{F}(c\mathbf{f}) = c\mathbf{F}(\mathbf{f}),$$

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

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

$$\mathbf{w}_i \circ \phi_k + \mathbf{w}_j \circ \phi_k \overset{\mathbf{F}}{\longmapsto} \mathbf{w}_i \otimes \phi_k + \mathbf{w}_j \otimes \phi_k$$
$$\mathbf{w}_i \circ \phi_j + \mathbf{w}_i \circ \phi_k \overset{\mathbf{F}}{\longmapsto} \mathbf{w}_i \otimes \phi_j + \mathbf{w}_i \otimes \phi_k$$
$$c(\mathbf{w}_i \circ \phi_j) \overset{\mathbf{F}}{\longmapsto} c(\mathbf{w}_i \otimes \phi_j),$$

where $\mathbf{w}_i \in \mathcal{L}(K \to W)$ 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-bilinearity[3] of $\otimes$:

$$\mathbf{w}_i \circ \phi_k + \mathbf{w}_j \circ \phi_k = (\mathbf{w}_i + \mathbf{w}_j) \circ \phi_k \overset{\mathbf{F}}{\longmapsto} (\mathbf{w}_i + \mathbf{w}_j) \otimes \phi_k = \mathbf{w}_i \otimes \phi_k + \mathbf{w}_j \otimes \phi_k$$
$$\mathbf{w}_i \circ \phi_j + \mathbf{w}_i \circ \phi_k = \mathbf{w}_i \circ (\phi_j + \phi_k) \overset{\mathbf{F}}{\longmapsto} \mathbf{w}_i \otimes \phi_j + \mathbf{w}_i \otimes \phi_k = \mathbf{w}_i \otimes (\phi_j + \phi_k)$$
$$c(\mathbf{w}_i \circ \phi_j) = (c\mathbf{w}_i) \circ \phi_j \overset{\mathbf{F}}{\longmapsto} (c\mathbf{w}_i) \otimes \phi_j = c(\mathbf{w}_i \otimes \phi_j).$$

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

The bijectivity of $\mathbf{F}$ now follows easily. $\mathbf{F}$ is surjective because any nonelementary tensor coorresponds to a "nonelementary composition", i.e., a linaer combination of elementary compositions. $\mathbf{F}$ is injective because it

[3] The 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, $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$)

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 3.15. ($\mathcal{L}(V \to W) \cong W \otimes V^*$ naturally).

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

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

This isomorphism is natural because it does not depend on a choice of basis. (See Definition 2.77).

Remark 3.16. (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 compositions", 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 3.17. (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 3.18. (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 $\mathbf{f} : V \to W$ 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 $\phi \in V^* = \mathcal{L}(V \to K)$ in the "elementary composition" $\mathbf{w} \circ \phi$.

In order 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 (p, q) tensor because they allow us to think of linear functions $V \to V$ as vectors. As was previously mentioned, every linear function $V \to V$ is a linear combination of elements of V^*.

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

Definition 3.19. (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 $V \to K$ under the operations of function addition and function scaling:

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

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

Definition 3.20. (Covariance and contravariance).

Let V be a vector space. For reasons that will be explained later, in Remark 4.34, 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, covaraint vectors themselves are indexed by upper subscripts. So, for example, we would write a linear combination of covariant vectors as $c_1\phi^1 + ... + c_n\phi^n$.

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 $c^1\mathbf{v}_1 + ... + c^n\mathbf{v}_n$.

The deeper meaning behind covariance and contravariance will be explained in Remark 4.34.

Bases for dual spaces

Derivation 3.21. (Induced dual basis).

Let V be a finite-dimensional vector space and let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be a basis for V. We can discover a basis for V^* by using the basis E for V to decompose an arbitrary $\phi \in V^*$ into a basis sum.

To achieve this decomposition, we utilize the correspondence between linear transformations and matrices. Any element $\phi \in V^* = \mathcal{L}(V \to K)$ is represented relative to the basis E by the (primitive) $1 \times n$ matrix $\phi(E)$ (see Derivation **??** and Remark **??**).

Recall from Definition 2.84 that the matrix $\phi(E)$ is

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

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

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

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

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

Now, we reinterpret each $\hat{\mathfrak{e}}_i^\top \mathbf{v}$ in the sum as the action of some linear linear function on $\mathbf{v}$. Specifically, we define $\phi_{\mathbf{e}_i}$ to be the element of V^* defined by $\phi_{\mathbf{e}_i}(\mathbf{v}) = \hat{\mathfrak{e}}_i^\top \mathbf{v}$. In other words, $\phi_{\mathbf{e}_i}$ is the linear function $V \to K$ represented by the matrix $\hat{\mathfrak{e}}_i^\top$.

With this new definition, the above becomes

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

So, in all, we have

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

We see that any $\phi \in V^*$ is a linear combination of the $\phi_{\mathbf{e}_i}$, so the $\phi_{\mathbf{e}_i}$ span V^*. The $\phi_{\mathbf{e}_i}$ are also linearly independent because they are represented by the linearly independent (primitive) row-matrices $\hat{\mathfrak{e}}_i^\top$. Thus, the set $E^* = \{\phi_{\mathbf{e}_1}, ..., \phi_{\mathbf{e}_n}\}$ is a basis for V^*. Since E^* depends on E, we call E^* the *dual basis for V^* induced by E*. We will often refer to E^* simply as the *induced dual basis*.

Theorem 3.22. (Dimension of dual space to a finite-dimensional vector space).

If V is a finite-dimensional vector space, then $\dim(V^*) = \dim(V)$. Applying this fact again, we see $\dim(V^{**}) = \dim(V^*) = \dim(V)$.

Proof. In the previous derivation, we started with the assumption that V is finite-dimensional, and eventually saw that the dual basis induced by a choice of basis for V contains n elements. □

Theorem 3.23. (Characterizations of an induced dual basis).

Let V be a finite-dimensional vector space. If $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is a basis for V, then the dual basis $E^* = \{\phi_{\mathbf{e}_1}, ..., \phi_{\mathbf{e}_n}\}$ for V^* induced by E is characterized by the following equivalent conditions.

For each $i \in \{1, ..., n\}$,

1. $\phi_{\mathbf{e}_i}$ is the element of V^* that is represented relative to E by the (primitive) matrix $\hat{\mathfrak{e}}_i^\top$. That is, $\phi_{\mathbf{e}_i}(\mathbf{v}) = \hat{\mathfrak{e}}_i^\top [\mathbf{v}]_E$.
2. $\phi_{\mathbf{e}_i}(\mathbf{v}) = ([\mathbf{v}]_E)^i$.
3. $\phi_{\mathbf{e}_i}(\mathbf{e}_j) = \delta_{ij}$.

Proof. The first and second items are obviously equivalent. We show that the second condition is equivalent to the third condition; we show $(\phi_{\mathbf{e}_i}(\mathbf{v}) = ([\mathbf{v}]_E)^i \iff \phi_{\mathbf{e}_i}(\mathbf{e}_j) = \delta_{ij})$. To prove the forward direction, substitute $\mathbf{v} = \mathbf{e}_j$. As for the reverse direction, we have $\phi_{\mathbf{e}_i}(\mathbf{v}) = \sum_{j=1}^n ([\mathbf{v}]_E)^j \phi(\mathbf{e}_j) = \sum_{j=1}^n ([\mathbf{v}]_E)^j \delta_{ij} = ([\mathbf{v}]_E)^i$. □

Remark 3.24. (An "unnatural" isomorphism $V \cong V^*$).

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

This isomorphism is *not* natural (see Definition 2.77) 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 onto when V is finite-dimensional, because when V is infinite-dimensional the cardinality of V^* is strictly greater than the cardinality of V.

Remark 3.25. (We don't always have to choose induced bases).

We don't have to pick a basis of V to pick a basis for V^*. Derivation 3.21 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^*.

Corresponding elements of dual spaces induced by bases

Definition 3.26. (Corresponding elements of dual spaces induced by bases).

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be the dual basis for V^* induced by E, and let $E^{**} = \{\Phi_{\phi^{\mathbf{e}_1}}, ..., \Phi_{\phi^{\mathbf{e}_n}}\}$ be the dual basis for V^{**} induced by E^*. Additionally, let $\mathbf{F} : V \to V^*$ be the isomorphism that sends $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ and let $\mathbf{G} : V^* \to V^{**}$ be the isomorphism that sends $\phi^{\mathbf{e}_i} \mapsto \Phi_{\phi^{\mathbf{e}_i}}$. We define the notation $\phi^{\mathbf{v}} := \mathbf{F}(\mathbf{v})$ and $\Phi_\phi := \mathbf{G}(\phi)$.

Theorem 3.27. ($\Phi_{\phi^{\mathbf{v}}} = \Phi_{\mathbf{v}}$).

Assume the hypotheses of the previous definition. We have $\Phi_{\phi^{\mathbf{v}}} = \Phi_{\mathbf{v}}$, where $\Phi_{\mathbf{v}}$ is defined by $\Phi_{\mathbf{v}}(\phi) := \phi(\mathbf{v})$ (this is the definition for $\Phi_{\mathbf{v}}$ given in Theorem 3.30).

Proof. To prove the theorem, we check that $\mathbf{G}(\mathbf{F}(\mathbf{v})) = \Phi_{\mathbf{v}}$, where $\Phi_{\mathbf{v}}(\phi) := \phi(\mathbf{v})$, holds for all $\mathbf{v} \in V$. Since $\mathbf{v} = \sum_{i=1}^n ([\mathbf{v}]_E)^i \mathbf{e}_i$, we have $\mathbf{F}(\mathbf{v}) = \sum_{i=1}^n ([\mathbf{v}]_E)^i \phi^{\mathbf{e}_i}$ and $\mathbf{G}(\mathbf{F}(\mathbf{v})) = \sum_{i=1}^n ([\mathbf{v}]_E)^i \Phi_{\phi^{\mathbf{e}_i}}$. Thus, $\Big(\mathbf{G}(\mathbf{F}(\mathbf{v}))\Big)(\phi) = \sum_{i=1}^n \Big(([\mathbf{v}]_E)^i \Phi_{\phi^{\mathbf{e}_i}}(\phi)\Big)$. Applying Theorem ..., we have $([\mathbf{v}]_E)^i = \phi^{\mathbf{e}_i}(\mathbf{v})$ and $\Phi_{\phi^{\mathbf{e}_i}}(\phi) = ([\phi]_{E^*})^i$, so $\sum_{i=1}^n \Big(([\mathbf{v}]_E)^i \Phi_{\phi^{\mathbf{e}_i}}(\phi)\Big) = \sum_{i=1}^n \Big(([\phi]_{E^*})^i \phi^{\mathbf{e}_i}(\mathbf{v})\Big) = \Big(\sum_{i=1}^n ([\phi]_{E^*})^i \phi^{\mathbf{e}_i}\Big)(\mathbf{v}) = \phi(\mathbf{v}) = \Phi_{\mathbf{v}}(\phi)$. □

Theorem 3.28. ($[\mathbf{v}]_E = [\phi^{\mathbf{v}}]_{E^*}$).

Assume the hypotheses of the previous definition. We have $[\mathbf{v}]_E = [\phi^{\mathbf{v}}]_{E^*}$ for any $\mathbf{v} \in V$.

Proof. The theorem is true precisely because $\mathbf{F}(\mathbf{e}_i) = \phi^{\mathbf{e}_i}$. A more explicit check is left to the reader. $\square$

Remark 3.29. (The misleading star notation for dual vectors).

Some authors use $\mathbf{v}^*$ to denote $\phi^{\mathbf{v}}$, while also using $\{\mathbf{e}_1^*, ..., \mathbf{e}_n^*\}$ to denote an *arbitrary* basis of V^*. This notation is misleading because it is suggestive of the false equation $\mathbf{e}_i^* = \mathbf{e}_i^*$, where the first $*$ is from the definition of $\mathbf{v}^*$ and where the second $*$ is part of a basis vector $\mathbf{e}_i^*$. The equation is false because it is equivalent to the following claim: "if $\{\epsilon^1, ..., \epsilon_n\}$ is an arbitrary basis of V^* then $\phi^{\mathbf{e}_i} = \epsilon^i$", which is false because it is possible to pick a basis for V^* that is not the induced dual basis.

The double dual

Theorem 3.30. ($V \cong V^{**}$ 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 $V \to V^{**}$ that sends $\mathbf{v} \mapsto \Phi_{\mathbf{v}}$, where $\Phi_{\mathbf{v}} : V^* \to K$ is the element of V^{**} defined by $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$.

Proof. Define $\mathbf{F} : V \to V^{**}$ by $\mathbf{F}(\mathbf{v}) = \Phi_{\mathbf{v}}$. We need to show that $\mathbf{F}$ is linear, one-to-one, and onto. Checking linearity is straightforward; $\mathbf{F}$ is linear regardless of the dimensionality of V. As for showing that $\mathbf{F}$ is one-to-one and onto, first recall from Theorem 3.22 that since V is finite-dimensional we have $\dim(V) = \dim(V^{**})$. Thus, to show that $\mathbf{F}$ is an isomorphism it suffices to show that $\mathbf{F}$ has a trivial kernel.

We have the following: if $\mathbf{F}(\mathbf{v}) = \mathbf{0}$, then $\Phi_{\mathbf{v}}$ is the zero function, and so $\Phi_{\mathbf{v}}(\phi)\phi(\mathbf{v}) = 0$ for all $\phi \in V^*$. One would think that this most recent statement implies $\mathbf{v} = \mathbf{0}$, and this is indeed the case, because the contrapositive of $((\forall \phi \in V^* \; \phi(\mathbf{v}) = 0) \implies \mathbf{v} = \mathbf{0})$, which is $(\mathbf{v} \neq \mathbf{0} \implies (\exists \phi \in V^* \; \phi(\mathbf{v}) \neq 0))$, is clearly true: if $\mathbf{v} \neq \mathbf{0}$, then for any basis E of V, some component $([\mathbf{v}]_E)^i \neq 0$ of $\mathbf{v}$ is nonzero, and thus the element of V^* defined by $\mathbf{v} \mapsto ([\mathbf{v}]_E)^i$ sends $\mathbf{v}$ to a nonzero scalar, as desired. $\square$

The dual transformation

Definition 3.31. (Dual transformation).

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

Theorem 3.32. (Dual transformation is represented by transpose matrix).

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

Proof. Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, $F = \{\mathbf{f}_1, ..., \mathbf{f}_m\}$, $E^* = \{\epsilon_1, ..., \epsilon_n\}$, $F^* = \{\delta_1, ..., \delta_n\}$. We will show that the ij entry of $[\mathbf{f}^*(F^*)]_{E^*}$ is the ji entry of $[\mathbf{f}(E)]_F$.

The jth column of $[\mathbf{f}^*(F^*)]_{E^*}$ is $[\mathbf{f}^*(\delta_j)]_{E^*}$. The ith entry of this column is $([\mathbf{f}^*(\delta_j)]_{E^*})_i$. By[4] Theorem 4.28, we have $([\mathbf{f}^*(\delta_j)]_{E^*})_i = \mathbf{f}^*(\delta_j)(\mathbf{e}_i)$. Then $\mathbf{f}^*(\delta_j)(\mathbf{e}_i) = (\delta_j \circ \mathbf{f})(\mathbf{e}_i) = \delta_j(\mathbf{f}(\mathbf{e}_i)) = ([\mathbf{f}(\mathbf{e}_i)]_F)_j$, which is the ji entry of $[\mathbf{f}(E)]_F$. $\square$

Remark 3.33. (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 3.31, is "given a linear function $\mathbf{f} : V \to W$, the dual transformation is the natural linear function $W^* \to V^*$". 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 3.34. (Dual of a composition).

Let U, V, W be finite-dimensional vector spaces, and let $\mathbf{f} : U \to V$, $\mathbf{g} : V \to W$ 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.

Proof. Left as an exercise. $\square$

[4] Though this theorem is from a later chapter, its proof is understandable now. If this is your first time reading this book, then, when reading the theorem, ignore the upper and lower indices, and pretend that all indices are lower.

3.4 (p, q) tensors

Before our detour into the world of dual spaces, we derived Theorem 3.15, and learned that when V and W are finite-dimensional vector spaces, there is a natural isomorphism $\mathbf{F}(V \to W) \cong W \otimes V^*$. One implication of this theorem is of particular interest: the vector space of linear maps $V \to V$ is naturally isomorphic to $V \otimes V^*$. The characterization of linear maps $V \to V$ as elements of $V \otimes V^*$ lends itself to generalization. We can generalize the notion of linear function by "tensor-producting in" more copies of V and V^*! This is the idea behind the following definition.

Definition 3.35. (Tensor space).

Let V be a vector space. We define a *tensor space on* V, or more colloquially, a *tensor space*, to be a vector space of the form $V_1 \otimes ... \otimes V_k$, where each V_i is either V or V^*.

The *type* of a tensor space on V is the sequence of positive integer superscripts and subscripts constructed in the pattern of the following examples[5, 6]:

- The type of the tensor space $V \otimes V^* \otimes V$ is ${}^1{}_1{}^1$.
- The type of the tensor space $V^* \otimes (V)^{\otimes 3}$ is ${}_1{}^3$.
- The type of the tensor space $(V)^{\otimes 2} \otimes V^* \otimes V$ is ${}^2{}_1{}^1$.

We additionally define a *tensor on* V, or more colloquially, a *tensor*, to be an element of a tensor space on V. The *type* of a tensor on V is the type of the tensor space of which the tensor is a member.

Definition 3.36. (Coordinates of a tensor).

Let V be a finite-dimensional vector space over a field K with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be a basis for V^*, and let $T = V_1 \otimes ... \otimes V_k$ be a tensor space on V. The *coordinates of a tensor* $\mathbf{T} \in T$ *relative to* E *and* E^* are the scalars indexed by $i_1, ..., i_k$ such that

- i_ℓ is a superscript iff $V_\ell = V$ and i_ℓ is a subscript iff $V_\ell = V^*$.
- When we omit subscripts and superscripts and denote the scalar indexed by $i_1, ..., i_k$ by $T(i_1, ..., i_k)$, we have

$$\mathbf{T} = \sum_{i_1, ..., i_k \in \{1, ..., n\}} T(i_1, ..., i_k)\ \mathbf{w}(i_1) \otimes ... \otimes \mathbf{w}(i_k),$$

  where $\mathbf{w}(i_\ell) = \mathbf{e}_{i_\ell}$ iff $V_\ell = V$ and $\mathbf{w}(i_\ell) = \epsilon^{i_\ell}$ iff $V_\ell = V^*$.

For example, the coordinates of a tensor $\mathbf{S} \in V^* \otimes V \otimes V^*$ are the $S_{i_1}{}^{i_2}{}_{i_3}$ such that

$$\mathbf{S} = \sum_{i_1, i_2, i_3 \in \{1, ..., n\}} S_{i_1}{}^{i_2}{}_{i_3}\ \epsilon^{i_1} \otimes \mathbf{e}_{i_2} \otimes \epsilon^{i_3}.$$

Each coordinate of a tensor can be thought of as occupying a position in a "multidimensional matrix", where each i_k is associated with an orthogonal axis. (E.g., if a tensor's type is ${}^1{}_1$ or ${}_1{}^1$, then that tensor's coordinates are stored in a two-dimensional matrix.

Definition 3.37. ((p, q) tensor).

Let V be a vector space. We define the vector space $T_{p,q}(V)$ of (p, q) *tensors on* V to be the tensor space on V of type ${}^p{}_q$. That is, the vector space $T_{p,q}(V)$ of (p, q) tensors on V is defined to be $T_{p,q}(V) := V^{\otimes p} \otimes (V^*)^{\otimes q}$.

Remark 3.38. ($(1, 1)$ tensors).

The vector space of $(1, 1)$ tensors on V is equal to the vector space of linear functions $V \to V$.

Remark 3.39. (Multilinearity and "recursive linearity").

[TO-DO: mention second physicist's definition of tensor, call it "recursive linearity". This defn is equivalent to above defn of (p, q) tensor when $V = \mathbb{R}^n$, due to the second boxed equation of Theorem 3.43.]

---

[5] It is possible to give a formal definition of this notion of "type", but any such definition will be verbose and not particularly illustrative of the concept.

[6] This convention goes against Definition 3.20, which sets the convention that vectors are associated with subscripts and covectors are associated with superscripts. This is necessary because we want the pattern of subscripts and superscripts in a tensor's type to mirror the pattern of subscripts and superscripts in a tensor's coordinates.

Remark 3.40. (δ^{ij} vs. $\delta^i{}_j$ vs. $\delta^i{}_j$).

One may be confused when they consider that there are three ways to write the Kronecker delta: $\delta^{ij}, \delta^i{}_j$, and δ_{ij}. The difference between these functions of i and j is straightforward: δ^{ij} is used for coordinates of a $(2,0)$ tensor, $\delta^i{}_j$ is used for a coordinates of a $(1,1)$ tensor, and δ_{ij} is used for coordinates of a $(0,2)$ tensor. All three functions of i and j are defined, as usual, to be 1 when $i = j$ and 0 otherwise.

Remark 3.41. (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 4.36). 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 (p,q) tensor to be a multilinear map $(V^*)^{\times p} \times V^{\times q} \to K$. This definition is equivalent to the one we have used (we see why in Theorem 3.43), 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 "nth order tensor" on V to be[7] 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 $V \to V$. This definition works because we have the natural isomorphism $T_1^1(V) = V \otimes V^* \cong \mathbf{F}(V \to V)$. 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 $T_0^2(V) \cong T_1^1(V) \cong T_2^0(V)$ due to the (unnatural) isomorphism $V \cong V^*$ that is obtained by choosing a basis (see Remark 3.24).

Definition 3.42. (Valence and order of a tensor).

The *valence* of a (p,q) tensor is the tuple (p,q). The *order* of a (p,q) tensor is $p+q$.

Theorem 3.43. (Four fundamental natural isomorphisms for (p,q) tensors).

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

$$\boxed{\begin{gathered}\mathcal{L}(V_1 \times \ldots \times V_k \to W) \cong \mathcal{L}(V_1 \otimes \ldots \otimes V_k \to W)\\ \mathcal{L}(V \to W) \cong W \otimes V^*\\ (V \otimes W)^* \cong V^* \otimes W^*\end{gathered}}$$

Importantly, application of the first and third line, and the fact that $Y \cong Y^{**}$ naturally for any vector space Y, yields

$$T_{p,q}(V) \cong (T_{p,q}(V))^{**} = (V^{\otimes p} \otimes (V^*)^{\otimes q})^{**} \cong ((V^*)^{\otimes p} \otimes V^{\otimes q})^* = \mathcal{L}((V^*)^{\otimes p} \otimes V^{\otimes q} \to K) \cong \mathcal{L}((V^*)^{\times p} \times V^{\times q} \to K).$$

so we have the natural isomorphism

$$\boxed{T_{p,q}(V) = V^{\otimes p} \otimes (V^*)^{\otimes q} \cong \mathcal{L}((V^*)^{\times p} \times V^{\times q} \to K)}$$

Proof. The first line in the first box is Theorem 3.10, and the second line in the first box is Theorem 3.15. 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 $\mathbf{f}_{\phi\otimes\psi} : V \otimes W \to K$ 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 3.44. (The four-fold nature of (p,q) tensors).

We have defined a (p,q) tensor to be an element of a tensor product space; a (p,q) tensor is a "multilinear element". Due to the important natural isomorphisms of the previous theorem we can think of (p,q) tensors in the four ways depicted by this diagram:

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

$$\begin{tikzcd}
\substack{\text{multilinear element} \\ \text{(element of a tensor product space)}} \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \text{multilinear function} \arrow[d, leftrightarrow] \\
\substack{\text{linear element} \\ \text{(element of a vector space)}} \arrow[r, leftrightarrow] & \text{linear function}
\end{tikzcd}$$

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 linear functions 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 3.45. (Other useful natural isomorphisms for (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.

4

Bilinear forms, metric tensors, and coordinates of tensors

The goal of this chapter is to present results regarding the coordinates of (p, q) tensors relative to bases. This is accomplished in the second part of this chapter. The first subsection of the second part, "Coordinates with a metric tensor", is the most important part of this entire chapter. Particularly important is Theorem 4.28, which describes how vectors and dual vectors can act on each other to produce each other's coordinates. The subsequent subsections of the second part of this chapter are less important, but still interesting. In these, we show how to change the bases of a (p, q) tensor and how composition of linear functions generalizes to *tensor contraction*. Of least (direct) importance is the discussion of the convention of *slanted indices*; we have included this because it is handy to know of when exploring literature on tensors.

We build up to these ideas about coordinates by investigating *bilinear forms* in the first part of the chapter. A special type of bilinear form with which the reader may be familiar is an *inner product*; we will define and investigate various facts about these. If you wish to be economical about reading the first part of this chapter, you can skip the middle two sections ("Inner products" and "Symmetric and orthogonal linear functions"); those sections are included for completeness.

4.1 Bilinear forms and metric tensors

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

Let $V_1, ..., V_k$ be vector spaces over a field K. A *linear k-form on* $V_1, ..., V_k$ is[1] a k-linear function $V_1 \times ... \times V_k \to K$.

Let V be a vector space over K. A *linear k-form on* V is a linear k-form on $V^{\times k}$.

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 $(0, k)$ tensors).

A linear k-form on V is an element of $\mathcal{L}(V^{\times k} \to K)$. Recalling Theorem 3.43, we have $\mathcal{L}(V^{\times k} \to K) \cong \mathcal{L}(V^{\otimes k} \to K) = (V^{\otimes k})^* \cong (V^*)^{\otimes k} = T^0_k(V)$. Therefore a linear k-form is naturally identified with a $(0, k)$ 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 $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ defined[2] by $\flat_1(\mathbf{v}) := B(\mathbf{v}, \cdot)$ and $\flat_2(\mathbf{w}) := B(\cdot, \mathbf{w})$. We denote $\mathbf{v}^{\flat_1} := \flat_1(\mathbf{v})$ and $\mathbf{w}^{\flat_2} := \flat_2(\mathbf{w})$.

What would it take for $\flat_1$ and $\flat_2$ to be linear isomorphisms? Well, if we knew that $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ 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 $\flat_1$ and $\flat_2$ would be one-to-one linear functions between finite-dimensional vector spaces of the same dimension, ontoness would follow automatically, and $\flat_1$ and $\flat_2$ would be linear isomorphisms (see Theorem 2.75).

[1] Unfortunately, the word "k-form" without the qualifier "linear" is reserved to mean *differential* k-form. We have not defined differential k-forms yet.

[2] $B(\mathbf{v}, \cdot)$ denotes the function $\mathbf{w} \mapsto B(\mathbf{v}, \mathbf{w})$ and $B(\cdot, \mathbf{w})$ denotes the function $\mathbf{v} \mapsto B(\mathbf{v}, \mathbf{w})$.

Therefore, if $\flat_1$ and $\flat_2$ are one-to-one, then they are linear isomorphisms. When are $\flat_1$ and $\flat_2$ one-to-one? This is the case if and only if their kernels are $\{\mathbf{0}\}$. In other words, $\flat_1$ and $\flat_2$ are isomorphisms iff the bilinear form B satisfies ($B(\mathbf{v},\mathbf{w}) = \mathbf{0} \iff \mathbf{v} = \mathbf{0}$ for all $\mathbf{w} \in W$) and $B((\mathbf{v},\mathbf{w}) = \mathbf{0} \iff \mathbf{w} = \mathbf{0}$ for all $\mathbf{v} \in V$), that is, iff

$$(B(\mathbf{v},\mathbf{w}) = \mathbf{0} \iff \mathbf{v} = \mathbf{0} \text{ or } \mathbf{w} = \mathbf{0}) \text{ for all } \mathbf{v} \in V, \mathbf{w} \in W.$$

A bilinear form B that satisfies the above condition is called *nondegenerate.*

We have contrived nondegenerate bilinear forms to be those for which $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ are natural linear isomorphisms. Note that when $\flat_1$ and $\flat_2$ are isomorphisms, they are indeed natural because they do not depend on a choice of basis (see Definition 2.77). When they are isomorphisms, $\flat_1$ and $\flat_2$ are called the *musical isomorphisms induced by B.* We denote the inverses of $\flat_1$ and $\flat_2$ by $\sharp_1$ and $\sharp_2$, respectively: $\sharp_1 = \flat_1^{-1} : W^* \to V$ and $\sharp_2 = \flat_2^{-1} : V^* \to W$.

Proof. We have $B(\mathbf{v},\mathbf{w}) = \mathbf{v}^{\flat_1}(\mathbf{w})$. Since the $1 \times m$ matrix $\mathbf{v}^{\flat_1}(F)$ of $\mathbf{v}^{\flat_1} : W \to K$ relative to F is $\mathbf{v}^{\flat_1}(F) = [\mathbf{v}^{\flat_1}]_{F^*}^\top$ (recall Theorem 4.29), we apply the characterizing property of [...] matrices (see Derivation [...]) and get $B(\mathbf{v},\mathbf{w}) = \mathbf{v}^{\flat_1}(\mathbf{w}) = \mathbf{v}^{\flat_1}(F)[\mathbf{w}]_F = [\mathbf{v}^{\flat_1}]_{F^*}^\top[\mathbf{w}]_F$. So $B(\mathbf{v},\mathbf{w}) = [\mathbf{v}^{\flat_1}]_{F^*}^\top[\mathbf{w}]_F$. We know $[\mathbf{v}^{\flat_1}]_{F^*} = \mathbf{B}[\mathbf{v}]_E$ from the fourth equation of Theorem 4.30, so $g(\mathbf{v},\mathbf{w}) = (\mathbf{B}[\mathbf{v}]_E)^\top[\mathbf{w}]_F = [\mathbf{v}]_E^\top\mathbf{g}[\mathbf{w}]_F$, which is the first equation of the first line. ☐

Theorem 4.4. (Induced bilinear form on the duals).

Let V and W be finite-dimensional vector spaces over a field K. If B is a nondegenerate bilinear form on V and W, then there is an induced nondegenerate bilinear form $\widetilde{B} : W^* \times V^* \to K$ on W^* and V^* defined by $\widetilde{B}(\psi,\phi) = B(\psi^{\sharp_1},\phi^{\sharp_2})$. The bilinear form $\widetilde{B}$ induces natural isomorphisms $\widetilde{\flat}_1 : V^* \to W^{**}$, $\widetilde{\flat}_2 : W^* \to V^{**}$ defined by $\phi^{\widetilde{\flat}_1}(\psi) = \widetilde{B}(\psi,\phi)$ and $\psi^{\widetilde{\flat}_2}(\phi) = \widetilde{B}(\psi,\phi)$. (Checking that $\widetilde{B}$ is actually a nondegenerate bilinear form is left an exercise).

Lemma 4.5. Let V and W be finite-dimensional vector spaces with bases E and F, let B be a nondegenerate bilinear form on V and W, and let $\widetilde{B}$ be the induced nondegenerate bilinear form on W^* and V^*. Then

$$\begin{aligned} B(\mathbf{v},\mathbf{w}) &= [\mathbf{v}]_E^\top\mathbf{B}[\mathbf{w}]_F = [\mathbf{w}]_F^\top\mathbf{B}^\top[\mathbf{v}]_E \\ \widetilde{B}(\psi,\phi) &= [\psi]_{F^*}^\top\widetilde{\mathbf{B}}[\phi]_{E^*} = [\phi]_{E^*}^\top\widetilde{\mathbf{B}}^\top[\psi]_{F^*} \end{aligned}$$

for all $\mathbf{v} \in V$, $\mathbf{w} \in W$, $\phi \in V^*$, and $\psi \in W^*$.

Proof. To show the first equation of the first line, we follow a derivation similar in spirit to that of a matrix relative to bases of a linear function between vector spaces. That is, we first consider the special case in which the vector spaces are K^n and K^m, where K is the field, and then use change of basis maps to generalize. So: consider the special case $V = K^n$ and $W = K^m$, where $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ and $\hat{\mathfrak{F}} = \{\hat{\mathfrak{f}}_1, ..., \hat{\mathfrak{f}}_m\}$ are the standard bases for K^n and K^m. We have $B(\mathbf{v},\mathbf{w}) = \sum_{ij}([\mathbf{v}]_{\hat{\mathfrak{E}}})^i([\mathbf{w}]_{\hat{\mathfrak{E}}})^j B_{ij} = \sum_{ij}([\mathbf{v}]_{\hat{\mathfrak{E}}})^i B_{ij}([\mathbf{w}]_{\hat{\mathfrak{E}}})^j = \sum_i([\mathbf{v}]_{\hat{\mathfrak{E}}})^i \sum_j B_{ij}([\mathbf{w}]_{\hat{\mathfrak{E}}})^j = \sum_i([\mathbf{v}]_{\hat{\mathfrak{E}}})^i([\mathbf{B}\mathbf{w}]_{\hat{\mathfrak{F}}})^i = \mathbf{v} \cdot (\mathbf{B}\mathbf{w}) = \mathbf{v}^\top\mathbf{B}\mathbf{w} = [\mathbf{v}]_{\hat{\mathfrak{E}}}^\top\mathbf{B}[\mathbf{w}]_{\hat{\mathfrak{F}}}^\top$, so the special case is proven. For the case when V and W are n- and m- dimensional vector spaces with bases E and F, consider the bilinear form C on K^n and K^m defined by $C(\mathbf{v},\mathbf{w}) := B([\mathbf{v}]_E, [\mathbf{w}]_F)$. By the special case, we have $B(\mathbf{v},\mathbf{w}) = [\mathbf{v}]_E^\top\mathbf{C}[\mathbf{w}]_F$, where $\mathbf{C} = (C(\mathbf{e}_i,\mathbf{f}_j))$. We have $\mathbf{C} = (C(\mathbf{e}_i,\mathbf{f}_j)) = (B([\mathbf{e}_i]_E,[\mathbf{f}_i]_F)) = (B(\hat{\mathfrak{e}}_i,\hat{\mathfrak{f}}_i)) = \mathbf{B}$, so $B(\mathbf{v},\mathbf{w}) = [\mathbf{v}]_E^\top\mathbf{B}[\mathbf{w}]_F$, as claimed.

The first equation of the second line is obtained by applying the first line to the induced nondegenerate bilinear form $\widetilde{B}$ on W^* and V^*, and the second equation on each line follows by transposing the first equation on each line. ☐

Theorem 4.6. (Matrices of the musical isomorphisms).

Let V and W be finite-dimensional vector spaces with bases E and F, let B be a bilinear form on V and W, and let E^* and F^* be the bases for V^* and W^* induced by E and F. We have[3]

$$[\flat_1(E)]_{F^*} = \mathbf{B}^\top \text{ and } [\flat_2(F)]_{E^*} = \mathbf{B}, \text{ where } \mathbf{B} := \Big(B(\mathbf{e}_i,\mathbf{e}_j)\Big)$$
$$[\widetilde{\flat_1}(F^*)]_{E^{**}} = (\widetilde{\mathbf{B}})^\top \text{ and } [\flat_2(E^*)]_{F^{**}} = \widetilde{\mathbf{B}}, \text{ where } \widetilde{\mathbf{B}} := \Big(\widetilde{B}(\mathbf{e}_i,\mathbf{e}_j)\Big).$$

[3]One can refer to $\mathbf{B}$ as "the matrix of B relative to E and F" because, as we will see, $\mathbf{B}$ satisfies a characterizing property similar to the one for the matrix $\mathbf{f}(E)$ of a linear function $\mathbf{f} : U \to K^m$, where U is a finite-dimensional vector space over a field K. Similarly, $\widetilde{\mathbf{B}}$ is the matrix of $\widetilde{B}$ relative to F^* and E^*.

Proof. The first line is true because we have $([\flat_1(E)]_{F^*})^i{}_j = ([\flat_1(\mathbf{e}_j)]_{F^*})^i = ([\mathbf{e}_j^{\flat_1}]_{F^*})^i = \mathbf{e}_j^{\flat_1}(\mathbf{f}_i) = B(\mathbf{e}_j, \mathbf{f}_i)$ and $([\flat_2(F)]_{E^*})^i{}_j = ([\flat_2(\mathbf{f}_j)]_{E^*})^i = ([\mathbf{f}_j^{\flat_2}]_{E^*})^i = \mathbf{f}_j^{\flat_2}(\mathbf{e}_i) = B(\mathbf{e}_i, \mathbf{f}_j)$. The second line follows by applying the first line to the bilinear form $\widetilde{B} : W^* \times V^* \to K$. □

Remark 4.7. (Indexing conventions for entries of matrices of bilinear forms).

Consider the hypotheses of the previous theorem and additionally assume that $V = W$. Then $B \in \mathcal{L}(V \times V \to K)$, so B can be identified with a $(0, 2)$ tensor on V, and $\widetilde{B} \in \mathcal{L}(V^* \times V^* \to K)$, so $\widetilde{B}$ can be identified with a $(2, 0)$ tensor on V. (Recall Remark 4.2). To make sure that we are following the convention regarding coordinates of tensors established in Definition 3.36, we use B_{ij} to denote the ij entry of the matrix $\mathbf{B}$ and use B^{ij} to denote the ij entry of the matrix $\widetilde{\mathbf{B}}$.

Theorem 4.8. ($\flat_1$ and $\widetilde{\flat}_1$ are "kind of" inverses).

Let V and W be finite-dimensional vector spaces over a field K with bases E and F, let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_n}\}$ be the bases for V^* and W^* induced by E and F, and let E^{**} and F^{**} be the bases for V^{**} and W^{**} induced by E^* and F^*. Lastly, let B be a nondegenerate bilinear form on V and W with induced musical isosmorphisms $\flat_1, \flat_2$, and let $\widetilde{B}$ be the induced nondegenerate bilinear form on W^* and V^* with induced musical isomorphisms $\widetilde{\flat}_1, \widetilde{\flat}_2$.

The musical isomorphisms $\flat_1$ and $\widetilde{\flat}_1$ are "kind of" inverses in the following sense:

1. $\widetilde{\flat}_1 \circ \flat_1 = (\mathbf{v} \mapsto \Phi_{\mathbf{v}})$
2. The matrices of $\widetilde{\flat}_1$ and $\flat_1$ relative to the appropriate bases are inverses: $[\widetilde{\flat}_1(F^*)]_{E^{**}}[\flat_1(E)]_{F^*} = \mathbf{I}$.
3. $\mathbf{B}^{-1} = \widetilde{\mathbf{B}}$, where $\widetilde{\mathbf{B}} = (\widetilde{B}(\psi^{\mathbf{f}_i}, \phi^{\mathbf{e}_j}))$.

Proof.

1. We have: $\widetilde{\flat}_1 \circ \flat_1 = (\mathbf{v} \mapsto \Phi_{\mathbf{v}}) \iff \Phi_{\mathbf{v}} = (\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1} \iff (\Phi_{\mathbf{v}}(\phi)(\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1}(\phi)$ for all $\phi \in V^*)$
$\iff (\phi(\mathbf{v}) = \widetilde{g}(\mathbf{v}^{\flat_1}, \phi) = g(\mathbf{v}, \phi^{\sharp_2})$ for all $\phi \in V^*)$
$\iff (\phi(\mathbf{e}_i) = g(\mathbf{e}_i, \phi^{\sharp_2})$ for all $\phi \in V^*) \iff (([\phi]_{E^*})_i = g(\mathbf{e}_i, \phi^{\sharp_2})$ for all $\phi \in V^*)$.
2. The matrix of a composition of a linear function is equal to the matrix-matrix product of the matrices (with respect to the appropriate bases) of the functions in the composition, so we have $[\widetilde{\flat}_1 \circ \flat_1]_{E^{**}} = [\widetilde{\flat}_1(F^*)]_{E^{**}}[\flat_1(E)]_{F^*}$. Since $\widetilde{\flat}_1 \circ \flat_1 = (\mathbf{v} \mapsto \Phi_{\mathbf{v}})$, we can prove the desired fact- that the matrices on the right side of this last equation are inverses- by showing that the matrix of $(\mathbf{v} \mapsto \Phi_{\mathbf{v}}) = \mathbf{F}$ relative to E and E^{**} is $\mathbf{I}$. To do so, define $\mathbf{F}(\mathbf{v}) = \Phi_{\mathbf{v}}$, so that the matrix of $(\mathbf{v} \mapsto \Phi_{\mathbf{v}}) = \mathbf{F}$ relative to E and E^{**} is $[\mathbf{F}(E)]_{E^{**}}$. The ij entry of this matrix is $([\mathbf{F}(\mathbf{e}_j)]_{E^{**}})^i = ([\Phi_{\mathbf{e}_j}]_{E^{**}})^i = \Phi_{\mathbf{e}_j}(\phi^{\mathbf{e}_i})$, where $\phi^{\mathbf{e}_i}$ denotes the ith basis vector in E^*. Since $\Phi_{\mathbf{e}_j}(\phi^{\mathbf{e}_i}) = \phi^{\mathbf{e}_i}(\mathbf{e}_j) = \delta^i{}_j$, the matrix itself is $\mathbf{I}$.
3. Theorem 4.6 tells us that $[\flat_1(E)]_{F^*} = \mathbf{B}^\top$ and $[\widetilde{\flat}_1(F^*)]_{E^{**}} = (\widetilde{\mathbf{B}})^\top$. Using these results with (2), we obtain $(\widetilde{\mathbf{B}})^\top \mathbf{B}^\top = \mathbf{I}$. Take the transpose of both sides to obtain $\widetilde{\mathbf{B}}\mathbf{B} = \mathbf{I}$, as desired. **before using transpose have to define symmetric linear function**

□

Theorem 4.9. (Self-dual $\iff$ orthonormal).

Let V be a finite-dimensional vector space with a nondegenerate bilinear form B and a basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. Let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^*. The musical isomorphism $\flat$ induced by B sends basis vectors to induced dual basis vectors, $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$, iff $B(\mathbf{e}_i, \mathbf{e}_j) = \delta^i_j$ for all i, j.

In the special case when B is an *inner product*, this means that $\flat$ sends basis vectors to induced dual basis vectors iff E is *orthonormal* with respect to B. (We have not defined what an "inner product" is or what "orthonormal" means yet, but the previous sentence should be immediately clear after we do touch on these concepts).

Proof. We have: $\mathbf{e}_i^\flat = \phi^{\mathbf{e}_i} \iff \mathbf{e}_i^\flat(\mathbf{v}) = \phi^{\mathbf{e}_i}(\mathbf{v})$ for all $\mathbf{v} \in V \iff \mathbf{e}_i^\flat(\mathbf{e}_j) = \phi^{\mathbf{e}_i}(\mathbf{e}_j) \iff B(\mathbf{e}_i, \mathbf{e}_j) = \phi^{\mathbf{e}_i}(\mathbf{e}_j) \iff B(\mathbf{e}_i, \mathbf{e}_j) = \delta^i{}_j$. □

Remark 4.10. (Naturality of self-duality).

The above theorem shows that the "unnatural" isomorphism $V \to V^*$ sending $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ that was discussed earlier in Remark 3.24 actually becomes natural exactly whenever we have a bilinear form B on V and $B(\mathbf{e}_i, \mathbf{e}_j) = \delta^i_j$.

Inner products

Definition 4.11. (Metric tensor).

Let V be a finite-dimensional vector space. A nondegenerate bilinear form on V that is also *symmetric*, in the sense that $g(\mathbf{v}_1, \mathbf{v}_2) = g(\mathbf{v}_2, \mathbf{v}_1)$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$, is called a *metric tensor on* V.

Definition 4.12. (The notation $\flat$ and $\sharp$).

When V is a finite-dimensional vector space and there is a metric tensor g on V, then the musical isomorphisms $\flat_1 : V \to V^*$ and $\flat_2 : V \to V^*$ induced by g are the same because g is symmetric. In this scenario, we define $\flat := \flat_1 = \flat_2$ and $\sharp := \sharp_1 = \flat_1^{-1} = \sharp_2 = \flat_2^{-1}$.

Definition 4.13. (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 $g(\mathbf{v}, \mathbf{v}) \geq 0$ for all $\mathbf{v} \in V$, with $g(\mathbf{v}, \mathbf{v}) = \mathbf{0}$ iff $\mathbf{v} = \mathbf{0}$.

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 4.14. (Inner product space).

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 4.15. (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 4.16. 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

Definition 4.17. (Length of a vector with respect to an inner product).

Let V be an inner product space. 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 2.134), 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 4.18. (Angle between vectors with respect to an inner product).

Let V be an inner product space. In analogy to the the fact that the angle between vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ is $\arccos(\hat{\mathbf{v}}_1 \cdot \hat{\mathbf{v}}_2)$, 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 $\theta := \arccos(\langle \hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2 \rangle)$. Note that $\hat{\mathbf{v}}_i = \frac{\mathbf{v}_i}{||\mathbf{v}_i||} = \frac{\mathbf{v}_i}{\sqrt{\langle \mathbf{v}_i, \mathbf{v}_i \rangle}}$.

Remark 4.19. (Geometric inner product).

Let V be an inner product space. 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 Theorem 2.140).

Theorem 4.20. (Cauchy-Schwarz inequality for vector spaces over $\mathbb{R}$).

Let V be a vector space over $\mathbb{R}$ with inner product. Then the *Cauchy-Schwarz inequality* holds: $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle \leq ||\mathbf{v}_1||\, ||\mathbf{v}_2||$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.

Note, the Cauchy-Schwarz inequality is equivalent to the statement that the angle θ in V between $\mathbf{v}_1$ and $\mathbf{v}_2$ with respect to the inner product on V satisfies $\theta \in [0, 2\pi)$.

Proof. Define $f : \mathbb{R} \to [0, \infty) \subseteq \mathbb{R}$ by $f(k) = \langle k\mathbf{v}_1 + \mathbf{v}_2, k\mathbf{v}_1 + \mathbf{v}_2 \rangle = k^2 \langle \mathbf{v}_1, \mathbf{v}_1 \rangle + 2k \langle \mathbf{v}_1, \mathbf{v}_2 \rangle + \langle \mathbf{v}_2, \mathbf{v}_2 \rangle$. Set $a := \langle \mathbf{v}_1, \mathbf{v}_1 \rangle$, $b := 2\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$, and $c := \langle \mathbf{v}_2, \mathbf{v}_2 \rangle$, so that $f(k) = ak^2 + bk + c$.

Since $\langle \cdot, \cdot \rangle$ is positive-definite, then f is nonnegative, and therefore must have either one or zero real roots. According to the quadratic formula, one real root occurs when $b^2 - 4ac = 0$, and zero real roots occur when $b^2 - 4ac < 0$. So, we must have $b^2 - 4ac \leq 0$.

Using our expressions for a, b, and c, we see that $(4\langle \mathbf{v}_1, \mathbf{v}_2 \rangle)^2 - 4(\langle \mathbf{v}_1, \mathbf{v}_2 \rangle)(\langle \mathbf{v}_2, \mathbf{v}_2 \rangle) \leq 0$. 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 4.21. (Orthogonality of vectors with respect to an inner product).

Let V be an inner product space. 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 4.22. (Orthonormal basis with respect to an inner product).

Let V be a finite-dimensional vector space with inner product $\langle\cdot,\cdot\rangle$. We say a basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ of V is *orthonormal (with respect to* $\langle\cdot,\cdot\rangle$*)* 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 $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta^i{}_j$ for all i, j.

Theorem 4.23. (Gram-Schmidt algorithm).

Let V be a finite-dimensional inner product space. Given any basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ for V, we can use the following *Gram-Schmidt algorithm* to convert E into an orthonormal basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$.

First, we "orthogonalize" the basis E into a basis $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$. Set $\mathbf{f}_1 := \mathbf{e}_1$, and, for $i \geq 2$, set

$$\mathbf{f}_i := \mathbf{e}_i - \text{proj}(\mathbf{f}_i \to \text{span}(\mathbf{e}_1, ..., \cancel{\mathbf{e}_i}, ..., \mathbf{e}_n)) = \mathbf{e}_i - \sum_{j \neq i} \text{proj}(\mathbf{f}_i \to \mathbf{e}_j) = \mathbf{e}_i - \sum_{j \neq i} \frac{\langle \mathbf{f}_i, \mathbf{e}_j \rangle}{\langle \mathbf{e}_j, \mathbf{e}_j \rangle}, \quad i \geq 2.$$

(In the last equality in the line above, we've used an analogue of Theorem 2.139 to express vector projections in terms of inner products). To obtain the orthonormal basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$, we just normalize the orthogonal basis F, and set $\hat{\mathbf{u}}_i := \frac{\mathbf{f}_i}{||\mathbf{f}_i||}$.

Symmetric and orthogonal linear functions

This subsection is presented only for 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 4.27 satisfied by an orthogonal linear function.

Derivation 4.24. (The adjoint of a linear function with respect to a nondegenerate bilinear form).

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 $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ induced by B.

There is an induced linear map $\mathbf{f}^\dagger : V \to W$, called the *adjoint of* $\mathbf{f}$ *(with respect to* B*)*, that is obtained by using the musical isomorphisms on the domain and codomain of the dual transformation $\mathbf{f}^* : W^* \to V^*$. We define $\mathbf{f}^\dagger$ to be the unique map for which the following diagram commutes:

$$\begin{tikzcd} V \arrow[r, "\mathbf{f}^\dagger"] \arrow[d, "\flat_1"'] & W \arrow[d, "\flat_2"] \\ W^* \arrow[r, "\mathbf{f}^*"'] & V^* \end{tikzcd}$$

Thus, $\mathbf{f}^\dagger = \flat_2^{-1} \circ \mathbf{f}^* \circ \flat_1$, or, equivalently, $\flat_2 \circ \mathbf{f}^\dagger = \mathbf{f}^* \circ \flat_1$.

Since $\flat_2 \circ \mathbf{f}^\dagger = \mathbf{f}^* \circ \flat_1$, then $\mathbf{f}^\dagger(\mathbf{v}_0)^{\flat_2} = \mathbf{f}^*(\mathbf{v}_0^{\flat_1})$. Using the definition of $\mathbf{f}^*$ (recall Definition 3.31), we have $\mathbf{f}^\dagger(\mathbf{v}_0)^{\flat_2} = \mathbf{v}_0^{\flat_1} \circ \mathbf{f}$. This is the same[4] as $B(\cdot, \mathbf{f}^\dagger(\mathbf{v}_0)) = B(\mathbf{v}_0, \cdot) \circ \mathbf{f}$. After evaluating both sides of this most recent equation on $\mathbf{v}_1 \in V$ and relabeling $\mathbf{v}_0$ as $\mathbf{v}_2$, we have:

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

Replacing the bilinear form of the above derivation with an inner product, we see that when we have an inner product instead of an arbitrary nondegenerate bilinear form, then $\mathbf{f}^\dagger$ is uniquely characterized by the condition that $\langle \mathbf{v}_1, \mathbf{f}^\dagger(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$. We will use this condition as a means to characterize special classes of linear functions. Specifically, we consider linear functions $\mathbf{f}$ for which $\mathbf{f} = \mathbf{f}^\dagger$ and for which $\mathbf{f}^\dagger = \mathbf{f}^{-1}$. Before investigating such linear functions, we quickly present a definition.

Definition 4.25. (Transpose of a matrix).

The *transpose* of an $m \times n$ matrix $(a^i{}_j)$ is the $n \times m$ matrix $(a^j{}_i)$. The tranpose of a matrix $\mathbf{A}$ is denoted $\mathbf{A}^\top$.

Now, we consider linear functions $\mathbf{f}$ for which $\mathbf{f} = \mathbf{f}^\dagger$ and for which $\mathbf{f}^\dagger = \mathbf{f}^{-1}$.

[4] Here we've used $B(\cdot, \mathbf{v}_2)$ to denote the function $\mathbf{v}_1 \mapsto B(\mathbf{v}_1, \mathbf{v}_2)$.

Definition 4.26. (Symmetric linear function).

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

1. $\mathbf{f} = \mathbf{f}^\dagger$.
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. If V is finite-dimensional and $\widehat{U}$ is an orthonormal basis of V, then the matrix of $\mathbf{f}$ relative to $\widehat{U}$ is equal to its transpose: $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top = [\mathbf{f}(\widehat{U})]_{\widehat{U}}$. (Matrices that are equal to their own transpose are called *symmetric*).

Proof. We need to show that the conditions are indeed equivalent.

$(1 \iff 2)$.

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

$(\impliedby)$. 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 $\langle \mathbf{v}_1, \mathbf{f}^\dagger(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle$. Therefore $\langle \mathbf{v}_1, \mathbf{f}(\mathbf{v}_2) \rangle = \langle \mathbf{v}_1, \mathbf{f}^\dagger(\mathbf{v}_2) \rangle$. Due to the cancelability of inner products (this follows from the positive-definiteness of inner products), we have $\mathbf{f}(\mathbf{v}_2) = \mathbf{f}^\dagger(\mathbf{v}_2)$ for all $\mathbf{v}_2 \in V$. So $\mathbf{f} = \mathbf{f}^\dagger$.

$(2 \iff 3)$.

$(\implies)$. If (2) is true, then the jth column of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$ is $[\mathbf{f}(\hat{\mathbf{u}}_j)]_{\widehat{U}}$, so the ${}^i_j$ entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$ is $\langle \mathbf{f}(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle$. Similarly, the ${}^i_j$ entry of the matrix of $\mathbf{f}^\dagger$ relative to $\widehat{U}$ is $\langle \mathbf{f}^\dagger(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle$. Due to the condition on $\langle \cdot, \cdot \rangle$ induced by the identification $V \cong V^*$, we have $\langle \mathbf{f}^\dagger(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle = \langle \hat{\mathbf{u}}_j, \mathbf{f}(\hat{\mathbf{u}}_i) \rangle = \langle \mathbf{f}^\dagger(\hat{\mathbf{u}}_i), \hat{\mathbf{u}}_j \rangle$. But this is the ${}^j_i$ entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$, i.e., the ${}^i_j$ entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$. Thus $[\mathbf{f}(\widehat{U})]_{\widehat{U}} = [\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$.

$(\impliedby)$. Let $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ be an orthonormal basis for V, and let the matrix $\mathbf{A}$ of $\mathbf{f}$ relative to $\widehat{U}$ satisfy $a_{ij} = a^j_i$. Since $a^i{}_j = \langle \mathbf{f}(\hat{\mathbf{u}}_i), \hat{\mathbf{u}}_j \rangle$, 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. □

Definition 4.27. (Orthogonal linear function).

Let V be a vector space over K (where[5] we have $K \neq \mathbb{Z}/2\mathbb{Z}$), consider a linear function $\mathbf{f} : V \to V$. We define $\mathbf{f}$ to be *orthogonal* iff the following equivalent conditions hold:

1. $\mathbf{f}^\dagger = \mathbf{f}^{-1}$.
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 $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^{-1} = [\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$.

Proof. We need to show that the conditions are indeed equivalent. To do so, 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 $\theta = \cos^{-1}\left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_1||||\mathbf{v}_2||}\right)$. 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 2.143 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 $(1) \iff (2) \iff (3) \iff (6) \iff (7)$.

$(1 \iff 2)$.

[5] This 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.

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

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

($2 \implies 3$). 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 $\mathbf{v}_1, \mathbf{v}_3 \in V$.

($3 \iff 6$).

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

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

($6 \iff 7$).

($\implies$). The ${}^i_j$ entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$ is (ith row of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$)$\cdot$($j$th column of $[\mathbf{f}(\widehat{U})]_{\widehat{U}})^\top$ = (ith row of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$)$\cdot$($j$th row of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$) $= \langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j)\rangle = \delta^i{}_j$. Therefore $[\mathbf{f}(\widehat{U})]_{\widehat{U}}[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top = \mathbf{I}$. A similar argument shows $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top[\mathbf{f}(\widehat{U})]_{\widehat{U}} = \mathbf{I}$.

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

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

4.2 Coordinates of (p, q) tensors

Theorem 4.28. (Coordinates of vectors and dual vectors).

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

We have

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

Recall from Theorem 3.30 that $\mathbf{v}$ is naturally identified with $\Phi_{\mathbf{v}} \in V^{**}$ defined by $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$.

When trying to remember this theorem, it may help to involve the contraction map $C : V \times V^* \to K$ defined by $C(\mathbf{v}, \phi) = \phi(\mathbf{v})$, since we have the following:

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

In other words, we take the ith coordinate of a vector by contracting that vector with the relevant dual basis vector, and we take the ith coordinate of a dual vector by contracting that dual vector with the relevant "regular" basis vector.

Proof. First, we prove the first equation in the first line, $([\mathbf{v}]_E)^i = \phi^{\mathbf{e}_i}(\mathbf{v})$:

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

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

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

Now we prove the second equality of the first line. Recall from Theorem 3.30 that $\Phi_{\mathbf{v}}$ is defined as $\Phi_{\mathbf{v}}(\phi) := \phi(\mathbf{v})$. Therefore $\phi^{\mathbf{e}_i}(\mathbf{v}) = \Phi_{\mathbf{v}}(\phi^{\mathbf{e}_i})$. By applying the second line, we then have $\Phi_{\mathbf{v}}(\phi^{\mathbf{e}_i}) = ([\Phi_{\mathbf{v}}]_{E^{**}})^i$. □

Theorem 4.29. (Matrix of dual vector as transposed coordinates).

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

Proof. The matrix of ϕ relative to E and $\hat{\mathfrak{E}}$ is

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

.

On the other hand, the previous theorem tells us that coordinates of ϕ relative to E^* are

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

Inspecting the two above results shows that $\phi(E) = [\phi]_{E^*}^\top$. □

Coordinates with nondegenerate bilinear form

Theorem 4.30. (Relationship between coordinates of vectors and dual vectors for vector spaces).

Let V and W be finite-dimensional vector spaces over a field K with bases E and F, let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_n}\}$ be the bases for V^* and W^* induced by E and F, and let E^{**} and F^{**} be the bases for V^{**} and W^{**} induced by E^* and F^*. Lastly, let B be a nondegenerate bilinear form on V and W with induced musical isosmorphisms $\flat_1, \flat_2$, and let $\widetilde{B}$ be the induced nondegenerate bilinear form on W^* and V^* with induced musical isomorphisms $\widetilde{\flat}_1, \widetilde{\flat}_2$.

Then we have

$$
\begin{aligned}
[\mathbf{v}]_E = \mathbf{B}^{-\top}[\mathbf{v}^{\flat_1}]_{F^*} \text{ and } [\mathbf{w}]_F = \mathbf{B}^{-1}[\mathbf{w}^{\flat_2}]_{E^*} \\
[\phi]_{E^*} = \mathbf{B}[\phi^{\sharp_2}]_F \text{ and } [\psi]_{F^*} = \mathbf{B}^\top[\psi^{\sharp_1}]_E,
\end{aligned}
$$

for all $\mathbf{v} \in V$, $\mathbf{w} \in W$, $\phi \in V^*$, and $\psi \in W^*$, where $\mathbf{B} = (B(\mathbf{e}_i, \mathbf{e}_j))$ and $\widetilde{\mathbf{B}} = (\widetilde{B}(\psi^{\mathbf{f}_i}, \phi^{\mathbf{e}_j}))$.

Proof. We will prove (1) that $[\psi]_{F^*} = \mathbf{B}[\psi^{\sharp_1}]_E$ for all $\psi \in W^*$ and (2) that $[\mathbf{v}]_E = \mathbf{B}^{-1}[\mathbf{v}^{\flat_1}]_{F^*}$ for all $\mathbf{v} \in V$. The equation involving $\phi \in V^*$ and $\mathbf{B}$ is obtained by applying (1) to the nondegenerate bilinear form C on W and V defined by $C(\mathbf{w}, \mathbf{v}) := B(\mathbf{v}, \mathbf{w})$, and the equation involving $\mathbf{w} \in W$ and $\mathbf{B}^{-1}$ is similarly obtained by applying (2) to the nondegenerate bilinear form $\widetilde{C}$ on V^* and W^* defined by $\widetilde{C}(\phi, \psi) := \widetilde{B}(\psi, \phi)$.

First, we prove that $[\psi]_{F^*} = \mathbf{B}[\psi^{\sharp_1}]_E$ for all $\psi \in W^*$. Recall from Definition [...] that the matrix $[\flat_1(E)]_{F^*}$ of $\flat_1 : V \to W^*$ relative to E and F^* satisfies the characterizing property $[\flat_1(\mathbf{v})]_{F^*} = [\flat_1(E)]_{F^*}[\mathbf{v}]_E$ for all $\mathbf{v} \in V$. That is, $[\mathbf{v}^{\flat_1}]_{F^*} = [\flat_1(E)]_{F^*}[\mathbf{v}]_E$ for all $\mathbf{v} \in V$. We know from Theorem 4.6 that $[\flat_1(E)]_{F^*} = \mathbf{B}^\top$, so we have $[\mathbf{v}^{\flat_1}]_{F^*} = \mathbf{B}^\top[\mathbf{v}]_E$. Since $\flat_1$ is an isomorphism, we can replace $\mathbf{v}^{\flat_1} \in W^*$ with an arbitrary $\psi \in W^*$ to obtain ($[\psi]_{F^*} = \mathbf{B}^\top[\psi^{\sharp_1}]_E$ for all $\psi \in W^*$), as desired.

Now we prove $[\mathbf{v}]_E = \mathbf{B}^{-\top}[\mathbf{v}^{\flat_1}]_{F^*}$ for all $\mathbf{v} \in V$. Applying (1) to the induced nondegenerate bilinear form $\widetilde{B} : W^* \times V^* \to K$, we obtain an analogous statement to (1) in which $\psi \in W^*$ is replaced with $\Phi \in V^{**}$, F^* is replaced with E^{**}, E is replaced with F^*, $\sharp_1 : W^* \to V$ is replaced with $\widetilde{\sharp}_1 : V^{**} \to W^*$, and $\mathbf{B}$ is replaced with $\widetilde{\mathbf{B}}$: we have $[\Phi]_{E^{**}} = \widetilde{\mathbf{B}}^\top[\Phi^{\sharp_1}]_{F^*}$ for all $\Phi \in V^{**}$. Substitute $\Phi_\mathbf{v}$ in for Φ and use the fact $([\Phi_\mathbf{v}]_{E^{**}})^i = ([\mathbf{v}]_E)^i$ from Theorem 4.28 to obtain the statement "$[\mathbf{v}]_E = \widetilde{\mathbf{B}}^\top[\Phi_\mathbf{v}^{\widetilde{\sharp}_1}]_{F^*}$ for all $\mathbf{v} \in V$". Notice that if we show that $\Phi_\mathbf{v}^{\widetilde{\sharp}_1} = \mathbf{v}^{\flat_1}$ for all $\mathbf{v} \in V$, then this equation involves only $\mathbf{v}$ and not $\Phi_\mathbf{v}$, and becomes $[\mathbf{v}]_E = \widetilde{\mathbf{B}}^\top[\mathbf{v}^{\flat_1}]_{F^*}$. This condition does hold: we have ($\Phi_\mathbf{v}^{\widetilde{\sharp}_1} = \mathbf{v}^{\flat_1}$ for all $\mathbf{v} \in V$) $\iff$ ($\Phi_\mathbf{v} = (\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1}$ for all $\mathbf{v} \in V$) $\iff$ ($(\mathbf{v} \mapsto \Phi_\mathbf{v}) = \widetilde{\flat}_1 \circ \flat_1$), where this last condition is just the first item of Theorem 4.8. So we know $[\mathbf{v}]_E = \widetilde{\mathbf{B}}^\top[\mathbf{v}^{\flat_1}]_{F^*}$ for all $\mathbf{v} \in V$. Use the fact that $\widetilde{\mathbf{B}} = \mathbf{B}^{-1}$ from Theorem 4.8 to obtain ($[\mathbf{v}]_E = (\mathbf{B}^{-1})^\top[\mathbf{v}^{\flat_1}]_{F^*} = (\mathbf{B}^{-\top})[\mathbf{v}^{\flat_1}]_{F^*}$ for all $\mathbf{v} \in V$, as desired. ☐

Remark 4.31. (Metric tensors and coordinates of vectors and dual vectors).

When tensors are used in physics, the above theorem is employed to the situation in which we have a vector space V with a metric tensor g. (Recall that a metric tensor on V is a nondegenerate symmetric bilinear form on V).

Since g is symmetric, we have $\flat := \flat_1 = \flat_2$, $\sharp := \sharp_1 = \sharp_2$, and $\mathbf{g}^\top = \mathbf{g}$, so the above equations simplify to

$$
\begin{aligned}
[\mathbf{v}]_E = \mathbf{g}^{-1}[\mathbf{v}^\flat]_{E^*} \iff [\mathbf{v}^\flat]_{E^*} = \mathbf{g}[\mathbf{v}]_E \\
[\phi]_{E^*} = \mathbf{g}[\phi^\sharp]_E \iff [\phi^\sharp]_E = \mathbf{g}^{-1}[\phi]_{E^*}.
\end{aligned}
$$

Physicists also make the definitions $v^i := ([\mathbf{v}]_E)^i$, $v_i := ([\mathbf{v}^\flat]_{E^*})_i$, $\phi_i := ([\phi]_{E^*})_i$, and $\phi^i := ([\phi^\sharp]_E)^i$. With these definitions, the above equations become

$$
\begin{aligned}
v^i = \sum_j g^{ij} v_j \iff v_i = \sum_j g_{ij} v^j \\
\phi_i = \sum_j g_{ij} \phi^j \iff \phi^i = \sum_j g^{ij} \phi_j.
\end{aligned}
$$

This notation has the advantage of being compact, and it's what I would personally use when doing calculations. However, it is best for reference materials such as this one to introduce the relations between v^i and v_i with notation that does involve explicit mention of the basis E.

investigate whether below theorem can be discovered without knowing the above thm about **B** and coordinates

[$T_{p,q}(V) \cong T_s^r(V)$ when $p+q=r+s$. Normally this isomorphism is not natural, and we have to obtain it by choosing a basis. But when we have a metric tensor g on V, then the musical isomorphism $\flat : V \to V^*$ achieves this isomorphism naturally.]

Theorem 4.32. (Using a metric tensor to convert between vectors and dual vectors in a (p,q) tensor).

double check whether $B_{ij} = B_{ji}$ is used or not in this theorem. i really hope it isn't...

seems like it's not. we'll just have to change $\flat$ to $\flat_1$ and state two extra results at end to generalize to case when multiplying by B_{jr} rather than B_{rj}

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let B be a metric tensor on V, with $\flat = \flat_1 = \flat_2$ being the musical isomorphism $V \to V^*$ (recall Definition 4.3), and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be the dual basis for V^* induced by E. (Note: recall from Remark 4.10 that $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ is a natural isomorphism, due to the presence of $\flat$).

We can send a basis (p,q) tensor $\mathbf{T} = \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_k} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes ... \otimes \phi^{\mathbf{e}_{j_q}} \in T_{p,q}(V)$ to a $(p-1, q+1)$ tensor by applying the map $\flat : V \to V^*$ to one of the p vectors in $\mathbf{T}$.

$$\begin{gathered}
\mathbf{T} = \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_k} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes ... \otimes \phi^{\mathbf{e}_{j_q}} \\
\longmapsto \\
\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_k}^{\flat} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes ... \otimes \phi^{\mathbf{e}_{j_q}}.
\end{gathered}$$

Using Theorem **??**, we compute $\mathbf{e}_{i_k}^{\flat}$ to be

$$\mathbf{e}_{i_k}^{\flat} = \sum_{r=1}^{n} ([\mathbf{e}_{i_k}^{\flat}]_{E^*})_r \phi^{\mathbf{e}_r} \underset{\text{Theorem } \mathbf{??}}{=} \sum_{r=1}^{n} \Big(\sum_{j=1}^{n} g_{rj} ([\mathbf{e}_{i_k}]_E)^j \Big) \phi^{\mathbf{e}_r} = \sum_{r=1}^{n} \Big(\sum_{j=1}^{n} g_{rj} \delta^j{}_{i_k} \Big) \phi^{\mathbf{e}_r} = \sum_{r=1}^{n} g_{i_k r} \phi^{\mathbf{e}_r},$$

so $\mathbf{T}$ is sent to

$$\begin{aligned}
&\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_{k-1}} \otimes \sum_{r=1}^{n} \Big(g_{i_k r} \phi^{\mathbf{e}_r} \Big) \otimes \mathbf{e}_{i_{k+1}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes ... \otimes \phi^{\mathbf{e}_{j_q}} \\
&= \sum_{r=1}^{n} g_{i_k r} \Big(\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_{k-1}} \otimes \phi^{\mathbf{e}_r} \otimes \mathbf{e}_{i_{k+1}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes ... \otimes \phi^{\mathbf{e}_{j_q}} \Big).
\end{aligned}$$

Thus, if the coordinates of $\mathbf{T}$ relative to E and E^* were originally $T^{i_1 \dots i_p}{}_{j_1 \dots j_q}$, then they get sent to

$$\sum_{r=1}^{n} g_{i_k r} T^{i_1 \dots i_{k-1}}{}_{r}{}^{i_{k+1} \dots i_p}{}_{j_1 \dots j_q}.$$

Following a similar process to above, we can use the other musical isomorphism, the sharp map $\sharp = \flat^{-1}$, to convert a (p,q) tensor to a $(p+1, q-1)$ tensor. This approach would send $T^{i_1 \dots i_p}{}_{j_1 \dots j_q}$ to

$$\sum_{r=1}^{n} g^{j_k r} T^{i_1 \dots i_p}{}_{j_1 \dots j_{k-1}}{}^{r}{}_{j_{k+1} \dots j_q}.$$

So we have the following "index lowering" and "index raising" mappings:

$$\boxed{\begin{aligned}
T^{i_1 \dots i_p}{}_{j_1 \dots j_q} &\mapsto \sum_{r=1}^{n} g_{i_k r} T^{i_1 \dots i_{k-1}}{}_{r}{}^{i_{k+1} \dots i_p}{}_{j_1 \dots j_q} \quad \text{(index lowering)} \\
T^{i_1 \dots i_p}{}_{j_1 \dots j_q} &\mapsto \sum_{r=1}^{n} g^{j_k r} T^{i_1 \dots i_p}{}_{j_1 \dots j_{k-1}}{}^{r}{}_{j_{k+1} \dots j_q} \quad \text{(index raising)}
\end{aligned}}$$

Change of basis for (p,q) tensors

Theorem 4.33. (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

$$\boxed{\begin{aligned} [\mathbf{v}]_F &= [\mathbf{E}]_F[\mathbf{v}]_E = [\mathbf{F}]_E^{-1}[\mathbf{v}]_E \\ [\phi]_{F^*} &= [\mathbf{E}]_F^{-\top}[\phi]_{E^*} = [\mathbf{F}]_E^{\top}[\phi]_{E^*} \\ [\mathbf{E}]_F &= [\mathbf{E}^*]_{F^*} = [\mathbf{F}]_E^{-1} \end{aligned}}$$

where $\mathbf{v} \in V$ and $\phi \in V^*$.

Proof. The first line of the boxed equation is Theorem 2.107, and the equation $[\mathbf{E}]_F = [\mathbf{F}]_E^{-1}$ from the third line is Theorem 2.109.

The equation $[\mathbf{E}]_F = [\mathbf{E}^*]_{F^*}$ from the third line is true because if $\mathbf{G} : V \to V^*$ is the isomorphism sending basis vectors to induced dual basis vectors, then we have $[\mathbf{v}]_E = [\mathbf{G}(\mathbf{v})]_{E^*}$ for any $\mathbf{v} \in V$ (see Theorem 3.28).

The second line follows by noticing that the first line implies $[\phi]_{F^*} = [\mathbf{E}^*]_{F^*}[\mathbf{v}]_{E^*}$, and then applying the equation $[\mathbf{E}]_F = [\mathbf{E}^*]_{F^*}$ of the third line. □

Remark 4.34. (What "covariance" and "contravariance" refer to).

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 applying the matrix transpose to 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 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 $[\mathbf{F}]_E^{-1}$. Thus, dual vectors "co-vary" *with* $[\mathbf{F}]_E$ when changing basis from E to E^*, and vectors "contra-vary" *against* $[\mathbf{F}]_E$ when changing basis from F to F^*.

Theorem 4.35. (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 $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_n}\}$ be the corresponding induced dual bases for V^*. We have

$$\boxed{\begin{aligned} \mathbf{f}_i &= \sum_{j=1}^n ([\mathbf{f}_i]_E)_j \mathbf{e}_j = \sum_{j=1}^n ([\mathbf{F}]_E)^j{}_i \mathbf{e}_j \\ \psi^{\mathbf{f}_i} &= \sum_{j=1}^n ([\psi^{\mathbf{f}_i}]_{E^*})_j \phi^{\mathbf{e}_j} = \sum_{j=1}^n ([\mathbf{F}]_E^{-\top})^j{}_i \phi^{\mathbf{e}_j} \end{aligned}}$$

Proof. The first line in the boxed equation follows directly from the definition of $[\cdot]_F$. (The first line is also Theorem 2.110)). 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 $\psi^{\mathbf{f}_i} = \sum_{j=1}^n ([\mathbf{F}^*]_{E^*})^j{}_i \phi^{\mathbf{e}_j}$, where we have $[\mathbf{F}^*]_{E^*} = [\mathbf{F}]_E^{-\top}$ due to the previous theorem. □

Theorem 4.36. (Change of basis for a (p,q) tensor).

Let V be a finite-dimensional vector space with bases $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_n}\}$ be the corresponding induced dual bases for V^*.

We now derive how to change the coordinates of a (p,q) tensor in $T_{p,q}(V)$. To do so, it is enough to relate the coordinates relative to F and F^* of the (p,q) tensor

$$\mathbf{T} = \sum_{\substack{i_1,...,i_p \in \{1,...,n\} \\ j_1,...,j_q \in \{1,...,n\}}} T^{i_1 ... i_p}{}_{j_1 ... j_q} \mathbf{f}_{i_1} \otimes ... \otimes \mathbf{f}_{i_p} \otimes \psi^{\mathbf{f}_{j_1}} \otimes ... \otimes \psi^{\mathbf{f}_{j_q}}$$

to the coordinates of $\mathbf{T}$ relative to E and E^*.

To obtain this relation, we apply the previous theorem to each basis vector in $\mathbf{T}$.

$$
\begin{aligned}
&\mathbf{f}_{i_1} \otimes ... \otimes \mathbf{f}_{i_p} \otimes \psi^{\mathbf{f}_{j_1}} \otimes ... \otimes \psi^{\mathbf{f}_{j_q}} \\
&= \Big(\sum_{j_1=1}^n ([\mathbf{F}]_E)^{j_1}_{i_1} \mathbf{e}_{j_1} \Big) \otimes ... \otimes \Big(\sum_{j_p=1}^n ([\mathbf{F}]_E)^{j_p}_{i_p} \mathbf{e}_{j_p} \Big) \otimes \Big(\sum_{i_1=1}^n ([\mathbf{F}_E]^{-1})^{j_1}_{i_1} \phi^{\mathbf{e}_{i_1}} \Big) \otimes ... \otimes \Big(\sum_{i_q=1}^n ([\mathbf{F}_E]^{-1})^{j_q}_{i_q} \phi^{\mathbf{e}_{i_q}} \Big) \\
&= \sum_{j_1=1}^n ... \sum_{j_p}^n \sum_{i_1=1}^n ... \sum_{i_q=1}^n \Big(([\mathbf{F}]_E)^{j_1}_{i_1} ... ([\mathbf{F}]_E)^{j_p}_{i_p} ([\mathbf{F}_E]^{-1})^{j_1}_{i_1} ... ([\mathbf{F}_E]^{-1})^{j_q}_{i_q} \mathbf{e}_{j_1} \otimes ... \otimes \mathbf{e}_{j_p} \otimes \phi^{\mathbf{e}_{i_1}} \otimes ... \otimes \phi^{\mathbf{e}_{i_q}} \Big).
\end{aligned}
$$

After substituting this expression back into the basis sum for $\mathbf{T}$, we see that an arbitrary (p,q) tensor with an $\left({}^{i_1 ... i_p}{}_{j_1 ... j_q}\right)$ component of $T^{i_1 ... i_p}{}_{j_1 ... j_q}$ relative to F and F^* has a $\left({}^{i_1 ... i_p}{}_{j_1 ... j_q}\right)$ component relative to E and E^* of

$$
\sum_{k_1=1}^n ... \sum_{k_p}^n \sum_{\ell_1=1}^n ... \sum_{\ell_q=1}^n \Big(([\mathbf{F}]_E)^{k_1}_{\ell_1} ... ([\mathbf{F}]_E)^{k_p}_{\ell_p} ([\mathbf{F}_E]^{-1})^{k_1}_{\ell_1} ... ([\mathbf{F}_E]^{-1})^{k_q}_{\ell_q} T^{k_1 ... k_p}_{\ell_1 ... \ell_q} \Big).
$$

(It is possible to "simplify" this expression by using the fact that $([\mathbf{F}]_E)^i_j ([\mathbf{F}]_E)^{-1})^i_j = \delta^i{}_j$. 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 $\left({}^{i_1 ... i_p}{}_{j_1 ... j_q}\right)$ component of $\mathbf{T}$ relative to E and E^* as

$$
([\mathbf{F}]_E)^{k_1}_{\ell_1} ... ([\mathbf{F}]_E)^{k_p}_{\ell_p} ([\mathbf{F}_E]^{-1})^{k_1}_{\ell_1} ... ([\mathbf{F}_E]^{-1})^{k_q}_{\ell_q} T^{k_1 ... k_p}_{\ell_1 ... \ell_q} \quad \text{(Einstein notation)}.
$$

Remark 4.37. (Tensors as "multidimensional matrices" that "transform like tensors").

As was mentioned in Remark 3.41, physicists often define tensors to be "multidimensional matrices" that follow the change of basis formula of the previous theorem.

Tensor contraction

Derivation 4.38. (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 a function $V \to W$ with a linear function $W \to Z$ is itself a bilinear map $\mathcal{L}(V \to W) \times \mathcal{L}(W \to Z) \overset{\circ}{\to} \mathcal{L}(V, Z)$. (Check this as an exercise!). Also recall from Section 3.2 that every element of $\mathcal{L}(V \to W)$ and $\mathcal{L}(W \to Z)$ is a linear combination of rank-1 compositions of linear functions, i.e., of "elementary compositions". Thus, 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 $\mathbf{w} \in \mathcal{L}(K \to W)$ defined by $\mathbf{w}(c) = c\mathbf{w}$. Then, under the composition map, $(\mathbf{z} \circ \phi, \mathbf{w} \circ \phi) \in \mathcal{L}(V \to W) \times \mathcal{L}(W \to Z)$ is sent to

$$
(\mathbf{w} \circ \phi, \mathbf{z} \circ \psi) \overset{\circ}{\longrightarrow} (\mathbf{z} \circ \psi) \circ (\mathbf{w} \circ \phi) = \mathbf{z} \circ (\psi \circ \mathbf{w}) \circ \phi.
$$

Now, notice that $\phi \circ \mathbf{w}$ is the linear map $K \to K$ sending $c \mapsto c\psi(\mathbf{w})$. If we extend the above notation (that uses $\mathbf{w}$ and $\mathbf{z}$ to denote linear maps) to elements of K, and denote the linear map $K \to K$ sending $c \mapsto c\psi(\mathbf{w})$ by $\psi(\mathbf{w})$, then we have

$$
(\mathbf{w} \circ \phi, \mathbf{z} \circ \psi) \overset{\circ}{\longrightarrow} \mathbf{z} \circ \psi(\mathbf{w}) \circ \phi = \psi(\mathbf{w}) \circ \mathbf{z} \circ \phi = \mathbf{z} \circ \phi \circ \psi(\mathbf{w})
$$

(In the last three equalities, we were able to commute $\psi(\mathbf{w})$ because it is a linear map $K \to K$).

In general, the action of $\psi \in W^*$ 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 4.39. (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 $C(\mathbf{v}, \phi) = \phi(\mathbf{v})$.

In a slight generalization of the natural pairing map, we define the (k, ℓ) *contraction* on elementary (p, q) tensors, and extend with multilinearity. The (k, ℓ) contraction of an elementary tensor is defined as follows:

$$\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_p \otimes \phi^1 \otimes ... \otimes \phi^q$$
$$\overset{(k,\ell)\text{ contraction}}{\longmapsto}$$
$$C(\mathbf{v}_k, \phi^\ell)(\mathbf{v}_1 \otimes ... \otimes \cancel{\mathbf{v}_k} \otimes ... \mathbf{v}_p \otimes \phi^1 \otimes ... \otimes \cancel{\phi^\ell} \otimes ... \otimes \phi^q)$$
$$=$$
$$\phi^\ell(\mathbf{v}_k)(\mathbf{v}_1 \otimes ... \otimes \cancel{\mathbf{v}_k} \otimes ... \mathbf{v}_p \otimes \phi^1 \otimes ... \otimes \cancel{\phi^\ell} \otimes ... \otimes \phi^q).$$

Remark 4.40. (Contraction with upper and lower indices).

Vectors can only ever get contracted against dual vectors, and dual vectors can only ever get contracted against vectors. Vectors cannot get contracted against vectors, and dual vectors cannot get contracted against dual vectors.

Since the convention we laid out in 3.20 requires that lower indices (e.g. those which appear in $\mathbf{v}_k$) be used on vectors and that upper indices (e.g. those which appear in ϕ^ℓ) be used on vectors, then it follows that lower indices can only be contracted against upper indices, and that upper indices can only be contracted against lower indices.

Remark 4.41. (Composition of linear functions with tensor contraction, revisited).

The map $\circ$ which composes linear functions is itself a bilinear map $\mathcal{L}(V \to W) \times \mathcal{L}(W \to Z) \overset{\circ}{\to} \mathcal{L}(V, Z)$. Due to Theorem 3.43, we have the natural isomorphism $\mathcal{L}(V \to W) \cong W \otimes V^*$, so $\circ$ can be identified with a linear map $\widetilde{\circ} : (W \otimes V^*) \otimes (Z^* \otimes W) \to Z \otimes V^*$. Following a similar argument as was presented in Derivaton 4.38, we see that $\widetilde{\circ}$ acts on elementary tensors by $(\mathbf{w} \otimes \phi) \otimes (\mathbf{z} \circ \psi) \overset{\widetilde{\circ}}{\mapsto} C(\mathbf{w}, \psi)(\mathbf{z} \otimes \phi) = \psi(\mathbf{w})(\mathbf{z} \otimes \phi)$.

Theorem 4.42. (Coordinates of a contracted tensor).

Let V be an n-dimensional vector space, let E be a basis for V, let g be a metric tensor on V, and let E^* be the induced dual basis for V^*. Consider a (p, q) tensor $\mathbf{T} \in T_{p,q}(V)$. If the $\left({}^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}\right)$ coordinate of $\mathbf{T}$ relative to E and E^* is $T^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}$, then the $\left({}^{i_1 \ldots i_{p-1}}{}_{j_1 \ldots j_{q-1}}\right)$ component of the (k, ℓ) contraction of $\mathbf{T}$ relative to E and E^* is **come back here** $\sum_{r=1}^n T^{i_1 \ldots i_{k-1}\ r\ i_k \ldots i_{p-1}}{}_{j_1 \ldots j_{\ell-1}\ r\ j_\ell \ldots j_{q-1}}$.

Proof. Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$. Assume $\mathbf{T}$ has a $\left({}^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}\right)$ component of $T^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}$ relative to E and E^*, so

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

Using that $\phi^{\mathbf{e}_{j_\ell}}(\mathbf{e}_{i_k}) = \mathbf{e}^\flat_{j_\ell}(\mathbf{e}_{i_k}) = g(\mathbf{e}_{j_\ell}, \mathbf{e}_{i_k}) = g_{j_\ell i_k} = g_{i_k j_\ell}$, we see that the (k, ℓ) contraction of $\mathbf{T}$ is

$$\sum_{\substack{i_1 \ldots, i_p \in \{1, \ldots, n\} \\ j_1, \ldots, j_q \in \{1, \ldots, n\}}} \phi^{\mathbf{e}_{j_\ell}}(\mathbf{e}_{i_k})\, T^{i_1 \ldots i_p}{}_{j_1 \ldots j_q} \mathbf{e}_{i_1} \otimes ... \otimes \cancel{\mathbf{e}_{i_k}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \cancel{\epsilon^{j_\ell}} \otimes ... \otimes \epsilon^{j_q}$$
$$= \sum_{\substack{i_1 \ldots, i_p \in \{1, \ldots, n\} \\ j_1, \ldots, j_q \in \{1, \ldots, n\}}} g_{i_k j_\ell}\, T^{i_1 \ldots i_p}{}_{j_1 \ldots j_q} \mathbf{e}_{i_1} \otimes ... \otimes \cancel{\mathbf{e}_{i_k}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \cancel{\epsilon^{j_\ell}} \otimes ... \otimes \epsilon^{j_q}$$
$$= \sum_{i_k, j_\ell \in \{1, \ldots, n\}} g_{i_k j_\ell} \sum_{\substack{i_1 \ldots, \cancel{i_k}, \ldots, i_p \in \{1, \ldots, n\} \\ j_1, \ldots, \cancel{j_\ell}, \ldots, j_q \in \{1, \ldots, n\}}} T^{i_1 \ldots i_p}{}_{j_1 \ldots j_q} \mathbf{e}_{i_1} \otimes ... \otimes \cancel{\mathbf{e}_{i_k}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \cancel{\epsilon^{j_\ell}} \otimes ... \otimes \epsilon^{j_q}$$
$$= \sum_{i_k, j_\ell \in \{1, \ldots, n\}} g_{i_k j_\ell} \sum_{\substack{i_1 \ldots, \cancel{i_k}, \ldots, i_p \in \{1, \ldots, n\} \\ j_1, \ldots, \cancel{j_\ell}, \ldots, j_q \in \{1, \ldots, n\} \\ r \in \{1, \ldots, n\}}} T^{i_1 \ldots i_{k-1}\ r\ i_{k+1} \ldots i_p}{}_{j_1 \ldots j_{\ell-1}\ r\ j_{\ell+1} \ldots j_q} \mathbf{e}_{i_1} \otimes ... \otimes \cancel{\mathbf{e}_{i_k}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \cancel{\epsilon^{j_\ell}} \otimes ... \otimes \epsilon^{j_q}$$

So, we can see that

$$\text{The } \binom{i_1 \ldots \cancel{i_k} \ldots i_p}{j_1 \ldots \cancel{j_\ell} \ldots j_p} \text{ component of the } (k,\ell) \text{ contraction of } \mathbf{T} \text{ is } \sum_{i_k, j_\ell} g_{i_k j_\ell} \sum_r T^{i_1 \ldots i_{k-1} \; r \; i_{k+1} \ldots i_p}{}_{j_1 \ldots j_{\ell-1} \; r \; j_{\ell+1} \ldots j_q}.$$

Equivalently, after shifting the indices $i_{k+1}, ..., i_p$ down by one (this is valid because the indices i_k and j_ℓ are no longer "occupied"), we see

$$\text{The } \left({}^{i_1 \ldots i_{p-1}}{}_{j_1 \ldots j_{q-1}}\right) \text{ component of the } (k,\ell) \text{ contraction of } \mathbf{T} \text{ is } \sum_{i_k, j_\ell} g_{i_k j_\ell} \sum_r T^{i_1 \ldots i_{k-1} \; r \; i_k \ldots i_{p-1}}{}_{j_1 \ldots j_{\ell-1} \; r \; j_\ell \ldots j_{q-1}}.$$

Additionally, when E is an orthonormal basis, we have $g_{i_k j_\ell} = \delta_{i_k j_\ell}$ and thus

$$\text{The } \left({}^{i_1 \ldots i_{p-1}}{}_{j_1 \ldots j_{q-1}}\right) \text{ component of the } (k,\ell) \text{ contraction of } \mathbf{T} \text{ is } \sum_r T^{i_1 \ldots i_{k-1} \; r \; i_k \ldots i_{p-1}}{}_{j_1 \ldots j_{\ell-1} \; r \; j_\ell \ldots j_{q-1}}$$
$$\text{when the basis for } V \text{ is orthonormal.}$$

In Einstein notation, this is stated as

$$\text{The } \left({}^{i_1 \ldots i_{p-1}}{}_{j_1 \ldots j_{q-1}}\right) \text{ component of the } (k,\ell) \text{ contraction of } \mathbf{T} \text{ is } T^{i_1 \ldots i_{k-1} \; r \; i_k \ldots i_{p-1}}{}_{j_1 \ldots j_{\ell-1} \; r \; j_\ell \ldots j_{q-1}}$$
$$\text{when the basis for } V \text{ is orthonormal.}$$

□

Theorem 4.43. Taking any (k,ℓ) contraction is basis-independent.

Proof. Recall that the definition of tensor contraction was phrased entirely in terms of tensor products of vectors and dual vectors; no bases were involved. □

Theorem 4.44. (The trace is the $(1,1)$ contraction of a $(1,1)$ tensor).

Let V be a finite-dimensional vector space over a field K.

The *trace* of a square matrix $(a^i{}_j)$ with entries in K is defined to be the sum of the matrix's diagonal entries: $\text{tr}(a^i{}_j) := \sum_{i=1}^n a^i_i$. We have that $\text{tr}(a^i{}_j)$ is the $(1,1)$ contraction of the $(1,1)$ tensor corresponding to $(a^i{}_j)$.

Thus, we see the trace is a special case of tensor contraction.

Proof. Let $\mathbf{f} : K^n \to K^n$ be the linear function satisfying $[\mathbf{f}(\hat{\mathfrak{E}})]_{\hat{\mathfrak{E}}} = (a^i{}_j)$, where $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ is the standard basis for K^n. Recall from Theorem 3.15 that if V is a vector space, then there is a natural isomorphism $\mathcal{L}(V \to V) \cong V \otimes V^*$. Using $V = K^n$, we see that $(a^i{}_j)$ can be identified with the $(1,1)$ tensor $\sum_{ij} a^i{}_j \hat{\mathfrak{e}}^i \otimes \phi^{\hat{\mathfrak{e}}_j}$, where $E^* = \{\phi^{\hat{\mathfrak{e}}_1}, ..., \phi^{\hat{\mathfrak{e}}_n}\}$ is the basis for $(K^n)^*$ induced by the standard basis $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ for K^n. The $(1,1)$ contraction of this $(1,1)$ tensor is $\sum_{ij} a^i{}_j \phi^{\hat{\mathfrak{e}}_j}(\hat{\mathfrak{e}}^i) = \sum_{ij} a^i{}_j \delta^i{}_j = \sum_i a^i_i = \text{tr}(a^i{}_j)$. □

5

Exterior powers, the determinant, and orientation

This chapter focuses on *antisymmetric tensors*. We need to know about these because differential forms, when evaluated a point, are antisymmetric tensors. Antisymmetric tensors are also closely related to the determinant; we will define the determinant and explore this relationship in the second part of this chapter. We will also show how to use antisymmetric tensors- or, more specifically, elements of a *top exterior power*- to give *orientation* to a finite dimensional vector space. Lastly, we investigate *pushing forward* and *pulling back* elements of top exterior powers, as this concept will be necessary for discussing integration of differential forms in Chapter 9.

5.1 Exterior powers

Antisymmetric tensors

Definition 5.1. (Permutations on $\{1, ..., n\}$).

Let X be any set. A *permutation on* X is a bijection $X \to X$. 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* $\{1, ..., n\}$. Formally, $S_n := \{\{1, ..., n\} \to \{1, ...n\}\}$.

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 $\text{sgn} : S_n \to \{-1, 1\}$ 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 $V_1, ..., V_k$ be vector spaces over the same field. Given a permutation $\sigma \in S_k$ and a tensor $\mathbf{T} \in V_1 \otimes ... \otimes V_k$, we define the map $(\cdot)^\sigma : V_1 \otimes ... \otimes V_k \to V_1 \otimes ... \otimes V_k$ sending $\mathbf{T} \mapsto \mathbf{T}^\sigma$ by specifying its action on elementary tensors and extending linearly. We define

$$(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k)^\sigma := \mathbf{v}_{\sigma(1)} \otimes ... \otimes \mathbf{v}_{\sigma(k)}.$$

Definition 5.4. (Antisymmetric tensor).

Let $V_1, ..., V_k$ be vector spaces over the same field. We say a tensor $\mathbf{T} \in V_1 \otimes ... \otimes V_k$ is *antisymmetric* iff $\mathbf{T}^\sigma = \text{sgn}(\sigma)\mathbf{T}$.

Definition 5.5. (Algebra).

Consider a tuple $(A, K, +, \cdot, \star)$, where $(A, K, +, \cdot)$ is a vector space. We say that A is an *algebra over* K iff $\star : A \times A \to A$ is bilinear with respect to $+$ and $\cdot$.

Remark 5.6. (We want to construct an algebra of antisymmetric tensors).

Let $V_1, ..., V_k$ be vector spaces over the same field. Observe that the tuple

$$(\{\text{antisymmetric tensors in } V_1 \otimes ... \otimes V_k\}, K, \cdot, \otimes)$$

is *not* an algebra, since the tensor product $\otimes$ of antisymmetric two tensors in $V_1 \otimes ... \otimes V_k$ is *not* necessarily another antisymmetric tensor. Consider, for example, the[1] antisymmetric $(1, 0)$ 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 $\wedge : (V_1 \otimes ... \otimes V_k) \times (V_1 \otimes ... \otimes V_k) \to V_1 \otimes ... \otimes V_k$, called the *wedge product*, such that

$$(\{\text{antisymmetric tensors in } V_1 \otimes ... \otimes V_k\}, K, \cdot, \wedge)$$

is an algebra.

Constructing the wedge product

Definition 5.7. (Antisymmetrization of elements of tensor product spaces).

Let $V_1, ..., V_k$ be vector spaces over the same field. We define an *antisymmetrizing function* $\text{alt} : V_1 \otimes ... \otimes V_k \to V_1 \otimes ... \otimes V$ that converts any tensor $\mathbf{T} \in V_1 \otimes ... \otimes V_k$ into an antisymmetric tensor in $V_1 \otimes ... \otimes V_k$. We define alt on elementary tensors and extend linearly: for an elementary tensor $\mathbf{T} \in V_1 \otimes ... \otimes V_k$, we define

$$\text{alt}(\mathbf{T}) := \frac{1}{k!} \sum_{\sigma \in S_k} \text{sgn}(\sigma)\mathbf{T}^\sigma.$$

When $\mathbf{T}$ is antisymmetric, then the sum argument is $\text{sgn}(\sigma)\mathbf{T}^\sigma = \text{sgn}(\sigma)^2\mathbf{T} = \mathbf{T}$, and so the sum is $k!\mathbf{T}$. Thus, the division by $k!$ ensures that $\text{alt}(\mathbf{T}) = \mathbf{T}$ when $\mathbf{T}$ is an antisymmetric tensor.

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 when given an arbitrary tensor $\mathbf{v}_1 \otimes \mathbf{v}_2$, we can form 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

$$\text{alt}(\mathbf{T})^\pi = \Big(\sum_{\sigma \in S_k} \text{sgn}(\sigma)(\mathbf{T}^\sigma) \Big)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma)(\mathbf{T}^\sigma)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma)\mathbf{T}^{\pi \circ \sigma}.$$

Since S_k is closed under taking the inverse of a permutation (recall, a permutation in S_k is a bijection on $\{1, ..., k\}$), 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

$$\sum_{\tau \in S_k} \text{sgn}(\pi^{-1} \circ \tau)\mathbf{T}^\tau = \text{sgn}(\pi^{-1}) \sum_{\tau \in S_k} \text{sgn}(\tau)\mathbf{T}^\tau = -\text{sgn}(\pi)\text{alt}(\mathbf{T}).$$

This shows $\text{alt}(\mathbf{T})^\pi = \text{sgn}(\pi)\text{alt}(\mathbf{T})$. □

Definition 5.8. (Wedge product).

Let $V_1, ..., V_k$ be vector spaces over the same field. We define the *wedge product*
$\wedge : (V_1 \otimes ... \otimes V_k) \otimes (V_1 \otimes ... \otimes V_k) \to \text{alt}((V_1 \otimes ... \otimes V_k) \otimes (V_1 \otimes ... \otimes V_k))$ by $\mathbf{T} \wedge \mathbf{S} := \text{alt}(\mathbf{T} \otimes \mathbf{S})$.

Lemma 5.9. (Lemma for associativity of wedge product).

Let $V_1, ..., V_k$ be vector spaces over the same field, and consider $\mathbf{T}, \mathbf{S} \in V_1 \otimes ... \otimes V_k$. If $\text{alt}(\mathbf{T}) = \mathbf{0}$, then $\mathbf{T} \wedge \mathbf{S} = \mathbf{0} = \mathbf{S} \wedge \mathbf{T}$.

[1] $\mathbf{v}_1$ is antisymmetric because its only permutation $\mathbf{v}_1^\sigma$ 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.

Proof. (This proof requires some abstract algebra. Understanding this proof is not really necessary to understand exterior powers, so you can take this theorem as an axiom if you want).

Assume $\text{alt}(\mathbf{T}) = \mathbf{0}$. Let $T = \mathbf{v}_{i_1} \otimes ... \otimes \mathbf{v}_{i_k}$ and $\mathbf{S} = \mathbf{v}_{j_1} \otimes ... \otimes \mathbf{v}_{j_k}$. 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 $j_1, ..., j_k$, 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

$$\begin{aligned}\text{alt}(\mathbf{T} \otimes \mathbf{S}) &= \sum_{[\pi]_\tau \sigma \in \{\text{right cosets}\}} \text{sgn}([\pi]_\tau \sigma)(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau \sigma} = \sum_{\sigma \in S_{2k}} \sum_{[\pi]_\tau \in H} \text{sgn}([\pi]_\tau \sigma)(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau \sigma} \\ &= \sum_{\sigma \in S_{2k}} \Big(\sum_{[\pi]_\tau \in H} \text{sgn}([\pi]_\tau)(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} \Big)^\sigma.\end{aligned}$$

Since $[\pi]_\tau \in H$, where H is the subgroup of S_{2k} whose elements fix all of $j_1, ..., j_\ell$, then $(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} = \mathbf{T}^{[\pi]_\tau} \otimes \mathbf{S}$. With this, the innermost sum becomes

$$\sum_{[\pi]_\tau \in H} \text{sgn}([\pi]_\tau)(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} = \sum_{[\pi]_\tau \in H} \text{sgn}([\pi]_\tau)\mathbf{T}^{[\pi]_\tau} \otimes \mathbf{S} = \Big(\sum_{[\pi]_\tau \in H} \text{sgn}([\pi]_\tau)\mathbf{T}^{[\pi]_\tau} \Big) \otimes \mathbf{S}.$$

Now define $\pi \in S_k$ by $\pi = \tau^{-1}[\pi]_\tau \tau$, where $\tau = (i_1, ..., i_k)$ (i.e. $\tau(i) = j_i$). Then the above is

$$\Big(\sum_{\pi S_k} \text{sgn}(\pi)\mathbf{T}^\pi \Big) \otimes \mathbf{S} = \text{alt}(\mathbf{T}) \otimes \mathbf{S} = \mathbf{0} \otimes \mathbf{S} = \mathbf{0}.$$

The last equality follows by the seeming-multilinearity of $\otimes$. □

Theorem 5.10. (Wedge product is associative).

Let $V_1, ..., V_k$ be vector spaces over the same field. Then for all $\mathbf{T}, \mathbf{S}, \mathbf{R} \in V_1 \otimes ... \otimes V_k$, 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 5.11. (Properties of the wedge product).

Let $V_1, ..., V_k$ be vector space over a field $K \neq \mathbb{Z}/2\mathbb{Z}$, and consider the tensor product space $V_1 \otimes ... \otimes V_k$. 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 $\mathbf{T}, \mathbf{S}, \mathbf{R} \in V_1 \otimes ... \otimes V_k$.
   1.2. $\mathbf{T} \wedge (\mathbf{S} + \mathbf{R}) = \mathbf{T} \wedge \mathbf{S} + \mathbf{T} \wedge \mathbf{R}$ for all $\mathbf{T}, \mathbf{S}, \mathbf{R} \in V_1 \otimes ... \otimes V_k$.
   1.3. $(c\mathbf{T}) \wedge \mathbf{S} = c(\mathbf{T} \wedge \mathbf{S}) = \mathbf{T} \wedge (c\mathbf{S})$ for all $\mathbf{T}, \mathbf{S} \in V_1 \otimes ... \otimes V_k$ 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 $\mathbf{T}, \mathbf{S} \in V_1 \otimes ... \otimes V_k$.
3. $\wedge$ looks as it it is an antisymmetric map: $\mathbf{T} \wedge \mathbf{S} = -\mathbf{S} \wedge \mathbf{T}$ for all $\mathbf{T}, \mathbf{S} \in V_1 \otimes ... \otimes V_k$.
4. $\wedge$ is *skew-commutative*: if $\mathbf{T} \in V_1 \otimes ... \otimes V_k$ and $\mathbf{S} \in V_1 \otimes ... \otimes V_\ell$, then $\mathbf{S} \wedge \mathbf{T} = (-1)^{k+\ell}(\mathbf{T} \wedge \mathbf{S})$.
5. $\mathbf{T} \wedge \mathbf{T} = \mathbf{0}$ for all $\mathbf{T} \in V_1 \otimes ... \otimes V_k$.

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 5.10. 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). □

In Remark 5.6, we said that we wanted to find a function $\wedge : (V_1 \otimes ... \otimes V_k) \times (V_1 \otimes ... \otimes V_k) \to V_1 \otimes ... \otimes V_k$ for which the tuple $(\{\text{antisymmetric tensors in } V_1 \otimes ... \otimes V_k\}, K, \cdot, \wedge)$ is an algebra. We have done so by constructing the wedge product $\wedge$, and now formalize this result with the next definitions.

Definition 5.12. (Wedge product space).

Let $V_1, ..., V_k$ be vector spaces over a field K. We define the *wedge product space* $V_1 \wedge ... \wedge V_k$ to be the algebra

$$V_1 \wedge ... \wedge V_k := (\text{alt}(V_1 \otimes ... \otimes V_k), K, +, \cdot, \wedge).$$

Theorem 5.13. (Basis and dimension for wedge product spaces).

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

$$\{\mathbf{e}_{1i_1} \otimes ... \otimes \mathbf{e}_{ki_k} \mid i_k \in \{1, ..., n_k\}\}.$$

Proof. See the proof of Theorem 3.8. □

Theorem 5.14. (Exterior powers).

Let V be a vector space. We define the *kth exterior power* $\Lambda^k(V)$ of V to be the wedge product space $\Lambda^k(V) := V^{\wedge k}$.

Theorem 5.15. (Basis and dimension of exterior powers).

Let V be an n-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. Then $\Lambda^k(V)$ is a $\binom{n}{k}$ dimensional vector space with basis

$$\{\mathbf{e}_{i_1} \wedge ... \wedge \mathbf{e}_{i_k} \mid k \in \{1, ..., n\} \text{ and } (i_1, ..., i_k) \text{ is strictly increasing}\}.$$

Note, $(i_1, ..., i_k)$ must be strictly increasing because $\mathbf{v} \wedge \mathbf{v} = \mathbf{0}$.

Proof. To show that this set spans $\Lambda^k(V)$, use the seeming-multilinearity of $\wedge$ as was done for $\otimes$ in the proof of Theorem 3.8.

For linear independence, note that if the sequence $i_1, ..., i_k$ were not strictly increasing, then we would have a $\mathbf{0}$ in the claimed basis, which would make our claimed basis a linearly dependent set. Since there is no $\mathbf{0}$ in the claimed basis, we can follow the proof of Theorem 3.8. □

Definition 5.16. (Alternating function).

Let $V_1, ..., V_k, W$ be vector spaces over the same field. We say a function $\mathbf{f} : V_1 \times ... \times V_k \to W$ is a *k-alternating function* iff for all $\sigma \in S_n$, we have $\mathbf{f}(\mathbf{v}_{\sigma(1)}, ..., \mathbf{v}_{\sigma(k)}) = \text{sgn}(\sigma)\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_k)$. Equivalently, $\mathbf{f}$ is k-alternating iff $\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_i, ..., \mathbf{v}_j, ..., \mathbf{v}_k) = -\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_j, ..., \mathbf{v}_i, ..., \mathbf{v}_k)$ for all $i \in \{1, ..., k\}$.

When k is clear from the context, k-linear functions are called *alternating functions.*

Remark 5.17. ("Antisymmetric" vs. "alternating").

Many authors refer to the "antisymmetric tensors" of Definition 5.4 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. (We have and will not discuss field characteristics). 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 5.18. (Vector space of alternating functions).

If $V_1, ..., V_k, W$ are vector spaces over a field K, then we use $(\text{alt}\mathcal{L})(V_1 \times ... \times V_k \to W)$ to denote the vector space over K formed by the set of k-alternating functions $V_1 \times ... \times V_k \to W$ under the operations of function addition and function scaling.

Theorem 5.19. (Universal property for exterior powers).

Let V_1, V_2, W be vector spaces over the same field, and let $\mathbf{f} : V_1 \times V_2 \to W$ be an alternating bilinear function. Then there exists a linear function $\mathbf{h} : V_1 \wedge V_2 \to W$ with $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ where $\mathbf{h}$ uniquely depends on $\mathbf{f}$, and where $\mathbf{g} : V_1 \times V_2 \to V_1 \otimes V_2$.

In Remark 5.17, 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 5.20. (Fundamental natural isomorphisms for exterior powers).

Theorem 3.43 stated that there are natural isomorphisms

$$\begin{aligned}
\mathcal{L}(V_1 \times ... \times V_k \to W) &\cong \mathcal{L}(V_1 \otimes ... \otimes V_k \to W) \\
\mathcal{L}(V \to W) &\cong W \otimes V^* \\
(V \otimes W)^* &\cong V^* \otimes W^* \\
T_{p,q}(V) &\cong T_{q,p}(V^*)
\end{aligned}$$

Analogously, there are natural isomorphisms

$$\boxed{\begin{aligned}
(\text{alt}\mathcal{L})(V_1 \times ... \times V_k \to W) &\cong (\text{alt}\mathcal{L})(V_1 \wedge ... \wedge V_k \to W) \\
(\text{alt}\mathcal{L})(V \to W) &\cong W \wedge V^* \\
(V \wedge W)^* &\cong V^* \wedge W^* \\
\Lambda^k(V)^* &\cong \Lambda^k(V^*)
\end{aligned}}$$

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 coorresponding 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 $\mathcal{L}(V \to W) \cong W \otimes V^*$. That is, take an element $\mathbf{f} \in (\text{alt}\mathcal{L})(V \to W)$, 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 3.43; 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.

$\square$

Pushforward and pullback

[segway]

Definition 5.21. (Pushforwards and pullbacks).

Let V and W be finite-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$ and its dual $\mathbf{f}^* : W^* \to V^*$. We define the following *pushforward* and *pullback* maps on $T_0^k(V) = V^{\otimes k}$ and $T_{0,k}(W) = (W^*)^{\otimes k}$, respectively, extending each with the seeming-multilinearity of $\otimes$:

$$\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k \in T_0^k(V) \overset{\otimes_{k,0}\mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \otimes ... \otimes \mathbf{f}(\mathbf{v}_k) \in T_0^k(W) \quad \text{(“pushforward on } T_0^k(V)\text{”)}$$

$$\psi_1 \otimes ... \otimes \psi_k \in T_{0,k}(W) \overset{\otimes_{0,k}\mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi_1) \otimes ... \otimes \mathbf{f}^*(\psi_k) \in T_{0,k}(V) \quad \text{(“pullback on } T_{0,k}(W)\text{”).}$$

We also define the following *pushforward* and *pullback* maps on $\Lambda^k(V) = V^{\wedge k}$ and $\Lambda^k(W^*) = (W^*)^{\wedge k}$, respectively, extending each with the seeming-multilinearity and antisymmetry of $\wedge$, for $k \leq n$:

$$\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k \in \Lambda^k(V) \overset{\Lambda^k\mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_k) \in \Lambda^k(W) \quad \text{(“pushforward on } \Lambda^k(V)\text{”)}$$

$$\psi^1 \wedge ... \wedge \psi^k \in \Lambda^k(W^*) \overset{\Lambda^k\mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1) \wedge ... \wedge \mathbf{f}^*(\psi^k) \in \Lambda^k(V^*) \quad \text{(“pullback on } \Lambda^k(W^*)\text{”).}$$

Remark 5.22. (Star notation for pushforward and pullback).

The above notation of using $\otimes_{0,k}\mathbf{f} : V \to W$ and $\Lambda^k\mathbf{f}$ for the pushforwards and using $\otimes_{k,0}\mathbf{f}^* : W^* \to V^*$ and $\Lambda^k\mathbf{f}^*$ for the pullbacks is nonstandard. It is more common to denote the pullbacks by $\mathbf{f}^* : T_{0,k}(W) \to T_{0,k}(V)$ and $\mathbf{f}^* : \Lambda^k(W) \to \Lambda^k(V)$. Somewhat less commonly, pushforwards are denoted by $\mathbf{f}_* : T_0^k(V) \to T_0^k(W)$ and $\mathbf{f}_* : \Lambda^k(V) \to \Lambda^k(W)$.

Once the above definitions of pushforward and pullback are understood, this “star notation” can be useful. But it can be difficult to understand what the various pushforwards and pullbacks are if this notation is used from the start, due to the potential for confusing the dual $\mathbf{f}^* : W^* \to V^*$ with either of the pullbacks $\mathbf{f}^* : T_{0,k}(W) \to T_{0,k}(V), \mathbf{f}^* : \Lambda^k(W) \to \Lambda^k(V)$.

We will use the star notation after we define a pullback of a differential form in Chapter 9. Until then, we do not use the star notation.

5.2 The determinant

Definition 5.23. (The determinant).

Let K be a field, and let $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ be the standard basis of K^n. We want to define a function $(K^n)^{\times n} \to K$ which, given $\mathbf{c}_1, ..., \mathbf{c}_n \in K^n$, returns the n-dimensional volume of the parallelapiped spanned by $\mathbf{c}_1, ..., \mathbf{c}_n$. We will denote this function by $\det : (K^n)^{\times n} \to K$. We require that det satisfy the following axioms:

1. $\det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$, 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. $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n) = 0$ when $\mathbf{c}_i = \mathbf{c}_j$ for all $\mathbf{c}_k \in K$, $k \in \{1, ..., n\}$. 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 $\mathbf{c}_1, ..., \mathbf{c}_j, ..., \mathbf{c}_i, ..., \mathbf{c}_n$ is the negation of the volume of the parallelapiped spanned by $\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n$.

The fact that the third axiom is logically equivalent to alternatingness also gives us a concise characterization of the determinant: $\det : (K^n)^{\times n} \to K$, when $K \neq \mathbb{Z}/2\mathbb{Z}$, is the unique multilinear alternating function satisfying $\det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$. **pretty sure we can't claim this until seeing permutation formula for** det

Definition 5.24. (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 5.25. (Consequent properties of the determinant).

4. det is an alternating function when $K \neq \mathbb{Z}/2\mathbb{Z}$.
5. det is invariant under linearly combining input vectors into a different input vector. That is, $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_n) = \det(\mathbf{c}_1, ..., \mathbf{c}_i + \sum_{j=1, j\neq i}^n d_j \mathbf{c}_j, ..., \mathbf{c}_n)$ for all $i \in \{1, ..., n\}$.
6. $\det(\mathbf{c}_1, ..., \mathbf{c}_n) = 0$ iff $\{\mathbf{c}_1, ..., \mathbf{c}_n\}$ is a linearly dependent set.

Proof.

5. Using the axiom $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n) = 0$ when $\mathbf{c}_i = \mathbf{c}_j$ together with the multilinearity of the determinant, we have

$$\begin{aligned}\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_n) &= \det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_n) + \sum_{j=1, j\neq i}^n \Big(d_j \det(\mathbf{c}_1, ..., \mathbf{c}_j, ..., \mathbf{c}_j, ..., \mathbf{c}_n)\Big) \\ &= \det\Big(\mathbf{c}_1, ..., \mathbf{c}_i + \sum_{j=1, j\neq i}^n (d_j \mathbf{c}_j), ..., \mathbf{c}_n\Big).\end{aligned}$$

6.
   ($\det(\mathbf{c}_1, ..., \mathbf{c}_n) = 0 \implies \{\mathbf{c}_1, ..., \mathbf{c}_n\}$ 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.
   ($\det(\mathbf{c}_1, ..., \mathbf{c}_n) = 0 \impliedby \{\mathbf{c}_1, ..., \mathbf{c}_n\}$ 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 $\det(\mathbf{c}_1, ..., \mathbf{c}_n) = 0$ for *all* $\mathbf{c}_1, ..., \mathbf{c}_n \in K^n$. This contradicts the first axiom that specifies $\det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$.

$\square$

Derivation 5.26. (Permutation formula for the determinant).

We now derive *permutation formula* for the determinant. The formula we obtain shows that the function det specified in Definition 5.23 exists; since this formula is derived from the axioms of the determinant, it is also a unique formula for the determinant.

Consider vectors $\mathbf{c}_1, ..., \mathbf{c}_n \in K^n$, and[2] set $(a^i_j) := ([\mathbf{c}_i]_{\hat{\mathfrak{E}}})^j$. Then...

$$
\begin{aligned}
\det((a^i_j)) = \det(\mathbf{c}_1, ..., \mathbf{c}_n) &= \det\Big(\sum_{i_1=1}^n a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., \sum_{i_n=1}^n a^n_{i_n}\hat{\mathfrak{e}}_{i_n}\Big) \\
&= \sum_{i_1=1}^n \det\Big(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., \sum_{i_n=1}^n a^n_{i_n}\hat{\mathfrak{e}}_{i_n}\Big) \\
&\vdots \\
&= \sum_{i_1=1}^n ... \sum_{i_n=1}^n \det(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., a^n_{i_n}\hat{\mathfrak{e}}_{i_n}) \\
&= \sum_{i_1=1}^n ... \sum_{i_n=1}^n \det(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., a^n_{i_n}\hat{\mathfrak{e}}_{i_n}), \text{ where } i_1, ..., i_n \text{ are distinct from each other} \\
&= \sum_{\sigma \in S_n} \det(a^1_{\sigma(1)}\hat{\mathfrak{e}}_{\sigma(1)}, ..., a^n_{\sigma(n)}\hat{\mathfrak{e}}_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} a^1_{\sigma(1)} ... a^n_{\sigma(n)} \det(\hat{\mathfrak{e}}_{\sigma(1)}, ..., \hat{\mathfrak{e}}_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} a^1_{\sigma(1)} ... a^n_{\sigma(n)} \mathrm{sgn}(\sigma) \det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) \\
&= \sum_{\sigma \in S_n} a^1_{\sigma(1)} ... a^n_{\sigma(n)} \mathrm{sgn}(\sigma)
\end{aligned}
$$

Therefore, we have

$$
\boxed{\det(\mathbf{c}_1, ..., \mathbf{c}_n) = \sum_{\sigma \in S_n} a^1_{\sigma(1)} ... a^n_{\sigma(n)} \mathrm{sgn}(\sigma)}
$$

In this derivation, we have mostly used the multilinearity of the determinant. Though, the expression labeled with "where $i_1, ..., i_n$ are distinct from each other" results from the previous line due to the third axiom of the determinant, $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n) = 0$ 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^i_j) 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 $i_1, ..., i_n$ 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 $i_1, ..., i_n$ 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 $\mathrm{sgn}(\sigma)$.

Theorem 5.27. (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

[2]There is no hidden meaning behind the upper and lower indices on a^i_j 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 $(1, 1)$ tensor rather those of a $(2, 0)$ or $(0, 2)$ tensor.

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 5.28. (Laplace expansion for the determinant).

Consider an $n \times n$ matrix $\mathbf{A} = (a^i_j)$, and let $\mathbf{A}^i_j$ denote the so-called *ij minor matrix* obtained by erasing the ith row and jth column of $\mathbf{A}$. We have

$$\det(\mathbf{A}) = \sum_{i=1}^n a^i_j \det(\mathbf{A}^i_j) \text{ for all } i \in \{1, ..., n\}$$
$$\det(\mathbf{A}) = \sum_{j=1}^n a^i_j \det(\mathbf{A}^i_j) \text{ for all } j \in \{1, ..., n\}$$

The first equation is called the *Laplace expansion for the determinant along the ith row*, and the second equation is called the *Laplace expansion for the determinant along the jth 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^i_j. Let $\mathbf{B}$ denote the shuffled diagonal matrix that corresponds to one of these terms. We can view $\det(\mathbf{B})$ as $\det(\mathbf{B}) = \pm a^i_j \det(\mathbf{B}^i_j)$, where $\mathbf{B}^i_j$ is the determinant of the matrix obtained by removing the ith column and jth row from $\mathbf{B}$. The $\pm$ sign is a result of the fact that the matrices $\mathbf{B}$ and $\mathbf{B}^i_j$ 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 $\mathbf{B}^i_j$ differ.

As a first step, note that the difference in the inversion count between $\mathbf{B}$ and $\mathbf{B}^i_j$ is the number of inversions that involve a^i_j. Thus, our problem reduces to determining an expression for the number of inversions that involve a^i_j. So, divide the matrix $\mathbf{B}$ into quadrants that are centered on a^i_j. 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^i_j 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 $\mathbf{B}^i_j$, 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,

$$\begin{aligned} a^i_j \det(\mathbf{B}^i_j) &= a^i_j \sum_{\pi \in S_n} a_1^{\pi(1)} ... \cancel{a_j^{\pi(i)}} ..., a_n^{\pi(n)} \text{sgn}(\pi) \\ &= a^i_j \sum_{\sigma \in S_n} a_1^{\pi(1)} ... \cancel{a_j^{\pi(i)}} ..., a_n^{\pi(n)} (-1)^{i+j} \text{sgn}(\sigma) \\ &= (-1)^{i+j} a^i_j \det(\mathbf{B}) \end{aligned}$$

Thus $a^i_j \det(\mathbf{B}^i_j) = (-1)^{i+j} a^i_j \det(\mathbf{B}) \iff \det(\mathbf{B}) = (-1)^{i+j} a^i_j \det(\mathbf{B}^i_j)$. Now sum all of the $\mathbf{B}$'s (the diagonal shuffled matrices) to get $\det(\mathbf{A}) = \sum_{j=1}^n a^i_j \det(\mathbf{A}^i_j)$. □

Definition 5.29. (Determinant of a linear function).

Let V and W be finite-dimensional vector spaces *of the same dimension*, and let E and F be be bases for V and W. We define the *determinant of a linear function* $\mathbf{f} : V \to W$ to be the determinant of the matrix of $\mathbf{f}$ relative to E and F, $\det(\mathbf{f}) := \det([\mathbf{f}(E)]_F)$.

Remark 5.30. We have not yet shown that the determinant of a linear function $V \to V$ 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 5.31. (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 $\mathbf{f} : V \to W$. Consider also the dual $\mathbf{f} : W^* \to V^*$ (recall Definition 4.24). Then $\det(\mathbf{f}^*) = \det(\mathbf{f})$.

Proof. Recall from[3] condition (3) of Definition 4.26 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 5.27,) we have $\det(\mathbf{f}) = \det(\mathbf{A}) = \det(\mathbf{A}^\top) = \det(\mathbf{f}^*)$. □

Theorem 5.32. (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 $\mathbf{f} : V \to W$ 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 $\Lambda^n\mathbf{f} : \Lambda^n(V) \to \Lambda^n(W)$ on the top exterior power, which (recall Definition **??**) is defined on elementary tensors by $\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n \mapsto \mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_n)$ 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 $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be a basis of V, and consider the action of the pushforward on the basis vectors of E,

$$\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n).$$

Because $\Lambda^n(\mathbf{f})$ is multilinear and alternating, the wedge product is analogous to the determinant $\det(\mathbf{f}(\mathbf{e}_1), ..., \mathbf{f}(\mathbf{e}_n)) = \det([\mathbf{f}(E)]_F) = \det(\mathbf{f})$. So, set $(a^i_j) = [\mathbf{f}(E)]_F$, 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

$$\begin{aligned}\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n) &= \sum_{\sigma \in S_n} \Big(a^1_{\sigma(1)} ... a^n_{\sigma(n)} \operatorname{sgn}(\sigma) \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n \Big) = \Big(\sum_{\sigma \in S_n} a^1_{\sigma(1)} ... a^n_{\sigma(n)} \operatorname{sgn}(\sigma) \Big) \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n \\ &= \det([\mathbf{f}(E)]_F)(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n) = \det(\mathbf{f})(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n).\end{aligned}$$

So, we have the statement on the basis E

$$\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n) = \det(\mathbf{f})(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n).$$

Using the seeming-multilinearity of $\wedge$, we can extend this fact to apply to any list of vectors in V:

$$\boxed{\mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_n) = \det(\mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n) \text{ for all } \mathbf{v}_1, ..., \mathbf{v}_n \in V}$$

We can explicitly involve the pullback $\Lambda^n\mathbf{f}$ and write the above as

$$(\Lambda^n\mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n) = \det(\mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n) \text{ for all } \mathbf{v}_1, ..., \mathbf{v}_n \in V.$$

Thus, $\Lambda^n(\mathbf{f})$ is multiplication by $\det(\mathbf{f})$.

Remark 5.33. (Determinant is well-defined).

The above theorem provides a unique characterization of $\det(\mathbf{f})$ that does not involve any bases. This shows that $\det(\mathbf{f})$ is basis-independent and thus that the determinant of a function is well-defined.

Theorem 5.34. (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 $\mathbf{f} : V \to W$. Consider additionally the dual $\mathbf{f}^* : W^* \to V^*$ (recall Definition 4.24). 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

$$(\Lambda^n\mathbf{f}^*)(\mathbf{w}_1 \wedge ... \wedge \mathbf{w}_n) = \det(\mathbf{f})(\mathbf{w}_1 \wedge ... \wedge \mathbf{w}_n) \text{ for all } \mathbf{w}_1, ..., \mathbf{w}_n \in V.$$

That is,

$$\boxed{\mathbf{f}^*(\mathbf{w}_1) \wedge ... \wedge \mathbf{f}^*(\mathbf{w}_n) = \det(\mathbf{f})(\mathbf{w}_1 \wedge ... \wedge \mathbf{w}_n)}$$

[3]Technically, the equivalent conditions of the definition we reference only apply to linear functions $V \to V$. 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 $\widetilde{\mathbf{f}} : V \to V$ 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 5.35. (Product rule for determinants). Let V, W and Z be finite-dimensional vector spaces of the same dimension, and consider linear functions $\mathbf{f} : V \to W$ and $\mathbf{g} : W \to Z$. 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

$$(\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1) \wedge ... \wedge (\mathbf{g} \circ \mathbf{f})(\mathbf{v}_n) = \det(\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n)$$

for all $\mathbf{v}_1, ..., \mathbf{v}_n \in V$.

Notice that the left side is

$$\mathbf{g}(\mathbf{f}(\mathbf{v}_1)) \wedge ... \wedge \mathbf{g}(\mathbf{f}(\mathbf{v}_n)) = \det(\mathbf{g})(\mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_n)) = \det(\mathbf{g}) \det(\mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n).$$

Thus

$$\det(\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n) = \det(\mathbf{g}) \det(\mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n).$$

This is a statement on the basis vector $\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n$ 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 $\det(\mathbf{g} \circ \mathbf{f}) - \det(\mathbf{g}) \det(\mathbf{f}) = 0$. □

Theorem 5.36. (Determinant of an inverse function).

Let V and W be finite-dimensional vector spaces of the same dimension, and consider a linear function $\mathbf{f} : V \to W$. 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 $\det(\mathbf{f}) \det(\mathbf{f}^{-1}) = 1$. □

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 define an *orientation on an inner product space* to be a choice of an ordered orthonormal basis. (We heavily rely on inner product spaces for their inner-product-induced orthonormality). This definition of orientation will only allow us to check the orientation of permutations of the chosen orthonormal basis, however. In order to give orientation to arbitrary ordered bases, we introduce rotations in n-dimensions, so that an arbitrary ordered basis can be given the orientation of a "close-by" permuted ordered basis.

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 5.37. (Ordered basis).

We will formalize the notion of orientation relying on the concept of an ordered basis. An *ordered basis* 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 5.38. (Permutation acting on an ordered basis).

Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be an ordered basis for some finite-dimensional vector space. Given a permutation $\sigma \in S_n$, we define $E^\sigma := \{\mathbf{e}_{\sigma(1)}, ..., \mathbf{e}_{\sigma(n)}\}$.

We now discover a consequence of imposing that the bases under consideration be ordered.

Derivation 5.39. (Intuition for the antisymmetry of ordered bases).

Consider the plane $\mathbb{R}^2$, and consider also two permutations of the standard ordered basis $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ for $\mathbb{R}^2$: $\{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ and $\{\hat{\mathfrak{e}}_1, -\hat{\mathfrak{e}}_2\}$. (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 $\{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ and $\{\hat{\mathfrak{e}}_2, \hat{\mathfrak{e}}_1\}$. Finally, consider the ordered bases $\{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ and $\{-\hat{\mathfrak{e}}_2, \hat{\mathfrak{e}}_1\}$ 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.*

We would now like to work towards a more precise statement of the antisymmetry of ordered bases. The notion of *rotational equivalence* is what will facilitate this formalization. Before we define rotational equivalence, however, we must formalize what a "rotation" is. The following definition accomplishes this.

Definition 5.40. (2-rotation).

Let V be a 2-dimensional inner product space, and let $\widehat{U}$ be an orthonormal ordered basis for V. A *2-rotation on the 2-dimensional inner product space V* is a linear function $\mathbf{R}_\theta : V \to V$ whose matrix relative to $\widehat{U}$ and $\widehat{U}$ 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 $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ is an orthonormal ordered basis for V. An *extension (to V) of a 2-rotation on the oriented*[4] *subspace* $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) \subseteq V$ *is a linear function $V \to V$ for which there exist i, j such that*:

[4] When we say "oriented subspace", we mean that the orientation of $\text{span}(\mathbf{u}_i, \mathbf{u}_j)$ is given by the ordered basis $\{\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j\}$.

- the map $(\mathbf{v}_i, \mathbf{v}_j) \mapsto \mathbf{R}_\theta(\mathbf{v}_1, ..., \mathbf{v}_i, ..., \mathbf{v}_j, ..., \mathbf{v}_n)$ is a 2-rotation on $\text{span}(\mathbf{v}_i, \mathbf{v}_j)$
- the map $(\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \cancel{\mathbf{v}_j}, ..., \mathbf{v}_n) \mapsto \mathbf{R}_\theta(\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \cancel{\mathbf{v}_j}, ..., \mathbf{v}_n)$ is the identity on $\text{span}(\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \cancel{\mathbf{v}_j}, ..., \mathbf{v}_n)$.

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 $i, j \in \{1, ..., n\}$,

$$\begin{array}{c} \begin{array}{cccccc} & & i\text{th column} & & j\text{th column} & \end{array} \\ \begin{pmatrix} 1 & \dots & \cos(\theta) & 0 & -\sin(\theta) & 0 \\ 0 & & 0 & \vdots & 0 & \vdots \\ 0 & & \vdots & 1 & \vdots & \vdots \\ \vdots & & 0 & \vdots & 0 & \vdots \\ 0 & \dots & \sin(\theta) & 0 & \cos(\theta) & 1 \end{pmatrix} \end{array}.$$

(The columns other than the ith and jth columns are the columns of the $n \times n$ identity matrix).

If $\mathbf{R}_\theta$ 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*, or a *2-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 5.41. (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 $\mathbf{R}_\theta$ for which $\widetilde{U} = \mathbf{R}_\theta(\widehat{U})$. (Recall Definition 2.84 for the meaning of $\mathbf{R}_\theta(\widehat{U})$).

Theorem 5.42. (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 $\mathbf{R}_\theta$ relative to $\widehat{U}$ and $\widehat{U}$ is

$$\pm \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Computing $\mathbf{R}_\theta(\widehat{U}) = \mathbf{R}_\theta(\{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\})$ for $\theta = -\frac{\pi}{2}, \frac{\pi}{2}$, we see that the following ordered bases are rotationally equivalent:

$$\begin{aligned} \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\} &\sim \{-\hat{\mathbf{u}}_2, \hat{\mathbf{u}}_1\} \sim \{\hat{\mathbf{u}}_2, -\hat{\mathbf{u}}_1\} \\ \{\hat{\mathbf{u}}_2, \hat{\mathbf{u}}_1\} &\sim \{-\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\} \sim \{\hat{\mathbf{u}}_1, -\hat{\mathbf{u}}_2\}. \end{aligned}$$

This is what we noticed in the informal discussion of Derivation 5.39.

We now generalize the equivalence under rotation and the antisymmetry of ordered bases to n-dimensional inner product spaces.

Definition 5.43. (Equivalence under rotation for n-dimensional inner product spaces).

We define orthonormal ordered bases $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ and $\widetilde{U} = \{\tilde{\mathbf{u}}_1, ..., \tilde{\mathbf{u}}_n\}$ 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 $\mathbf{R} = \mathbf{R}_k \circ ... \circ \mathbf{R}_1$, 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.

Theorem 5.44. (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 $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ of V, we have

$$\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}, ..., \hat{\mathbf{u}}_n\} \sim -\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_{i+1}, \hat{\mathbf{u}}_i, ..., \hat{\mathbf{u}}_n\}.$$

Equivalently, for any orthonormal ordered basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ of V and any permutation $\sigma \in S_n$,

$$\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\} \sim \mathrm{sgn}(\sigma)\{\hat{\mathbf{u}}_{\sigma(1)}, ..., \hat{\mathbf{u}}_{\sigma(n)}\}.$$

Definition 5.45. (Orientation of permuted ordered bases).

Let V be an n-dimensional inner product space, and fix an orthonormal ordered basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ for V. Suppose $\widehat{U}^\sigma = \{\hat{\mathbf{u}}_{\sigma(1)}, ..., \hat{\mathbf{u}}_{\sigma(n)}\}$, where $\sigma \in S_n$, is a permutation of $\widehat{U}$ that is not rotationally equivalent to $\widehat{U}$ (so choose any σ with $\mathrm{sgn}(\sigma) = -1$). By the antisymmetry of ordered bases, any other permutation $\widehat{U}^\pi = \{\hat{\mathbf{u}}_{\pi(1)}, ..., \hat{\mathbf{u}}_{\pi(n)}\}$, $\pi \in S_n$, of $\widehat{U}$ is rotationally equivalent either to $\widehat{U}$ or to $\widehat{U}^\sigma$. In other words, there are only two equivalence classes[5] 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 $\widehat{U}$ for V. When V is given the orientation $\widehat{U}$, then the *orientation of a permutation of* $\widehat{U}$ *(relative to* $\widehat{U}$*)* is said to be *positive* iff that permutation of $\widehat{U}$ is rotationally equivalent to $\widehat{U}$, and is said to be *negative* otherwise. Per the previous paragraph, every permutation of $\widehat{U}$ is either positively oriented or negatively oriented relative to $\widehat{U}$.

Remark 5.46. (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 $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2, \hat{\mathfrak{e}}_3\}$ such that $\hat{\mathfrak{E}}$ can be rotated so that $\hat{\mathfrak{e}}_1$ points out of the page, $\hat{\mathfrak{e}}_2$ points to the right, and $\hat{\mathfrak{e}}_3$ points upwards on the page. In this visual convention, the direction of each basis vector corresponds to its position in $\hat{\mathfrak{E}}$. Counterclockwise rotational configurations are also called "right handed coordinate systems".

A *clockwise rotational configuration* then corresponds to the ordered bases which are not rotationally equivalent to $\hat{\mathfrak{E}}$. One such ordered basis, $\{-\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2, \hat{\mathfrak{e}}_3\}$, can be depicted using the visual convention just established by drawing $\hat{\mathfrak{e}}_1$ as pointing into the page (i.e. $-\hat{\mathfrak{e}}_1$ points out of the page), $\hat{\mathfrak{e}}_2$ as pointing to the right, and $\hat{\mathfrak{e}}_3$ as pointing upwards on the page. Clockwise rotational configurations are also called "left handed coordinate systems".

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 5.47. (n-rotation).

An *n-rotation on an n-dimensional inner product space* is a composition of extensions of 2-rotations on that space.

Theorem 5.48. Every n-rotation is an orthogonal linear function.

Proof. Every n-rotation is a composition of 2-rotations, which are orthogonal linear functions. □

Lemma 5.49. (In three dimensions, every 2-dimensional orthonormal ordered basis is "rotationally close" to a permuted orthonormal basis).

Let V be a inner product space with $\dim(V) \geq 3$, and consider 3-dimensional subspaces $\widetilde{W}$ and W of V. If $\widetilde{U} = \{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2\}$ and $\widehat{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ are orthonormal ordered bases of $\widetilde{W}$ and W, respectively, then there exists a 3-rotation which takes $\widetilde{U}$ to $\widehat{U}^\sigma$ 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 $\mathbf{R}_\gamma$ 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 $\mathbf{R}(\widetilde{U}) = \widehat{U}$. To do so, we first choose α, β, γ such that $\mathbf{R}(\tilde{\mathbf{u}}_1) = \hat{\mathbf{u}}_1$. 2-rotations are orthogonal linear functions, so they preserve the orthonormality of bases. Thus, $\mathbf{R}(\tilde{\mathbf{u}}_2)$ must have length 1 and be orthogonal to $\mathbf{R}(\tilde{\mathbf{u}}_1) = \hat{\mathbf{u}}_1$, so $\mathbf{R}(\tilde{\mathbf{u}}_2)$ is either $\hat{\mathbf{u}}_2$ or $-\hat{\mathbf{u}}_2$. Therefore, $\mathbf{R}(\widetilde{U}) = \{\hat{\mathbf{u}}_1, \pm\hat{\mathbf{u}}_2\}$. More formally, $\mathbf{R}(\widetilde{U}) = \widehat{U}^\sigma$ for some $\sigma \in S_2$.

□

[5] I 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.

Remark 5.50. 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 5.51. (In n dimensions, any ordered basis is "rotationally close" to a permuted orthonormal basis).

Let V be an n-dimensional vector space, let $k \leq n$, and consider k-dimensional subspaces $\widetilde{W}$ and W of V. If $\widetilde{U} = \{\widetilde{\mathbf{u}}_1, ..., \widetilde{\mathbf{u}}_n\}$ and $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ are orthonormal ordered bases of $\widetilde{W}$ and W, respectively, then there is an n-rotation taking $\widetilde{U}$ to some permutation $\widehat{U}^\sigma$ 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 $i \in \{1, ..., n\}$, let $\mathbf{R}_i$ be the 3-rotation taking $\{\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}\}$ 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 $\mathbf{R}_{\frac{n}{2}} \circ ... \circ \mathbf{R}_1$ is a composition of 3-rotations taking $\widetilde{U}$ to $\widehat{U}^\sigma$, 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 $\{\hat{\mathbf{u}}_{n-2}, \hat{\mathbf{u}}_{n-1}, \hat{\mathbf{u}}_n\}$ 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 $\mathbf{R}_{\frac{n-3}{2}} \circ ... \circ \mathbf{R}_1$ of 3-rotations taking $\widetilde{U} - \{\widetilde{\mathbf{u}}_{n-2}, \widetilde{\mathbf{u}}_{n-1}, \widetilde{\mathbf{u}}_n\}$ to $\widehat{U} - \{\hat{\mathbf{u}}_{n-2}, \hat{\mathbf{u}}_{n-1}, \hat{\mathbf{u}}_n\})^\sigma$, for some permutation $\sigma \in S_n$. Thus $\mathbf{R} \circ \mathbf{R}_{\frac{n-3}{2}} \circ ... \circ \mathbf{R}_1$ is a composition of 3-rotations taking $\widetilde{U}$ to $\widehat{U}^\pi$, where $\pi(i) = \sigma(i)$ when $i \in \{1, ..., n-3\}$ and $\pi(i) = \sigma_{n-2}(i)$ when $i \in \{n-2, n-1, n\}$. □

Completing the definition of orientation for inner product spaces

Definition 5.52. (Orientation of arbitrary ordered bases).

Let V be an n-dimensional inner product space, and fix an orthonormal basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ 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 a permutation of $\widehat{U}$ to be the orientation of the unique permuted orthonormal basis $\widehat{U}^\sigma$, $\sigma \in S^n$, of $\widehat{U}$ for which there exists an n-rotation taking E to $\widehat{U}^\sigma$.

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 4.23).

Theorem 5.53. (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 $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ 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 $\det([\mathbf{E}]_{\widehat{U}}) = \det([\mathbf{e}_1]_{\widehat{U}}, ..., [\mathbf{e}_n]_{\widehat{U}})$, and perform Gram-Schmidt on $\{[\mathbf{e}_1]_{\widehat{U}}, ..., [\mathbf{e}_n]_{\widehat{U}}\}$. In the ith step of Gram-Schmidt, a linear combination of the vectors $[\mathbf{e}_1]_{\widehat{U}}, ..., \cancel{[\mathbf{e}_i]_{\widehat{U}}}, ..., [\mathbf{e}_n]_{\widehat{U}}$ is added to $[\mathbf{e}_i]_{\widehat{U}}$. Recall from Theorem 5.25 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 $\widehat{U}_E = \{\widetilde{\mathbf{u}}_1, ..., \widetilde{\mathbf{u}}_n\}$ is the orthonormal basis obtained by performing Gram-Schmidt on E, then

$$\det([\mathbf{E}]_{\widehat{U}}) = \det([\mathbf{e}_1]_{\widehat{U}}, ..., [\mathbf{e}_n]_{\widehat{U}}) = \det([\widetilde{\mathbf{u}}_1]_{\widehat{U}}, ..., [\widetilde{\mathbf{u}}_n]_{\widehat{U}}) = \det([\widehat{U}_E]_{\widehat{U}}).$$

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 $\widehat{U}^\sigma$ of $\widehat{U}$.

Theorem 5.51 says that there is a n-rotation $\mathbf{R}$ taking $\widehat{U}_E$ to $\widehat{U}^\sigma$, for some $\sigma \in S_n$, and Theorem 5.48 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 $\det([\widehat{U}^\sigma]_{\widehat{U}}) = \mathrm{sgn}(\sigma)\det([\widehat{U}]_{\widehat{U}})$; once we have shown this, we are done, since $\mathrm{sgn}(\sigma)\det([\widehat{U}]_{\widehat{U}}) = \mathrm{sgn}(\sigma)\det(\mathbf{I}) = \mathrm{sgn}(\sigma)$. Since any permutation is a composition of "swaps" (a "swap" is a permutation defined on a two-element set), then $\widehat{U}^\sigma$ 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 $\mathrm{sgn}(\sigma)$ factor in the equation $\det([\widehat{U}^\sigma]_{\widehat{U}}) = \mathrm{sgn}(\sigma)\det([\widehat{U}]_{\widehat{U}})$. □

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 5.54. (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 5.55. (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 5.56. (Orientation with top degree wedges).

Let V be a finite-dimensional vector space with an orientation $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. 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 $\mathrm{sgn}(\sigma) < 0$.

Notice, we can identify $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ with $\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n \in \Lambda^n(V)$, because the antisymmetry of ordered bases is manifested in elements of $\Lambda^n(V)$ due to the antisymmetry of $\wedge$. 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 ordered 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 $\Lambda^n(V)$. The fact that the pushforward on the top exterior power is multiplication by the determinant (recall Theorem 5.32) plays nicely into this interpretation: once an orientation $\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n$ of V has been chosen, then we have $\mathbf{f}_1 \wedge ... \wedge \mathbf{f}_n = \det(\mathbf{f})\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n$, where $\mathbf{f}$ is the linear function $V \to V$ sending $\mathbf{e}_i \mapsto \mathbf{f}_i$.

As a last sidenote, the following theorem gives some justification as to why our definition of n-rotation was a good definition. (Strictly speaking, though, the justification is somewhat circular, since we used the notion of n-rotations to explain how the determinant- which is involved in the justification- tracks orientation).

Theorem 5.57. If V is an n-dimensional inner product space, then

$$\{n\text{-rotations on } V\} = \{\text{orthogonal linear functions } V \to V \text{ with determinant } 1\}$$

Proof. Proof idea is from jagr2808.

($\subseteq$). This is just Theorem 5.48.

($\supseteq$). If $\mathbf{f} : V \to V$ is an orthogonal linear function, then $\mathbf{f}$ sends an arbitrary orthonormal basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ of V to another orthonormal basis $\mathbf{f}(\widehat{U}) = \{\mathbf{f}(\hat{\mathbf{u}}_1), ..., \mathbf{f}(\hat{\mathbf{u}}_n)\}$.

Let $\mathbf{R}_1$ be the 2-rotation sending $\mathbf{f}(\hat{\mathbf{u}}_1)$ to $\hat{\mathbf{u}}_1$. Then $\mathbf{R}_2 := \mathbf{R}_1 \circ \mathbf{f} : V \to V$ fixes $\hat{\mathbf{u}}_1$, so we can think of it as a map from $\mathrm{span}(\hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_n)$ to $\mathrm{span}(\hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_n)$.

We now use the above idea finitely many times. Define $\mathbf{R}_{i+1}$ to be the 2-rotation sending $(\mathbf{R}_i \circ \mathbf{R}_{i-1})(\hat{\mathbf{u}}_i)$ to $\hat{\mathbf{u}}_i$. We know that such a 2-rotation always exists because $\det(\mathbf{f}) = 1$ implies that $\{\mathbf{f}(\hat{\mathbf{u}}_1), ..., \mathbf{f}(\hat{\mathbf{u}}_n)\}$ has the same orientation as $\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$. (Recall, having the same orientation involves "rotational equivalence" via n-rotations, which are compositions of 2-rotations). By induction, $\mathbf{R}_n \circ \mathbf{f}$ is a composition of 2-rotations.

Thus, since $\mathbf{f} = \mathbf{R}_n^{-1} \circ (\mathbf{R}_n \circ \mathbf{f})$, we see $\mathbf{f}$ is a composition of the 2-rotations $\mathbf{R}_n^{-1}$ and $\mathbf{R}_n^{-1} \circ \mathbf{f}$.

$\square$

5.4 Exterior powers as vector spaces of functions

When we first encountered (p, q) tensors, we learned that (p, q) tensors can be identified with multilinear functions; we saw that $T_{p,q}(V) \cong \mathcal{L}((V^*)^{\times p} \times V^{\times q} \to K)$ is a natural isomorphism when V is a finite-dimensional vector space. When $V_1, ..., V_k$ are finite-dimensional vector spaces, we similarly[6] have the natural isomorphism $V_1^* \otimes ... \otimes V_k^* \cong \mathcal{L}(V_1 \times ... \times V_k \to K)$.

Wedge product spaces and exterior powers can also be interpreted as vector spaces of functions. We will need to interpret exterior powers as spaces of functions in the setting of differential forms. Our two goals in this subsection are to (1) present that a k-wedge of covectors in V^* can act on k vectors from V to produce a scalar and to (2) give a new presentation of the pullback of a k-wedge of covectors.

To achieve our goals, we first formalize some notation about the alternative interpretations of tensor product spaces, (p, q) tensors, wedge product spaces, and exterior powers.

Derivation 5.58. (Tensor product that operates on tensors that are treated as functions).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces, and consider that $V_1^* \otimes ... \otimes V_k^* \cong \mathcal{L}(V_1 \times ... \times V_k \to K)$ naturally. Observe that the involvement of $\otimes$ in $V_1^* \otimes ... \otimes V_k^*$ induces a binary operation $\widetilde{\otimes}$ on $\mathcal{L}(V_1 \times ... \times V_k \to K)$. We construct $\widetilde{\otimes}$ by explicitly constructing a natural isomorphism $V_1^* \otimes ... \otimes V_k^* \cong \mathcal{L}(V_1 \times ... \times V_k \to K)$, and giving new notation to the output of this isomorphism.

The natural isomorphism $V_1^* \otimes ... \otimes V_k^* \cong \mathcal{L}(V_1 \times ... \times V_k \to K)$ is defined on the elementary tensor $\phi^1 \otimes ... \otimes \phi^k \in V_1^* \otimes ... \otimes V_k^*$, and extended with the seeming-multilinearity of $\otimes$ and the corresponding actual multilinearity of the newly-defined $\widetilde{\otimes}$. The isomorphism sends

$$\phi^1 \otimes ... \otimes \phi^k \mapsto \phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k,$$

where $\phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k : V^{\times k} \to K$ is the multilinear function defined by

$$(\phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k)(\mathbf{v}_1, ..., \mathbf{v}_k) := \phi^1(\mathbf{v}_1) ... \phi^k(\mathbf{v}_k).$$

Note, we have essentially reused the idea of the natural isomorphism from the third bullet point of Theorem 3.43 (this isomorphism is discussed in the proof of the referenced theorem).

Definition 5.59. (Tensor product spaces as vector spaces of functions).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces over a field K. We define $V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$ to be the vector space spanned by elements of the form $\phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k$, where $\phi^i \in V_i^*$. Thus,

$$V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^* = \mathcal{L}(V_1^* \times ... \times V_k^* \to K).$$

We also define $V_1 \widetilde{\otimes} ... \widetilde{\otimes} V_k := V_1^{**} \widetilde{\otimes} ... \widetilde{\otimes} V_k^{**} = \mathcal{L}(V_1^{**} \times ... \times V_k^{**} \to K)$.

Theorem 5.60. (Natural isomorphisms relating $\otimes$ and $\widetilde{\otimes}$).

For finite-dimensional vector spaces V and W, we have $V^* \widetilde{\otimes} W^* \cong V^* \otimes W^*$ naturally. From this, it follows that $V \widetilde{\otimes} W \cong V \otimes W$ naturally. Induction then gives that $V^{\widetilde{\otimes} k} \cong V^{\otimes k}$.

Proof. To prove the first statement, send elementary tensors to elementary tensors via the linear map $\phi \widetilde{\otimes} \psi \mapsto \phi \otimes \psi$, and extend this map with linearity. The rest follows easily. □

Definition 5.61. ((p, q) tensors as functions).

We define $\widetilde{T}_{p,q}(V) := (V^*)^{\widetilde{\otimes} p} \widetilde{\otimes} V^{\widetilde{\otimes} q}$. Note that, due to the previous theorem, there is a natural isomorphism $T_{p,q}(V) \cong \widetilde{T}_{p,q}(V)$.

Derivation 5.62. (Alternization of elements of $V_1 \widetilde{\otimes} ... \widetilde{\otimes} V_k$).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces, and consider the wedge product space $V_1^* \wedge ... \wedge V_k^*$. Due to the existence of the alt function on $V_1^* \wedge ... \wedge V_k^*$, the isomorphism $V_1^* \otimes ... \otimes V_k^* \cong V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$ induces a function $\widetilde{\text{alt}} : V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^* \to V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$:

$$\widetilde{\text{alt}}(\mathbf{T}) = \frac{1}{k!} \sum_{\sigma \in S_k} \text{sgn}(\sigma) \mathbf{T}^\sigma,$$

where $(\cdot)^\sigma$ is a permutation on elements of $V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$ induced by a permutation on elements of $V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$. To be extra clear, $(\cdot)^\sigma : V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^* \to V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$ is defined on elementary "tensors", and extended with multilinearity, by $(\phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k)^\sigma = \phi^{\sigma(1)} \widetilde{\otimes} ... \widetilde{\otimes} \phi^{\sigma(k)}$.

[6] $V_1^* \otimes ... \otimes V_k^* \cong (V_1 \otimes ... \otimes V_k)^* = \mathcal{L}(V_1 \otimes ... \otimes V_k \to K) \cong \mathcal{L}(V_1 \times ... \times V_k \to K)$.

Derivation 5.63. (Wedge product that operates on tensors that are treated as functions).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces. The wedge product $\wedge : (V_1^* \otimes ... \otimes V_k^*)^{\times 2} \to \text{alt}((V_1^* \otimes ... \otimes V_k^*)^{\otimes 2})$ induces a wedge product $\widetilde{\wedge} : (V_1 \widetilde{\otimes} ... \widetilde{\otimes} V_k^*)^{\times 2} \to \widetilde{\text{alt}}((V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*)^{\widetilde{\otimes} 2})$ defined by $\mathbf{T} \widetilde{\wedge} \mathbf{S} = \widetilde{\text{alt}}(\mathbf{T} \widetilde{\otimes} \mathbf{S})$. The wedge product $\widetilde{\wedge}$ is an alternating multilinear function because $\wedge$ is antisymmetric and appears multilinear.

Definition 5.64. (Wedge product spaces as vector spaces of functions).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces over a field K. We define $V_1^* \widetilde{\wedge} ... \widetilde{\wedge} V_k^*$ to be the vector space spanned by elements of the form $\phi^1 \widetilde{\wedge} ... \widetilde{\wedge} \phi^k$, where $\phi^i \in V_i^*$. Thus,

$$V_1^* \widetilde{\wedge} ... \widetilde{\wedge} V_k^* = (\text{alt}\mathcal{L})(V_1 \times ... \times V_k \to K).$$

We also define $V_1 \widetilde{\wedge} ... \widetilde{\wedge} V_k := V_1^{**} \widetilde{\wedge} ... \widetilde{\wedge} V_k^{**} = (\text{alt}\mathcal{L})(V_1^{**} \times ... \times V_k^{**} \to K)$.

Theorem 5.65. (Natural isomorphisms relating $\wedge$ and $\widetilde{\wedge}$).

For finite-dimensional vector spaces V and W, we have $V^* \widetilde{\wedge} W^* \cong V^* \wedge W^*$ naturally. From this, it follows that $V \widetilde{\wedge} W \cong V \wedge W$ naturally. Induction then gives that $V^{\widetilde{\wedge} k} \cong V^{\wedge k}$.

Proof. To prove the first statement, send elementary tensors to elementary tensors via the linear map $\phi \widetilde{\wedge} \psi \mapsto \phi \wedge \psi$, and extend this map with linearity and antisymmetry. The rest follows easily. □

Definition 5.66. (Exterior powers as vector spaces of functions).

Let V be a finite-dimensional vector space over a field K. Because the previous definition shows $\Lambda^k(V^*) = (V^*)^{\wedge k} \cong (V^*)^{\widetilde{\wedge} k}$ naturally, we define $\widetilde{\Lambda}^k(V^*) := (V^*)^{\widetilde{\wedge} k} = (\text{alt}\mathcal{L})(V^{\times k} \to K)$.

We have now laid out the landscape for the interpretation of tensor product spaces and wedge product spaces as vector spaces of functions. In doing so, we described explicitly how $\widetilde{\otimes}$ acts on multilinear functions to produce a multilinear function. To complete this section, we show how $\widetilde{\wedge}$ acts on alternating multilinear functions to produce an alternating multilinear function, and present the pullback of elements of exterior powers in the context where exterior powers are thought of as vector spaces of functions.

Lemma 5.67. (Pushforward on the dual is multiplication by $\det(\phi^i(\mathbf{e}_j))$).

Let V be an n-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^*. For all $\phi^1, ..., \phi^n \in V^*$, we have

$$\phi^1 \wedge ... \wedge \phi^n = \det(\mathbf{f}^*)\phi^{\mathbf{e}_1} \wedge ... \wedge \phi^{\mathbf{e}_n},$$

because the pushforward on a top exterior power is multiplication by the determinant (recall Theorem 5.32). Since the matrix of $\mathbf{f}^*$ relative to E^* and E^* has ith column $[\phi^i]_{E^*}$, we have

$$\begin{aligned} \phi^1 \wedge ... \wedge \phi^n &= \det \begin{pmatrix} [\phi^1]_{E^*} & \dots & [\phi^1]_{E^*} \end{pmatrix} \phi^{\mathbf{e}_1} \wedge ... \wedge \phi^{\mathbf{e}_n} \\ &= \det(\phi^i(\mathbf{e}_j))\phi^{\mathbf{e}_1} \wedge ... \wedge \phi^{\mathbf{e}_n}. \end{aligned}$$

To obtain the last expression, recall the fact that for any $\phi \in V^*$ and basis F^* of V^* induced by a basis $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$ of V, we have $([\phi]_{F^*})_i = \phi(\mathbf{f}_i)$ (see Theorem 4.28).

Remark 5.68. Interestingly, the fact $\det(\mathbf{f}^*) = \det(\mathbf{f})$ does *not* come in useful for the current line of argument (the "current line of argument" started with the previous lemma).

Lemma 5.69. (Action of dual n-wedge on dual basis).

Let V be an n-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^*. Consider an element $\phi^1 \widetilde{\wedge} ... \widetilde{\wedge} \phi^n \in \widetilde{\Lambda}^n(V^*)$. We have

$$(\phi^{\mathbf{e}_1} \widetilde{\wedge} ... \widetilde{\wedge} \phi^{\mathbf{e}_n})(\mathbf{e}_1, ..., \mathbf{e}_n) = 1.$$

Proof. Using a similar argument to the one that showed the permutation formula for the determinant on the $\phi^{\mathbf{e}_i}$, we have

$$\phi^{\mathbf{e}_1} \widetilde{\wedge} ... \widetilde{\wedge} \phi^{\mathbf{e}_n} = \widetilde{\text{alt}}(\phi^{\mathbf{e}_1} \widetilde{\wedge} ... \widetilde{\otimes} \phi^{\mathbf{e}_n}) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \phi^{\mathbf{e}_{\sigma(1)}} \widetilde{\otimes} ... \widetilde{\otimes} \phi^{\mathbf{e}_{\sigma(n)}}.$$

Therefore

$$(\phi^{\mathbf{e}_1}\widetilde{\wedge}...\widetilde{\wedge}\phi^{\mathbf{e}_n})(\mathbf{e}_1,...,\mathbf{e}_n) = \Big(\sum_{\sigma \in S_n} \text{sgn}(\sigma)\phi^{\mathbf{e}_{\sigma(1)}}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\mathbf{e}_{\sigma(n)}}\Big)(\mathbf{e}_1,...,\mathbf{e}_n) = \sum_{\sigma \in S_n} \Big(\text{sgn}(\sigma)(\phi^{\mathbf{e}_{\sigma(1)}}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\mathbf{e}_{\sigma(n)}})(\mathbf{e}_1,...,\mathbf{e}_n)\Big).$$

Now we focus on the inner term, $(\phi^{\mathbf{e}_{\sigma(1)}}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\mathbf{e}_{\sigma(n)}})(\mathbf{e}_1,...,\mathbf{e}_n)$. By definition of $\widetilde{\otimes}$, we have

$$(\phi^{\mathbf{e}_{\sigma(1)}}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\mathbf{e}_{\sigma(n)}})(\mathbf{e}_1,...,\mathbf{e}_n) = \phi^{\mathbf{e}_{\sigma(1)}}(\mathbf{e}_1)...\phi^{\mathbf{e}_{\sigma(n)}}(\mathbf{e}_n)$$

Since $\epsilon^{\sigma(i)}(\mathbf{e}_j) = \delta^{\sigma(i)}{}_j$, 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

$$(\phi^{\mathbf{e}_1}\widetilde{\wedge}...\widetilde{\wedge}\phi^{\mathbf{e}_n})(\mathbf{e}_1,...,\mathbf{e}_n) = \text{sgn}(i) \cdot 1 = 1 \cdot 1 = 1.$$

□

Theorem 5.70. (Action of dual n-wedge on vectors).

Let V be an n-dimensional vector space with basis $E = \{\mathbf{e}_1,...,\mathbf{e}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1},...,\phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^*. Consider an element $\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n \in \widetilde{\Lambda}^n(V^*)$. We have

$$(\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n)(\mathbf{e}_1,...,\mathbf{e}_n) = \det(\phi^i(\mathbf{e}_j)).$$

By extending with antisymmetry and multilinearity, we have

$$\boxed{(\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n)(\mathbf{v}_1,...,\mathbf{v}_n) = \det(\phi^i(\mathbf{v}_j)) \text{ for all } \mathbf{v}_1,...,\mathbf{v}_n \in V}$$

The action of $\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n$ on $\mathbf{v}_1,...,\mathbf{v}_n$ can be interpreted as follows. First notice that the above is nonzero only when $\mathbf{v}_1,...,\mathbf{v}_n$ are linearly independent, i.e., when $E := \{\mathbf{v}_1,...,\mathbf{v}_n\}$ is a basis[7]. Thus, when $\mathbf{v}_1,...,\mathbf{v}_n$ are linearly independent, we can interpret $\phi^i(\mathbf{v}_j)$ in the style of Lemma 5.67 as $\phi^i(\mathbf{v}_j) = ([\phi^i]_{E^*})^j$, where $E^* := \{\phi^{\mathbf{v}_1},...,\phi^{\mathbf{v}_n}\}$ is the dual basis for V^* induced by the basis E for V:

$$(\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n)(\mathbf{v}_1,...,\mathbf{v}_n) = \det\begin{pmatrix}[\phi^1]_{E^*} & \dots & [\phi^1]_{E^*}\end{pmatrix} \text{ when } \mathbf{v}_1,...,\mathbf{v}_n \text{ are linearly independent.}$$

Proof. The first of the previous two lemmas implies $\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n = \det(\mathbf{f}^*)\phi^{\mathbf{e}_1}\widetilde{\wedge}...\widetilde{\wedge}\phi^{\mathbf{e}_n}$. (We have just put $\sim$'s over the $\wedge$'s of the first lemma). Combine this with the second of the previous two lemmas. □

So, we have shown how a k-wedge of elements from V^*, when treated as a function, can act on vectors from V. Lastly, we present the pushforward and pullback of elements of exterior powers when exterior powers are interpreted to be spaces of functions. (We are interested in the pullback, but present the pushforward for completeness).

Pushforward and pullback on vector spaces of functions

Derivation 5.71. (Pushforward on $\widetilde{T}_0^k(V)$)

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f}: V \to W$. Definition 5.21 defined the pushforward $\otimes_{k,0}\mathbf{f}: T_0^k(V) \to T_0^k(W)$ by

$$\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k \xmapsto{\widetilde{\otimes}_0^k \mathbf{f}} \mathbf{f}(\mathbf{v}_1) \otimes ... \otimes \mathbf{f}(\mathbf{v}_k).$$

The natural isomorphisms $T_0^k(V) \cong \widetilde{T}_0^k(V)$ and $T_0^k(W) \cong \widetilde{T}_0^k(W)$ induce a pushforward map $\widetilde{\otimes}_0^k\mathbf{f}: \widetilde{T}_0^k(V) \to \widetilde{T}_0^k(W)$ defined by

[7]Quick proof that (linearly independent) $\implies$ (basis) in this scenario. If $E := \{\mathbf{v}_1,...,\mathbf{v}_n\}$ is not a basis for V, then E is a linearly independent set of n vectors that doesn't span V, so we need to add more linearly independent vectors to E in order to obtain a basis; that is, $\dim(V) > n$, which is a contradiction.

$$\mathbf{v}_1 \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{v}_k \overset{\widetilde{\otimes}^k_0 \mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{f}(\mathbf{v}_k).$$

Notice that by definition of $\widetilde{\otimes}$, we have $\mathbf{f}(\mathbf{v}_1) \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{f}(\mathbf{v}_k) = (\mathbf{f}^{\widetilde{\otimes} k})(\mathbf{v}_1, ..., \mathbf{v}_i)$. Therefore

$$\mathbf{v}_1 \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{v}_k \overset{\widetilde{\otimes}^k_0 \mathbf{f}}{\longmapsto} \mathbf{f}^{\widetilde{\otimes} k}(\mathbf{v}_1, ..., \mathbf{v}_k).$$

Derivation 5.72. (Pullback on $\widetilde{T}_{0,k}(W)$).

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \rightarrow W$. Definition **??** defined the pullback $\otimes_{k,0} \mathbf{f}^* : T_{0,k}(W) \rightarrow T_{0,k}(V)$ by

$$\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k \overset{\Lambda^k \mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1) \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{f}^*(\psi^k).$$

The natural isomorphisms $T_{0,k}(W) \cong \widetilde{T}_{0,k}(W)$ and $T_{0,k}(V) \cong \widetilde{T}_{0,k}(V)$ induce a pullback map $\widetilde{\otimes}^0_k \mathbf{f}^* : \widetilde{T}_{0,k}(W) \rightarrow \widetilde{T}_{0,k}(V)$ defined by

$$\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k \overset{\widetilde{\otimes}^0_k \mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1) \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{f}^*(\psi^k) = (\psi^1 \circ \mathbf{f}) \widetilde{\otimes} ... \widetilde{\otimes} (\psi^k \circ \mathbf{f})$$

It can be checked using the definition of $\widetilde{\otimes}$ that $(\psi^1 \circ \mathbf{f}) \widetilde{\otimes} ... \widetilde{\otimes} (\psi^k \circ \mathbf{f}) = (\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k) \circ \mathbf{f}$. Therefore

$$\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k \overset{\widetilde{\otimes}^0_k \mathbf{f}^*}{\longmapsto} (\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k) \circ \mathbf{f}.$$

The above is a statement on the elementary tensor $\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k$. Extending this statement using the multilinearity of $\widetilde{\otimes}$, we have

$$\mathbf{T} \in \widetilde{T}_{0,k}(W) \overset{\widetilde{\otimes}^0_k \mathbf{f}^*}{\longmapsto} \mathbf{T} \circ \mathbf{f} \in \widetilde{T}_{0,k}(V) \text{ for all } \mathbf{T} \in \widetilde{T}_{0,k}(W).$$

Since $\mathbf{T} \circ \mathbf{f} = \mathbf{f}^*(\mathbf{T})$, this is equivalent to

$$\mathbf{T} \in \widetilde{T}_{0,k}(W) \overset{\widetilde{\otimes}^0_k \mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\mathbf{T}) \in \widetilde{T}_{0,k}(V) \text{ for all } \mathbf{T} \in \widetilde{T}_{0,k}(W).$$

Since elements of $\widetilde{T}_{0,k}(V)$ act on k vectors from V, we now ask, how does $\otimes_{0,k} \mathbf{f}^*$ act on k vectors from V? Well,

$$\otimes_{0,k} \mathbf{f}^*(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{f}^*(\mathbf{T})(\mathbf{v}_1, ..., \mathbf{v}_k) = (\mathbf{T} \circ \mathbf{f})(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k)).$$

Thus $\otimes_{0,k} \mathbf{f}^*$ acts on k vectors from V by

$$\boxed{\otimes_{0,k} \mathbf{f}^*(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k))}$$

Derivation 5.73. (Pushforward on $\widetilde{\Lambda}^k(V)$).

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \rightarrow W$. Definition **??** defined the pushforward $\Lambda^k \mathbf{f} : \Lambda^k(V) \rightarrow \Lambda^k(W)$ by

$$\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k \overset{\Lambda^k \mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_k).$$

Equivalently,

$$\text{alt}(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k) \overset{\Lambda^k \mathbf{f}}{\longmapsto} \text{alt}(\otimes_{k,0} \mathbf{f}(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k)).$$

Since the pushforward $\otimes_{k,0} : T^k_0(V) \to T^k_0(W)$ induces the pushforward $\widetilde{\otimes}^k_0 : \widetilde{T}^k_0(V) \to \widetilde{T}^k_0(W)$, then we obtain a pushforward $\widetilde{\Lambda}^k\mathbf{f} : \widetilde{\Lambda}^k(V) \to \widetilde{\Lambda}^k(W)$ defined by

$$\mathbf{v}_1\widetilde{\wedge}...\widetilde{\wedge}\mathbf{v}_k = \widetilde{\text{alt}}(\mathbf{v}_1\widetilde{\otimes}...\widetilde{\otimes}\mathbf{v}_k) \overset{\widetilde{\Lambda}^k\mathbf{f}}{\longmapsto} \widetilde{\text{alt}}(\widetilde{\otimes}^k_0\mathbf{f}(\mathbf{v}_1\widetilde{\otimes}...\widetilde{\otimes}\mathbf{v}_k)) = \mathbf{f}^{\widetilde{\wedge}k}(\mathbf{v}_1, ..., \mathbf{v}_k).$$

That is,

$$\mathbf{v}_1\widetilde{\wedge}...\widetilde{\wedge}\mathbf{v}_k \overset{\widetilde{\Lambda}^k\mathbf{f}}{\longmapsto} \mathbf{f}^{\widetilde{\wedge}k}(\mathbf{v}_1, ..., \mathbf{v}_k).$$

Derivation 5.74. (Pullback on $\widetilde{\Lambda}^k(W^*)$).

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Definition **??** defined the pullback $\Lambda^k\mathbf{f}^* : \Lambda^k(W^*) \to \Lambda^k(V^*)$ by

$$\psi^1 \wedge ... \wedge \psi^k \overset{\Lambda^k\mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1) \wedge ... \wedge \mathbf{f}^*(\psi^k).$$

Equivalently,

$$\text{alt}(\psi^1 \otimes ... \otimes \psi^k) \overset{\Lambda^k\mathbf{f}^*}{\longmapsto} \text{alt}(\otimes_{0,k}\mathbf{f}^*(\psi^1 \otimes ... \otimes \psi^k)).$$

Since the pullback $\otimes_{0,k}\mathbf{f}^* : T_{0,k}(W) \to T_{0,k}(V)$ induces the pullback $\widetilde{\otimes}^0_k\mathbf{f}^* : \widetilde{T}_{0,k}(W) \to \widetilde{T}_{0,k}(V)$, then we obtain a pullback $\widetilde{\Lambda}^k\mathbf{f}^* : \widetilde{\Lambda}^k(W^*) \to \widetilde{\Lambda}^k(V^*)$ defined by

$$\begin{aligned}\psi^1\widetilde{\wedge}...\widetilde{\wedge}\psi^k = \widetilde{\text{alt}}(\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k) &\overset{\widetilde{\Lambda}^k\mathbf{f}^*}{\longmapsto} \widetilde{\text{alt}}(\widetilde{\otimes}^0_k\mathbf{f}^*(\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k)) \\ &= \widetilde{\text{alt}}((\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k) \circ \mathbf{f}) = \widetilde{\text{alt}}(\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k) \circ \mathbf{f} = (\psi^1\widetilde{\wedge}...\widetilde{\wedge}\psi^k) \circ \mathbf{f}.\end{aligned}$$

(The equality between the rightmost expression of the first line and the leftmost expression of the second line uses the fact that $\widetilde{\otimes}\mathbf{f}^*(\mathbf{T}) = \mathbf{f}^*(\mathbf{T})$, which is presented in Derivation 5.72. The validity of the second equals sign in the second line has not been proven, but it is quickly checked).

Overall, the above line reads

$$\psi^1\widetilde{\wedge}...\widetilde{\wedge}\psi^k \overset{\widetilde{\Lambda}^k\mathbf{f}^*}{\longmapsto} (\psi^1\widetilde{\wedge}...\widetilde{\wedge}\psi^k) \circ \mathbf{f}.$$

Extending with the multilinearity and alternatingness of $\widetilde{\wedge}$, we extend this statement to

$$\mathbf{T} \in \Lambda^k(W^*) \overset{\widetilde{\Lambda}^k\mathbf{f}^*}{\longmapsto} \mathbf{T} \circ \mathbf{f} = \mathbf{f}^*(\mathbf{T}) \in \Lambda^k(V^*).$$

Therefore, $\widetilde{\Lambda}^k\mathbf{f}^* : \widetilde{\Lambda}^k(W^*) \to \widetilde{\Lambda}^k(V^*)$ acts on k vectors from V by

$$\boxed{\widetilde{\Lambda}^k\mathbf{f}^*(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k))}$$

In appearance, it seems that $\widetilde{\Lambda}^k\mathbf{f}^*$ acts on $\mathbf{v}_1, ..., \mathbf{v}_k$ exactly as does $\widetilde{\otimes}^0_k\mathbf{f}^*$, since $\widetilde{\otimes}^0_k\mathbf{f}^*(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k))$. The distinction between the two definitions is that $\widetilde{\Lambda}^k\mathbf{f}^*$ acts on alternating multilinear maps $\mathbf{T}$, while $\widetilde{\otimes}^0_k\mathbf{f}^*$ acts on multilinear (not necessarily alternating) maps $\mathbf{T}$.

5.5 The cross product

At the end of Chapter 2, we presented an explanation for the dot product that remedies two common problematic ways to explain the dot product. The cross product also comes with two common pedagogical problems. We describe and remedy those here.

The first pedagogical problem with the cross product is that the complicated algebraic formula for the cross product is rarely explained. (The formula is $\mathbf{v} \times \mathbf{w} = \Big(([\mathbf{v}_1]_E)^2([\mathbf{v}_2]_E)^3 - ([\mathbf{v}_1]_E)^3([\mathbf{v}_2]_E)^2\Big) - \Big(([\mathbf{v}_1]_E)^1([\mathbf{v}_2]_E)^3 - ([\mathbf{v}_1]_E)^3([\mathbf{v}_2]_E)^1\Big) + \Big(([\mathbf{v}_1]_E)^1([\mathbf{v}_2]_E)^2 - ([\mathbf{v}_1]_E)^2([\mathbf{v}_2]_E)^1\Big)$).

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 conventions about orientation.

Derivation 5.75. (Cross product).

In Section 2.6, we presented the dot product by starting with an intuitive geometric definition of the dot product and building onto that definition. If we were to define the cross product in the same manner, we would start with this geometric definition:

The cross product of $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ is the vector $(\mathbf{v}_1 \times \mathbf{v}_2) \in \mathbb{R}^3$ such that

- $||\mathbf{v}_1 \times \mathbf{v}_2||$ is the area of the parallelogram spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$
- $\widehat{\mathbf{v}_1 \times \mathbf{v}_2}$ is the vector perpendicular to $\mathbf{v}_1$ and $\mathbf{v}_2$ such that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_1 \times \mathbf{v}_2\}$ is positively oriented (this condition is equivalent to the "right hand rule")

Starting with the above definition is better than most non-explanations of the cross product, but is still problematic, as doing so begs the question, "Why is it natural for some vector to satisfy these two conditions?". A better way to introduce the cross product is to stumble across a vector that satisfies these two conditions, and then define that vector to be $\mathbf{v}_1 \times \mathbf{v}_2$. This is what we will do.

Consider the linear function $\mathbf{f}_{\mathbf{v}_1,\mathbf{v}_2} : \mathbb{R}^3 \to \mathbb{R}$ defined by $\mathbf{f}_{\mathbf{v}_1,\mathbf{v}_2}(\mathbf{v}) = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v})$. Since $\mathbf{f}_{\mathbf{v}_1,\mathbf{v}_2} : \mathbb{R}^3 \to \mathbb{R}$, then $\mathbf{f}_{\mathbf{v}_1,\mathbf{v}_2}$ must be represented by a 1×3 matrix: $\mathbf{f}_{\mathbf{v}_1,\mathbf{v}_2}(\mathbf{v}) = \mathbf{c}^\top \mathbf{v}$ for some $\mathbf{c} \in \mathbb{R}^3$. We define the *cross product of* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ to be this $\mathbf{c}$. That is, we define the cross product of $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ to be the unique vector $(\mathbf{v}_1 \times \mathbf{v}_2) \in \mathbb{R}^3$ such that

$$(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}) \text{ for all } \mathbf{v} \in \mathbb{R}^3.$$

We now prove that $\mathbf{v}_1 \times \mathbf{v}_2$ satisfies the previously mentioned geometric properties.

Theorem 5.76. (Magnitude, direction of cross product).

The cross product $\mathbf{v}_1 \times \mathbf{v}_2$ of $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ is such that:

- $||\mathbf{v}_1 \times \mathbf{v}_2||$ is the area of the parallelogram spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$
- $\widehat{\mathbf{v}_1 \times \mathbf{v}_2}$ is the vector perpendicular to $\mathbf{v}_1$ and $\mathbf{v}_2$ such that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_1 \times \mathbf{v}_2\}$ is positively oriented

Proof.

- We have $(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v} = ||\text{proj}(\mathbf{v} \to \mathbf{v}_1 \times \mathbf{v}_2)||\ ||\mathbf{v}_1 \times \mathbf{v}_2||$. By the definition of the cross product, we have $||\text{proj}(\mathbf{v} \to \mathbf{v}_1 \times \mathbf{v}_2)||\ ||\mathbf{v}_1 \times \mathbf{v}_2|| = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v})$, so $||\mathbf{v}_1 \times \mathbf{v}_2|| = \frac{\det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v})}{||\text{proj}(\mathbf{v} \to \mathbf{v}_1 \times \mathbf{v}_2)||}$. This is equal to the volume of the parallelapiped spanned by $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}$ divided by the height of the same parallelapiped, which is the same as the area of the base of the parallelapiped, i.e., the area of the parallelogram spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$.
- Consider $\mathbf{v}_i$, $i \in \{1, 2\}$. We have $(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_i = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_i)$. Since $\mathbf{v}_i \in \{\mathbf{v}_1, \mathbf{v}_2\}$, then two vectors in the determinant are the same. This implies the determinant is zero.

□

The below is in-progress

Definition 5.77. (Signed counterclockwise angle).

$\theta_{\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2)$ for full formal notation, where $\hat{\mathbf{n}}$ is one of the two vectors perpendicular to $\mathbf{v}_1$, $\mathbf{v}_2$; $\theta(\mathbf{v}_1, \mathbf{v}_2)$ for more colloquial

we have $\theta_{\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2) = 2\pi - \theta_{-\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2)$

Theorem 5.78. (Geometric formula for magnitude of the cross product).

The magnitude of the cross product $\mathbf{v}_1 \times \mathbf{v}_2$ of $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ is

$$||\mathbf{v}_1 \times \mathbf{v}_2|| = ||\mathbf{v}_1||\, ||\mathbf{v}_2||\, |\sin(\theta(\mathbf{v}_1, \mathbf{v}_2))|,$$

where θ is the signed counterclockwise angle from $\mathbf{v}_1$ to $\mathbf{v}_2$.

Proof. Basic trigonometry shows this result. The absolute value around $\sin(\theta(\mathbf{v}_1, \mathbf{v}_2))$ ensures the RHS is always nonnegative. □

Derivation 5.79. (Right hand rule).

Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$, and consider the ordered basis $\{\mathbf{v}_1, \mathbf{v}_2, \hat{\mathbf{n}}\}$, where $\hat{\mathbf{n}}$ is either one of the two vectors perpendicular to $\mathbf{v}_1$ and $\mathbf{v}_2$. (So $\hat{\mathbf{n}}$ is either $\widehat{\mathbf{v}_1 \times \mathbf{v}_2}$ or $-\widehat{\mathbf{v}_1 \times \mathbf{v}_2}$).

Since $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_1 \times \mathbf{v}_2\}$ is positively oriented, $\widehat{\mathbf{v}_1 \times \mathbf{v}_2} = \hat{\mathbf{n}}$ iff $\{\mathbf{v}_1, \mathbf{v}_2, \hat{\mathbf{n}}\}$ is positively oriented. And $\{\mathbf{v}_1, \mathbf{v}_2, \hat{\mathbf{n}}\}$ is positively oriented iff $\theta_{\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2) \in (0, \pi)$.

So $\widehat{\mathbf{v}_1 \times \mathbf{v}_2} = \hat{\mathbf{n}}$ iff $\theta_{\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2) \in (0, \pi)$. By similar reasoning, $\widehat{\mathbf{v}_1 \times \mathbf{v}_2} = -\hat{\mathbf{n}}$ iff $\theta_{\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2) \in (\pi, 2\pi)$.

In all:

- $\widehat{\mathbf{v}_1 \times \mathbf{v}_2} = \hat{\mathbf{n}}$ iff $\theta_{\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2) \in (0, \pi)$
- $\widehat{\mathbf{v}_1 \times \mathbf{v}_2} = -\hat{\mathbf{n}}$ iff $\theta_{\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2) \in (\pi, 2\pi)$

Think of $\hat{\mathbf{n}}$ as designating which side of the plane spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$ is the "top" side. Then the above two bullet points tell us that the cross product $\mathbf{v}_1 \times \mathbf{v}_2$ points "up" (i.e. to the top side) when $\theta_{\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2) \in (0, \pi)$ and points "down" (i.e. to the bottom side) when $\theta_{\hat{\mathbf{n}}}(\mathbf{v}_1, \mathbf{v}_2) \in (0, 2\pi)$.

Theorem 5.80. (Cyclic pattern with the cross product).

$\mathbf{e}_{\sigma(1)} \times \mathbf{e}_{\sigma(2)} = \text{sgn}(\sigma)\mathbf{e}_{\sigma(3)}$

so $\hat{\mathfrak{e}}_1 \times \hat{\mathfrak{e}}_2 = \hat{\mathfrak{e}}_3$, $\hat{\mathfrak{e}}_2 \times \hat{\mathfrak{e}}_3 = \hat{\mathfrak{e}}_1$, $\hat{\mathfrak{e}}_3 \times \hat{\mathfrak{e}}_1 = \hat{\mathfrak{e}}_2$

Derivation 5.81. (Cross product is Hodge dual of wedge product).

$\Lambda^2(V) \cong V$ iff $\dim(\Lambda^2(V)) = \dim(V)$ iff $\binom{\dim(V)}{2} = \dim(V)$ iff $\dim(V) = 3$. In all, $\Lambda^2(V) \cong V$ iff $\dim(V) = 3$. This shows in particular that $\Lambda^2(\mathbb{R}^3)$ and $\mathbb{R}^3$ are isomorphic.

One isomorphism $\perp\colon \Lambda^2(\mathbb{R}^3) \to \mathbb{R}^3$ is defined by $\perp(\hat{\mathfrak{e}}_{\sigma(1)} \wedge \hat{\mathfrak{e}}_{\sigma(2)}) = \text{sgn}(\sigma)\hat{\mathfrak{e}}_{\sigma(3)}$. We call $\perp$ the *Hodge dual.*

$$\perp(\mathbf{v}_1 \wedge \mathbf{v}_2) = \perp\Big(\sum_{i=1}^n \Big(([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i \hat{\mathfrak{e}}_i\Big) \wedge \sum_{i=1}^n \Big(([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_i \hat{\mathfrak{e}}_i\Big)\Big) = ... = \mathbf{v}_1 \times \mathbf{v}_2$$

The previous theorem motivates us to ask if $\perp(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_{n-1})$ satisfies a condition that generalizes the defining property of the cross product. The following theorem provides the answer: "yes!".

Theorem 5.82. (Hodge dual generalizes cross product).

Let $\mathbf{v}_1, ..., \mathbf{v}_{n-1} \in \mathbb{R}^n$. Then $\perp(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_{n-1}) \in \mathbb{R}^n$ is the unique vector satisfying

$$\perp(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_{n-1}) \cdot \mathbf{v} = \det(\mathbf{v}_1, ..., \mathbf{v}_{n-1}, \mathbf{v}) \text{ for all } \mathbf{v} \in \mathbb{R}^n.$$

Proof. Note that *if* the condition is satisfied, then uniqueness easily follows, since the function $\mathbf{v} \mapsto \det(\mathbf{v}_1, ..., \mathbf{v}_{n-1}, \mathbf{v})$, being a linear function, has a unique matrix.

We now show that the condition is satisfied. $(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_{n-1}) \mapsto \perp(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_{n-1}) \cdot \mathbf{v}$ and $(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_{n-1}) \mapsto \det(\mathbf{v}_1, ..., \mathbf{v}_{n-1}, \mathbf{v})$ are both multilinear alternating functions, so they must be scalar multiples of each other[8]. Thus, to show the condition, it suffices to show that the two maps agree on a particular input. We show $\perp(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_{n-1}) \cdot \mathbf{v} = \det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_{n-1}, \mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^n$. This condition is logically equivalent to the condition "$\hat{\mathfrak{e}}_n \cdot \mathbf{v} = \det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_{n-1}, \mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^n$", which is in turn logically equivalent to "$\hat{\mathfrak{e}}_n \cdot \mathbf{v} = \det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_{n-1}, \mathbf{v})$ for all $\mathbf{v} \in \hat{\mathfrak{E}}$, where $\hat{\mathfrak{E}}$ is the standard basis for $\mathbb{R}^n$". When $\mathbf{v} = \hat{\mathfrak{e}}_i \in \hat{\mathfrak{E}}$, we have $\hat{\mathfrak{e}}_n \cdot \hat{\mathfrak{e}}_i = \delta^i{}_n = \det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_{n-1}, \hat{\mathfrak{e}}_i)$, so this last condition (and thus all of them) are true. □

[8] In Section 5.2, we proved that the determinant $\det : (K^n)^{\times n} \to K$ is the unique multilinear alternating function satisfying $\det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$. An arbitrary multilinear alternating function $f : (K^n)^{\times n} \to K$ is such that $\frac{1}{c}f$ satisfies $(\frac{1}{c}f)(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$, so such an f is $f = c\det$.

Remark 5.83. (Interpretation of k-wedges as oriented volumes).

We can actually define the Hodge dual in a more general context.

Definition 5.84. (Hodge dual).
Let V be a finite-dimensional vector space with a metric tensor g. We define the *Hodge dual (induced by g)* to be the function $\perp\colon \Lambda^k(V) \to \Lambda^{n-k}(V)$ defined by

$$\perp (\hat{\mathbf{u}}_{i_1} \wedge ... \wedge \hat{\mathbf{u}}_{i_k}) := \mathrm{sgn}(\sigma)\ \hat{\mathbf{u}}_{i_{k+1}} \wedge ... \wedge \hat{\mathbf{u}}_{i_{k+(n-k)}},$$
$$\text{where } \{i_{k+1}, ..., i_{k+(n-k)}\} = \{1, ..., n\} - \{i_1, ..., i_k\} \text{ and } \sigma = (i_1, ..., i_n).$$

[Extending with multilinearity and alternatingness], we see that

$$\perp (\mathbf{v}_{i_1} \wedge ... \wedge \mathbf{v}_{i_k}) = \mathrm{sgn}(\sigma) \det \Big((g(\mathbf{v}_{i_k}, \mathbf{v}_{j_\ell})) \Big)\ \mathbf{v}_{i_{k+1}} \wedge ... \wedge \mathbf{v}_{i_n},$$
$$\text{where } \{i_{k+1}, ..., i_n\} = \{1, ..., n\} - \{i_1, ..., i_k\} \text{ and } \sigma = (i_1, ..., i_n).$$

In the above, $\det \Big((\langle \mathbf{v}_{i_k}, \mathbf{v}_{j_\ell} \rangle) \Big)$ denotes the determinant of the matrix with $k\ell$ entry $\langle \mathbf{v}_{i_k}, \mathbf{v}_{j_\ell} \rangle$, and is called the *Gram determinant*. Interestingly enough, the function $\widetilde{g}$ that sends $(\mathbf{v}_{i_1} \wedge ... \mathbf{v}_{i_k}, \mathbf{w}_{i_1} \wedge ... \mathbf{w}_{i_k})$ to the Gram determinant is a metric tensor on $\Lambda^k(V)$. We can use this new metric tensor to restate the above as

$$\perp \mathbf{v} = \perp (\mathbf{v}_{i_1} \wedge ... \wedge \mathbf{v}_{i_k}) = \mathrm{sgn}(\sigma) \widetilde{g}(\mathbf{v}, \mathbf{v})\ \mathbf{v}_{i_{k+1}} \wedge ... \wedge \mathbf{v}_{i_n},$$
$$\text{where } \{i_{k+1}, ..., i_n\} = \{1, ..., n\} - \{i_1, ..., i_k\} \text{ and } \sigma = (i_1, ..., i_n).$$

Most authors use $*$ to denote the Hodge dual, and the Hodge dual is often known as the "Hodge star". We've chosen not to use this notation so as to avoid confusion with the $*$ that appears when notating dual vector spaces.

Remark 5.85. The above definition of the Hodge dual depends on a basis, so we do not yet know that the Hodge dual is well-defined.

Theorem 5.86. (Double Hodge dual).
Let V be a finite-dimensional vector space with metric tensor g. The Hodge dual $\perp\colon \Lambda^k(V) \to \Lambda^{n-k}(V)$ satisfies $\perp\perp \mathbf{v} = (-1)^{k(n-k)} \mathbf{v}$.

Proof. It suffices to prove the above when $\mathbf{v}$ is an elementary wedge, $\mathbf{v} = \mathbf{v}_{i_1} \wedge ... \wedge \mathbf{v}_{i_k}$. Using Definition ..., we have $\perp \mathbf{v} = \mathrm{sgn}(\sigma) g(\mathbf{v}, \mathbf{v}) \mathbf{v}_{i_{k+1}} \wedge ... \wedge \mathbf{v}_{i_n}$, where $\sigma = (i_1, ..., i_n)$. Applying $\perp$ again, we have $\perp\perp \mathbf{v} = \mathrm{sgn}(\tau)\mathrm{sgn}(\sigma) \mathbf{v}_{i_1} \wedge ... \wedge \mathbf{v}_{i_k} = \mathrm{sgn}(\tau)\mathrm{sgn}(\sigma)\mathbf{v}$, where $\tau = (i_{k+1}, ..., i_n, i_1, ..., i_k)$. The key is to notice that $\mathrm{sgn}(\tau) = (-1)^{k(n-k+1)} \mathrm{sgn}(\sigma)$, since, to obtain σ from τ, we need to move each of the k indices $i_1, ..., i_k$ past the $n-k$ many indices $i_{k+1}, ..., i_n$. Thus, we have $\perp\perp \mathbf{v} = (-1)^{k(n-k+1)} \mathrm{sgn}(\sigma)^2 \hat{\mathbf{v}} = (-1)^{k(n-k+1)} \hat{\mathbf{v}}$. □

Theorem 5.87. Let V be an n-dimensional vector space with metric tensor g. For all $k \leq n$ is a metric tensor $\widetilde{g}$ on $\Lambda^k(V)$ defined on elementary wedges by $\widetilde{g}(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k, \mathbf{w}_1 \wedge ... \wedge \mathbf{w}_k) = \det(g(\mathbf{v}_i, \mathbf{w}_i))$.

Theorem 5.88. Let V be an n-dimensional vector space with metric tensor g. If $k \leq n$, then for all $\omega, \eta \in \Lambda^k(V)$ we have

$$\omega \wedge (\perp \eta) = \widetilde{g}(\omega, \eta) \omega_{\mathrm{vol}},$$

where ω_{vol} is the volume form on V and $\widetilde{g}$ is the metric tensor on $\Lambda^k(V)$ from the previous definition. Therefore, the Hodge dual is basis independent and thus well-defined.

Part II

Calculus and basic topology

6

Review of calculus

This chapter presents a review of calculus, particularly multivariable 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 Notational conventions in single-variable calculus

This section formalizes common notational conventions used in single-variable calculus. These conventions ripple up to multivariable calculus, so they are worth reviewing.

First, we must establish some unambigiuous notation.

Definition 6.1. (The derivative).

$$f'(t) := \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}.$$

We use the above "prime" notation as a starting point for explaining the various ways of notating derivatives.

Definition 6.2. (Common notation for derivatives).

Suppose $U \subseteq \mathbb{R}$ is an open set and $f : U \subseteq \mathbb{R} \to \mathbb{R}$ is a differentiable function. We define the *Leibniz* and *operator* notations for differentiation.

Leibniz notation

$$\frac{df}{dt} := f'$$

$$\frac{df}{dt}\Big|_{t_0} = f'(t_0)$$

Operator notation

$$\frac{d}{dt} f := f'$$

$$\frac{d}{dt} f(t) := \Big(\frac{d}{ds} f\Big)\Big|_{s=t} = f'(t)$$

$$\frac{df(t)}{dt} := \frac{d}{dt} f(t)$$

Remark 6.3. ($\frac{df}{dt}$ vs. $\frac{df(t)}{dt}$).

It is important not to confuse $\frac{df}{dt}$, which is a function, with $\frac{df(t)}{dt}$, which is the result of evaluating $\frac{df}{dt}$ at t.

A more pedantic distinction is that if $t \neq a \neq b \neq c \neq ... \neq z$, then $\frac{df}{dt} = \frac{df}{da} = \frac{df}{db} = \frac{df}{dc} = ... = \frac{df}{dz}$, as the letter in the "denominator" is just a "piece of notation". Contrastingly, if $t \neq a \neq b \neq c \neq ... \neq z$, it is not necessarily true that $\frac{df(t)}{dt} = \frac{df(a)}{da} = \frac{df(b)}{db} = \frac{df(c)}{dc} = ... = \frac{df(z)}{dz}$. This is because the letter in the "denominator" of $\frac{df(t)}{dt}$ is used as input; $\frac{df(t)}{dt}$ is the result of taking the derivative of f and plugging in t.

Definition 6.4. (Derivative with respect to a function).

This definition formalizes a convention that is often used but rarely explained.

Suppose $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ and $x : V \subseteq \mathbb{R} \to \mathbb{R}$ satisfy the differentiability conditions of the chain rule, so that $f \circ x$ is differentiable. We define $\frac{df}{dx} : U \to \mathbb{R}$ to be the function defined by

$$\frac{df}{dx}\Big|_s := \frac{df}{dt}\Big|_{t=x(s)} = f'(x(s)).$$

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:

$$\frac{d(f \circ x)}{dt}\Big|_t = \frac{df(u)}{du}\frac{du(t)}{dt}, \text{ where } u = x(t).$$

This above statement is bad for two reasons. Firstly, when applying the above statement to a composite function such as $f(t) = \sin^2(t)$ (here, $f(x) = x^2$ and $x(t) = \sin(t)$), one is not allowed to plug $x(t)$ in for u until a closed form has been found for $\frac{df(u)}{du}$. If one were to plug in $x(t)$ for u before such a closed form were found, then they would obtain an expression involving $\frac{df(x(t))}{dx(t)}$, which has no meaning. (Yes, $\frac{df(x(t))}{dx(t)}$ is reminiscent of the notion of "derivative with respect to a function", defined above, but no calculus textbook I have seen actually defines this notion[1]!). Secondly, since one obtains an expression involving $\frac{df(x(t))}{dx(t)}$ when trying to reason about functions in the abstract, this statement of the chain rule will not work for all scenarios.

Remark 6.5. (Letters in the denominator).

The definition $\frac{df}{dt} := f'$ from Definition 6.2 technically implies that $\frac{df}{da} = \frac{df}{db} = \frac{df}{dc} = ... = \frac{df}{dz} = f'$; it does not matter which letter is used in the "denominator".

On the other hand, when the letter in the "denominator" represents a function $\mathbb{R} \to \mathbb{R}$, 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 $f : U \subseteq \mathbb{R} \to \mathbb{R}$ 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 $f : U \subseteq \mathbb{R} \to \mathbb{R}$ and $x : V \subseteq \mathbb{R} \to \mathbb{R}$ 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.

[1]And, anyways, $\frac{df(x(t))}{dx(t)}$ looks more like a "derivative with with respect to a function that has been evaluated at an input". I suppose one could define "derivative with with respect to a function that has been evaluated at an input" to mean "derivative with with respect to a function, where the final result is evaluated at that input". The point stands, this substitution business is ugly!

6.2 Multivariable calculus

Lemma 6.6. (Multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}$).

Let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ be sufficiently differentiable, and set $\mathbf{x}(t_0) = \mathbf{p}$. It can be proved that

$$\frac{d(f \circ \mathbf{x})}{dt}\Big|_{t_0} = \frac{\partial f(\mathbf{x})}{\partial x_1}\Big|_{\mathbf{p}} \frac{dx_1}{dt}\Big|_{t_0} + ... + \frac{\partial f(\mathbf{x})}{\partial x_n}\Big|_{\mathbf{p}} \frac{dx_n}{dt}\Big|_{t_0}.$$

In other words,

$$\frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{p}} = (\underset{\mathbf{x}}{\nabla} f)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}(t)}{dt}\Big|_{t_0},$$

where we have defined the *gradient of f with respect to the function* $\mathbf{x}$ to be

$$\boxed{\nabla_{\mathbf{x}} f := \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}}$$

Note that since $\mathbf{x}(\mathbf{p}) = (x_1(\mathbf{p}), ..., x_n(\mathbf{p}))^\top$, 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 6.4. 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

$$\boxed{\frac{d(f \circ \mathbf{x})}{dt} = (\underset{\mathbf{x}}{\nabla} f) \cdot \frac{d\mathbf{x}(t)}{dt}}$$

Definition 6.7. (Directional derivative of a differentiable function $\mathbb{R}^n \to \mathbb{R}$).

Consider $f : \mathbb{R}^n \to \mathbb{R}$. Let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be the curve with $\mathbf{x}(t_0) = \mathbf{p}$ and $\frac{d\mathbf{x}}{dt}\Big|_{t_0} = \mathbf{v}$. Let $\mathbf{x}$ and f be sufficiently differentiable. We define the *directional derivative* $\frac{\partial f}{\partial \mathbf{v}}$ *of f in the direction of* $\mathbf{v}$ to be

$$\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} := \frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{p}} = (\underset{\mathbf{x}}{\nabla} f)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}(t)}{dt}\Big|_{t_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v}.$$

Therefore the directional derivative is expressed as

$$\boxed{\begin{aligned} &\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} \\ &\frac{\partial f}{\partial \mathbf{v}} = \nabla f \cdot \mathbf{v} \end{aligned}}$$

In the second line, we interpret ∇ as the function sending $\mathbf{x} \mapsto \nabla_{\mathbf{x}}$.

Most authors denote $\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$ as $D_{\mathbf{p}} f(\mathbf{v})$ or as $Df[\mathbf{v}](\mathbf{p})$.

Remark 6.8. [Lee, p. 282, 283] (Infinitesimal displacement versus infinitesimal time).

We typically think of the directional derivative $\frac{\partial f}{\partial \mathbf{v}}$ as being the "infinitesimal change" in f resulting from traveling an "infinitesimal displacement" in the direction of $\mathbf{v}$, even though the directional derivative is defined with the notion of "infinitesimal time" (the time derivative). So, in some sense, the infinitesimal displacement interpretation is indirect, since infinitesimal time is lurking behind the scenes.

Actually, there is there is a way to "directly" justify the infinitesimal displacement notion! The statement of the multivariable Taylor theorem does all the work for us: if $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, then $(\Delta_{\mathbf{v}} f)|_{\mathbf{p}} = f(\mathbf{p} + \mathbf{v}) - f(\mathbf{p})$ is well approximated as $\Delta_{\mathbf{v}} f \approx \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} ([\mathbf{v}]_{\hat{\mathfrak{E}}})_i = (\nabla_{\mathbf{x}} f) \cdot \mathbf{v} = (D_{\mathbf{p}} f)(\mathbf{v})$ when $||\mathbf{v}||$ is small.

Theorem 6.9. (Directional derivative as a limit).

It is sometimes useful to express the directional derivative as a limit.

To do so, first consider the curve $\mathbf{x}$ from the definition of the directional derivative. Since $\mathbf{x}(t_0) = \mathbf{p}$ and $\frac{d\mathbf{x}}{dt}\Big|_{t_0} = \mathbf{v}$, then $\mathbf{x}(t) = \mathbf{p} + \mathbf{v}t$. Thus

$$\begin{aligned}\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} &= \frac{d(f \circ \mathbf{x})(t)}{dt}\Big|_{t_0} = \frac{df(\mathbf{p}+\mathbf{v}t)}{dt}\Big|_{t_0} = \lim_{h \to 0}\Big(\frac{f(\mathbf{p}+\mathbf{v}(t+h)) - f(\mathbf{p}+vt)}{h}\Big)\Big|_{t_0} \\ &= \lim_{h \to 0}\Big(\Big(\frac{f(\mathbf{p}+\mathbf{v}(t+h)) - f(\mathbf{p}+vt)}{h}\Big)\Big|_{t_0}\Big) = \lim_{h \to 0}\frac{f(\mathbf{p}+h\mathbf{v}) - f(\mathbf{v})}{h}.\end{aligned}$$

Therefore

$$\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} = \lim_{h \to 0}\frac{f(\mathbf{p}+h\mathbf{v}) - f(\mathbf{v})}{h}.$$

Many authors define the directional derivative using this formula.

Theorem 6.10. (Gradient is direction of greatest increase).

Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a sufficiently differentiable function. Then, at each $\mathbf{p} \in R^3$, the gradient $(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}$ is the direction of greatest increase in f at $\mathbf{p}$. When $||\mathbf{v}|| = 1$, then $||(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}||$ is the magnitude of this greatest increase.

Proof. The previous theorem shows that $\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v}$. We know $(\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} = ||(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}||\ ||\mathrm{proj}(\mathbf{v} \to (\nabla_{\mathbf{x}} f)|_{\mathbf{p}})||$. The dot product is maximized when the projection of $\mathbf{v}$ onto $(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}$ is $\mathbf{v}$ itself. Thus, when the directional derivative is maximized, $\mathbf{v} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}}$.

The magnitude of this maximal directional derivative is $(\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} = ||(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}||\ ||\mathrm{proj}(\mathbf{v} \to (\nabla_{\mathbf{x}} f)|_{\mathbf{p}})||$. When $||\mathbf{v}|| = 1$, this reduces to $||(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}||$. □

Remark 6.11. (Directional derivative simplifies to partial derivative).

We have $\frac{\partial}{\partial \hat{\mathfrak{e}}_i} = \frac{\partial}{\partial x_i}$.

Lemma 6.12. (Multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}^m$).

Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be sufficiently differentiable, and set $\mathbf{p} = \mathbf{x}(t_0)$.

$$\frac{d(\mathbf{f} \circ \mathbf{x})(t)}{dt}\Big|_{\mathbf{p}} = \begin{pmatrix} \frac{d}{dt} f_1(\mathbf{x}(t)) \\ \vdots \\ \frac{d}{dt} f_m(\mathbf{x}(t)) \end{pmatrix} = \begin{pmatrix} (\nabla_{\mathbf{x}} f_1)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}}{dt}\Big|_{t_0} \\ \vdots \\ (\nabla_{\mathbf{x}} f_m)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}}{dt}\Big|_{t_0} \end{pmatrix} = \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix}\Bigg|_{\mathbf{p}} \frac{d\mathbf{x}}{dt}\Big|_{t_0}.$$

In terms of functions, we have

$$\frac{d(\mathbf{f} \circ \mathbf{x})(t)}{dt} = \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix} \frac{d\mathbf{x}}{dt}$$

Recall from Derivation **??** and Theorem 2.98 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:

$$\begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix} \frac{d\mathbf{x}}{dt} = \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \cdot \frac{d\mathbf{x}}{dt} \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \cdot \frac{d\mathbf{x}}{dt} \end{pmatrix} = \sum_{i=1}^{n} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_i} \frac{dx_i(t)}{dt}.$$

Definition 6.13. (The Jacobian).

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 6.4), we define the *Jacobian matrix* $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ to be

$$\boxed{\frac{\partial \mathbf{f}}{\partial \mathbf{x}} := \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix}}$$

Alternative notation for this matrix is $\left(\frac{\partial f_i}{\partial x_j}\right)$.

Using the Jacobian, the multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}^m$ is now succintly stated as

$$\boxed{\frac{d(\mathbf{f} \circ \mathbf{x})}{dt} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt}}$$

Definition 6.14. (Directional derivative of a differentiable function $\mathbb{R}^n \to \mathbb{R}^m$).

The directional derivative of a differentiable function $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is defined analogously to that of a differentiable function $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}$. Indeed, in the special case of $m = 1$, the two definitions are equivalent.

As was done previously, let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be the curve with $\mathbf{x}(t_0) = \mathbf{p}$ and $\frac{d\mathbf{x}}{dt}\Big|_{t_0} = \mathbf{v}$. We define the *directional derivative* $\frac{\partial f}{\partial \mathbf{v}}$ *of* f *in the direction of* $\mathbf{v}$ to be

$$\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\Big|_{\mathbf{p}} := \frac{d(\mathbf{f} \circ \mathbf{x})}{dt}\Big|_{\mathbf{p}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{p}} \frac{d\mathbf{x}}{dt}\Big|_{t_0}$$

So this most general definition of directional derivative is expressed as

$$\boxed{\begin{aligned} \frac{\partial \mathbf{f}}{\partial \mathbf{v}}\Big|_{\mathbf{p}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{p}} \frac{d\mathbf{x}}{dt}\Big|_{t_0} \\ \frac{\partial \mathbf{f}}{\partial \mathbf{v}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt} \end{aligned}}$$

Remark 6.15. (Linearity with respect to the $\mathbf{v}$ in the denominator).

The directional derivative $\frac{\partial \mathbf{f}}{\partial \mathbf{v}}$ is linear with respect to $\mathbf{v}$.

A technical theorem

Theorem 6.16. (The integral is "linear" with respect to the region of integration).

If $D = \cup_{i=1}^k D_i$, where $D_1, ..., D_k$ are domains of integration such that $D_i \cap C_j$ has measure zero[2] for all $i \neq j$, then

$$\int_D f = \sum_{i=1}^k \int_{D_i} f.$$

[2]Informally, a subset of $\mathbb{R}^n$ has *measure zero* iff its volume is zero.

7

Basic topology

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 $\mathbb{R}^n$" at each point.

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 reader 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 "what is going on".

The standard topology on $\mathbb{R}$

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 how many sets occur in the union: there can be finitely many, countably infinitely many, or uncountably infinitely many.

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 $\exists$open $U_x \ni x$ $U_x \subseteq A$.

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 $\mathbb{R}$).

A subset $A \subseteq \mathbb{R}$ is open iff each $x \in A$ is an interior point, i.e., iff $A = \text{int}(A)$. Equivalently, A is open iff $A \subseteq \text{int}(A)$, since $A \supseteq \text{int}(A)$ is always true.

Proof. We show that (A is open) $\iff$ (each $x \in A$ 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!

$$
\begin{aligned}
&U \text{ is open} \\
U = \text{int}(U) \iff U \subseteq \text{int}(U) \iff &\forall x \in U \; x \in \text{int}(U) \\
&\forall x \in U \; \exists \text{open } U_x \ni x \; U_x \subseteq U \\
&\forall x \in U \; \exists \text{open } U_x \ni x \; U_x \cap U^c = \emptyset \\
&\forall x \; x \notin U^c \implies \text{not}(\forall \text{open } U_x \ni x \;\; U_x \cap U^c \neq \emptyset) \\
&\forall x \; x \notin U^c \implies \text{not}(\forall \text{open } U_x \ni x \;\; U_x \cap U^c - \{x\} \neq \emptyset) \\
&\forall x \; x \notin U^c \implies x \notin (U^c)' \\
&\forall x \; x \in (U^c)' \implies x \in U^c \\
&(U^c)' \subseteq U^c.
\end{aligned}
$$

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. Line 7 implies line 8 due to the contrapositive.

In line 7, the notation "$(U^c)'$" relies upon the notion of a limit point of a subset of $\mathbb{R}$, in the following way. We say $x \in \mathbb{R}$ is a *limit point of* $A \subseteq \mathbb{R}$ iff $\forall$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 $\forall$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 will be 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".

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 in $\mathbb{R}$ is an open set in $\mathbb{R}$. DeMorgan's laws then imply that an arbitrary intersection of closed sets in $\mathbb{R}$ is a closed set in $\mathbb{R}$.

What about intersections of open sets $\mathbb{R}$- or, equivalently, by DeMorgan's laws- unions of closed sets in $\mathbb{R}$? Well, any infinite union of closed sets in $\mathbb{R}$ 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 (i.e., perhaps a finite union of sets that contain their limit points itself contains all its limit points)? This is indeed the case.

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 7.10. (Covers and generating sets).

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 $U = \cup_{\alpha \in I} C_\alpha$, where $\{C_\alpha\}_{\alpha \in I} \subseteq \mathcal{C}$.

We can now define topological spaces.

Definition 7.11. (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.
    - Note: 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 general version of the interior point characterization of open sets, which is in turn equivalent to the general version of the limit point characterization of closed sets.
        - Interior points and limit points will be defined analogously as to how they were before, and the "general versions" of the interior and limit point characterizations, which are obtained from the versions we've seen by substituting in a general topological space X in for where we before wrote $\mathbb{R}$, are indeed true due to analogous arguments.
    - (1) is also equivalent to: "arbitrary unions of open sets are open", which is equivalent to "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 7.12. ($\mathbb{R}^n$ with the standard topology).

In the section "Topology on $\mathbb{R}$", we were really doing topology on $(\mathbb{R}, \text{std})$, where std is the *standard topology on* $\mathbb{R}$, std $:= \{$subsets of $\mathbb{R}$ that are arbitrary unions of open intervals in $\mathbb{R}\}$. In general, the *standard topology on* $\mathbb{R}^n$ is $\{$subsets of $\mathbb{R}^n$ that are arbitrary unions of open intervals in $\mathbb{R}^n\}$.

Definition 7.13. (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 7.14. ($\emptyset, X \in \tau$).

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. Obviously, $X \in \tau$, since τ generates X. It's less easy to see that $\emptyset$ is also generated by τ, but this is true too, since the empty set is the empty union of any collection of open sets (an empty union is a union that employs the empty set as its indexing set).

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 finite intersections of elements of $\mathcal{B}$ are guaranteed to also be in $\mathcal{B}$. If we find such requirements, then $\mathcal{B}$ will induce a topology on X, in the sense that the set τ whose elements are arbitrary unions of elements of $\mathcal{B}$ will be a topology on X. The following definition and theorem present this condition.

Definition 7.15. (Refining with interior points).

If $\mathcal{B}$ is a cover of some set X, we say that $\mathcal{B}$ *refines with interior points* iff
$\forall B_1, B_2 \in \mathcal{B} \;\; x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B} \; x \in B_3$.

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.

Proof.

($\Longrightarrow$). Assume that finite intersections of open sets are 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, it is a union of basis elements, $U = \cup_{i \in \{3,4,...\}} B_i$. Since x is in U, it must be in B_j, for some $j \in \{3, 4, ...\}$. Relabel B_j to be B_3 to match the syntax of the theorem.

($\Longleftarrow$). Assume $\mathcal{B}$ is a basis; that is, assume that $\mathcal{B}$ 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 $V = \cap_{i=1}^n U_i$, 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 $x \in \cap_{i=1}^n B_i$. Using induction on the fact that $\mathcal{B}$ refines with interior points, there is a $B_x \in \mathcal{B}$ $x \in B_x \subseteq \cap_{i=1}^n B_i$. We have $\cap_{i=1}^n B_i = \cup_{x \in V} B_x$, so $\cap_{i=1}^n B_i$ 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
$\forall B_1, B_2 \in \mathcal{B}\ \ x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B}\ x \in B_3$.

The reader may wonder whether topological bases satisfy a condition similar to linear independence, since bases of vector spaces are linearly independent sets. They do not; topological bases are not like bases of vector spaces in this regard.

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 $\exists$open $U_x \ni x\ U_x \subseteq A$.

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, i.e., iff $A = \text{int}(A)$. Equivalently, A is open iff $A \subseteq \text{int}(A)$, since $A \supseteq \text{int}(A)$ is always true.

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 $\forall$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.

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 7.25. (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 7.26. Let (X, τ_X) be a topological space, and consider $Y \subseteq X$. If $A \subseteq Y$, then A is closed in Y iff $A = C \cap Y$, where C 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 that is open in X. Thus $A = Y - (Y - A) = Y - (Y \cap U) = Y - U = Y \cap (X - U)$. So, $A = (X - U) \cap Y$, where $X - U$ is closed in X.

($\impliedby$). Suppose $A = C \cap Y$, where C is closed in X. Now, we show that $Y - A$ is open in Y. We have $Y - A = Y - (C \cap Y) = Y \cap (X - C)$. Thus $Y - A = (X - C) \cap Y$, where $X - C$ is open in X, meaning $(X - C) \cap Y$ is open in Y. $\square$

Interior, closure, and boundary

For the purposes of differential forms, understanding this section in depth is not necessary. Some familiarity with this section on a intuitive level is required, though.

Theorem 7.27. (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 X. $\square$

Definition 7.28. (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 7.29. (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 7.30. (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 7.31. (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 and homeomorphisms

Limits and continuity for functions $A \subseteq \mathbb{R} \to \mathbb{R}$

Definition 7.32. (Open balls).

We denote the *open ball in* $\mathbb{R}^n$ *of radius* r *centered at* $\mathbf{x} \in \mathbb{R}^n$ by $B(r, x)$: $B(r, \mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n \mid d(\mathbf{x}, \mathbf{y}) < r\}$, where $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|$, and where $|| \cdot ||$ is the norm on $\mathbb{R}^n$ induced by the dot product on $\mathbb{R}^n$.

Definition 7.33. (Limit of a function $A \subseteq \mathbb{R} \to \mathbb{R}$).

Let $A \subseteq \mathbb{R}$, and consider a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$. We write $\lim_{x \to x_0} f(x) = L$ iff $\forall \epsilon > 0 \; \exists \delta > 0 \; x \in B(\delta, x_0) - \{x_0\} \implies f(x) \in B(\epsilon, L)$.

Definition 7.34. (Continuity of a function $A \subseteq \mathbb{R} \to \mathbb{R}$ at a point).

Let $A \subseteq \mathbb{R}$, and consider a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$. We say that f *is continuous at* $x_0 \in A$ iff $\forall \epsilon > 0 \; \exists \delta > 0 \; x \in B(\delta, x_0) \implies f(x) \in B(\epsilon, f(x_0))$

Theorem 7.35. (Condition for continuity for functions $A \subseteq \mathbb{R} \to \mathbb{R}$).

Let $A \subseteq \mathbb{R}$. A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is continuous at $x_0 \in A$ iff $\lim_{x \to x_0} f(x) = f(x_0)$.

Proof.

($\implies$). If f is continuous at x_0, then we immediately know $\lim_{x \to x_0} f(x) = f(x_0)$ 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. $x \in B(\delta, x) \cap A - \{x_0\} \implies f(x) \in B(\epsilon, f(x_0))$. 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 7.36. (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 $f : A \subseteq X \to Y$. We write $\lim_{x \to x_0} f(x) = L$ iff $\forall$open $V \ni L, V \subseteq Y$ $\exists$open $U \ni x_0, U \subseteq X$ $f(U - \{x_0\}) \subseteq V$.

Definition 7.37. (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 $f : A \subseteq X \to Y$. We say that f *is continuous at* $x_0 \in A$ iff $\forall$open $V \ni f(x_0), V \subseteq Y$ $\exists$open $U \ni x_0, U \subseteq X$ $f(U) \subseteq V$.

Remark 7.38. (Topological limits and continuity generalize real analytical notions).

Note, the earlier definitions of the limit of a function $A \subseteq \mathbb{R} \to \mathbb{R}$ and continuity at a point for a function $A \subseteq \mathbb{R} \to \mathbb{R}$ can both be viewed as a special case of the just-stated corresponding definitions for 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 7.39. (Condition for continuity for functions on subsets of topological spaces).

Let X and Y be topological spaces, and let $A \subseteq X$. A function $f : A \subseteq X \to Y$ is continuous at $x_0 \in A$ iff $\lim_{x \to x_0} f(x) = f(x_0)$.

Proof.

($\implies$). If f is continuous at x_0, then we immediately know $\lim_{x \to x_0} f(x) = f(x_0)$ because $f(U) \subseteq V$ is a weaker condition than $f(U - \{x_0\}) \subseteq V$.

($\impliedby$). Suppose that $\forall$open $V \ni f(x_0), V \subseteq Y$ $\exists$open $U \ni x_0, U \subseteq X$ $f(U - \{x_0\}) \subseteq V$. 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 7.40. (Continuous function).

Let X and Y be topological spaces, and let $A \subseteq X$. A function $f : A \subseteq X \to Y$ is called *continuous* iff it is continuous at every $x \in A$.

Theorem 7.41. (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 $f : A \subseteq X \to Y$ 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. Left as an exercise. □

Homeomorphisms

Definition 7.42. (Homeomorphism).
Let (X, τ_X) and (Y, τ_Y) be topological spaces. A *homeomorphism* is an injective function $X \to Y$ that is continuous and has a continuous inverse.

Compact and Hausdorff topological spaces

This section briefly details the last two topological notions we require to define manifolds.

Compactness

Definition 7.43. (Compact topological space).
Let (X, τ) be a topological space. A cover $\mathcal{C}$ for X is said to be an *open cover* iff every set $A \in \mathcal{C}$ is open. We say that (X, τ) is *compact* iff for every open cover $\mathcal{O}$ of X, there exists a finite open cover $\mathcal{O}'$ of X with $\mathcal{O}' \subseteq \mathcal{O}$. More succinctly, (X, τ) is *compact* iff every open cover of X admits a finite subcover.

Definition 7.44. (Compactness in $\mathbb{R}^n$ with the standard topology).
A set $A \subseteq \mathbb{R}^n$, where $\mathbb{R}^n$ has the standard topology, is compact iff A is closed and bounded[1].

Proof. See any textbook on topology. □

Hausdorff spaces

We need to know what Hausdorff spaces are because manifolds are special types of Hausdorff topological spaces.

Definition 7.45. (Hausdorff space).
Let (X, τ) be a topological space. We say that (X, τ) *is Hausdorff* iff $\forall x \in X\ \forall y \in X\ \exists \text{open } U, V\ U \cap V = \emptyset$. Intuitively, a topological space is Hausdorff iff every two points in that space can be "separated" by taking open sets about each point.

[1] A subset of $\mathbb{R}^n$ is *bounded* iff there exists some open ball that contains it.

Part III

Differential forms

8

Manifolds

This chapter follows various parts of [Lee] and sets the stage for differential forms. Differential forms will live on and be integrated over *smooth manifolds*, which can be thought of as "multidimensional surfaces". We begin the chapter by working simply with *manifolds*, then introduce the notions of *boundary* and *corners* on a manifold.

We will discuss the *tangent space* at a point on a manifold, and use the machinery of *tangent vectors* to set up a coordinate-free method of differentiation on smooth manifolds, called the *differential.* In Theorems 8.39 and 8.40, we see a duality between coordiantes of tangent vectors and *tangent covectors* that is reminiscent of the duality between coordinates of vectors and covectors. Lastly, we discuss orientations of manifolds in preparation for the integration of the next chapter.

8.1 Introduction to manifolds

Definition 8.1. [Lee, p. 2] (Manifold).

An *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

$$\mathbb{H}^n := \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^n \geq 0\},$$

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 Section 7.1 of Chapter 7), are

$$\text{int}(\mathbb{H}^n) = \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^n > 0\}$$
$$\partial\mathbb{H}^n = \{(x^1, ..., 0) \in \mathbb{R}^n\}.$$

Recall from Section 7.1 of Chapter 7 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 $\mathbf{p} \in \partial H^n$.

Now let M be a manifold, and suppose that for some open subset $U \subseteq M$ containing a point $\mathbf{p} \in U$, there is a homeomorphism $\mathbf{x} : U \subseteq M \to \mathbb{R}^n$, where $\mathbf{x}(\mathbf{p}) = (x^1(\mathbf{p}), ..., x^{n-1}(\mathbf{p}), 0)$ is in $\partial \mathbb{H}^n$. Then, in analogy to the notion of "locally Eulidean" 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 Section 7.1 of Chapter 7). 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. For example, 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.

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 $\mathbf{x} : U \to V \subseteq \mathbb{R}^n$ 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. [Lee, p. 33] (Coordinate representations of functions on manifolds).

Let M and N be manifolds, and consider a function $\mathbf{F} : M \to N$. Let $(U, \mathbf{x})$ be a chart on M about $\mathbf{p} \in M$, and let $(V, \mathbf{y})$ be a chart on N about $\mathbf{F}(\mathbf{p}) \in N$. The *coordinate representation of* $\mathbf{F} : U \subseteq M \to V \subseteq N$ *relative to the charts* $(U, \mathbf{x})$ *and* $(V, \mathbf{y})$ is the function $\widetilde{\mathbf{F}}_{(U,\mathbf{x}),(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^n \to \mathbb{R}^m$.

In the case that M or N is $\mathbb{R}^k$, we do not require a chart on whichever manifold is $\mathbb{R}^k$, and the coordinate representation of $\mathbf{F}$ is said to be "relative" to only a single chart. In these cases, we the coordinate representation is either $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^k \to N$ or $\widetilde{\mathbf{F}}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{F} : M \to \mathbb{R}^k$.

8.3 Smooth manifolds

Definition 8.10. [Lee, p. 11] (Differentiability classes, smooth functions, and diffeomorphisms).

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be of *differentiability class* C^k iff f's partial derivatives of orders $0, 1, ..., k$ are all 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 $f : M \to \mathbb{R}$ has *differentiability class* C^∞ iff $f \in C^k(M)$ for all $k \in \{0, 1, ...\}$.

A function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^k$ is said to be of *differentiability class* C^k iff each component function F^i of $\mathbf{F}$ is of differentiability class C^k, in the previous sense of C^k for functions $\mathbb{R}^n \to \mathbb{R}$. Similarly, a function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ is said to be of *differentiability class* C^∞ iff each component function F^i of $\mathbf{F}$ is of differentiability class C^∞, in the previous sense of C^∞ for functions $\mathbb{R}^n \to \mathbb{R}$.

We define $C^k(\mathbb{R}^n \to \mathbb{R}^k)$ to be the set of functions $\mathbb{R}^n \to \mathbb{R}^k$ that are of differentiability class C^k, and define $C^\infty(\mathbb{R}^n \to \mathbb{R}^k)$ to be the set of functions $\mathbb{R}^n \to \mathbb{R}^k$ that are of differentiability class C^∞. We will use the word "smooth" to mean C^∞. Following this convention, the set $C^\infty(\mathbb{R}^n \to \mathbb{R})$ is called the set of *smooth real-valued functions on* $\mathbb{R}^n$.

Lastly, we say that a function $\mathbb{R}^n \to \mathbb{R}$ is a *diffeomorphism* iff it is smooth, bijective, and has a smooth inverse. Note that every diffeomorphism is a topological homeomorphism.

Smooth manifolds with or without boundary

Derivation 8.11. [Lee, p. 27, 28] (Smooth manifold with or without boundary).

Let M be an n-manifold with or without boundary. We say that a function $\mathbf{F} : M \to \mathbb{R}^k$ is *smooth relative to a chart* $(U, \mathbf{x})$ of M iff the coordinate representation $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1} : \mathbf{x}(U) \subseteq \mathbb{R}^n \to \mathbb{R}^k$ is a smooth function.

Consider two charts $(U, \mathbf{x})$ and $(V, \mathbf{y})$. If $U \cap V \neq \emptyset$, then the *transition map* $\mathbf{y} \circ \mathbf{x}^{-1} : \mathbf{x}(U \cap V) \to \mathbf{y}(U \cap V)$ is a homeomorphism. We would like to define a notion of *smooth* so that if M is a *smooth manifold with or without boundary*, then

$$\mathbf{F} : M \to \mathbb{R}^k \text{ is smooth relative to some chart } (U, \mathbf{x})$$
$$\Longrightarrow$$
$$\mathbf{F} \text{ is smooth relative to all other charts } (V, \mathbf{y}) \text{ that overlap } (U, \mathbf{x}),\ U \cap V \neq \emptyset.$$

In our present situation, this is not the case. If $(U, \mathbf{x})$ and $(V, \mathbf{y})$ intersect, $U \cap V \neq \emptyset$, 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 $\widetilde{\mathbf{F}}_{(U,\mathbf{x})}$ as $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \widetilde{\mathbf{F}}_{(V,\mathbf{y})} \circ (\mathbf{y} \circ \mathbf{x}^{-1})$. If $\widetilde{\mathbf{F}}_{(V,\mathbf{y})}$ is smooth, then $\widetilde{\mathbf{F}}_{(U,\mathbf{x})}$ is not guaranteed to be smooth, since composing with a homeomorphism (such as $\mathbf{y}$ or $\mathbf{x}^{-1}$) does not preserve smoothness.

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 compatible[1]. A chart that is an element of a smooth atlas is called a *smooth chart*.

Definition 8.12. (Smooth manifold with or without boundary)

A manifold is called a *smooth manifold with or without boundary* iff every atlas of that manifold is a smooth atlas.

Definition 8.13. (Smooth functions on manifolds).

Let M be an n-manifold, and let $(U, \mathbf{x})$ be a chart on M. A function $\mathbf{F} : U \subseteq M \to \mathbb{R}^n$ is said to be of *differentiability class* C^k *on* U iff the coordinate representation $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^n \to \mathbb{R}$ is of differentiability class C^k, and is said to be of *differentiability class* C^∞ *on* U, or to be *smooth*, iff the coordinate representation $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^n \to \mathbb{R}$ is of differentiability class C^∞. The sets of functions on U of differentiability class C^k and C^∞ are denoted $C^k(U \to \mathbb{R}^n)$ and $C^\infty(U \to \mathbb{R}^n)$, respectively.

Example 8.14. [Lee, p. 20] Graph of smooth functions into $\mathbb{R}^k$ are smooth manifolds.

[1]Note, 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.

Manifolds with or without boundary or corners

Definition 8.15. [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

$$\text{cl}(\mathbb{H}^n) = \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^i \geq 0 \text{ for all } i\}.$$

We say that a topological space M is a *manifold with or without boundary or corners* iff it 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 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 8.16. [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 8.17. [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 8.18. [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...

- *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 cl*$(\mathbb{H}^n)$; a point $\mathbf{p} \in \text{cl}(\mathbb{H}^n)$ is a *corner point of 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 8.19. (Manifolds with/without boundary or corners).

We will frequently use the acronym "WWBOC" as shorthand to mean "with/without boundary or corners".

Remark 8.20. [Lee, p. 415] (Manifolds with corners are *topologically* the same as manifolds with boundary).

The title of this remark 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 the "smoothly compatible" criterion for charts with corners is different from the analagous criterion for boundary charts.

8.4 Tangent vectors

In multivariable calculus, one can consider tangent vectors that reside in plane that is tangent to a surface. We will present a coordinate-free generalization of this concept for manifolds. Surprisingly, we will see that tangent vectors anchored at a point can be identified with "directional derivatives" at that point.

Tangent vectors in $\mathbb{R}^n$

To start, we formalize the notion of what it means to consider a vector that is anchored at a point.

Definition 8.21. [Lee, p. 51] (Vectors "anchored" at a point).

Let M be a manifold WWBOC. Given $\mathbf{p} \in M$, we define $\mathbb{R}^n_{\mathbf{p}}$ to be the vector space $\mathbb{R}^n \times \{\mathbf{p}\}$. We use the notation $\mathbf{v}_{\mathbf{p}} := (\mathbf{v}, \mathbf{p})$ to denote a typical element of $\mathbb{R}^n_{\mathbf{p}}$. Elements of $\mathbb{R}^n_{\mathbf{p}}$ are often called "geometric tangent vectors".

With the previous notion formalized, we now discover, strangely enough, that geometric tangent vectors can be identified with "directional derivative functions".

Theorem 8.22. (Geometric tangent vectors are naturally isomorphic to "directional derivative functions").

Let $\frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$ denote the function $f \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}(f)$. The map $\mathbf{v}_{\mathbf{p}} \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$ is a natural linear isomorphism $\mathbb{R}^n_{\mathbf{p}} \cong T_{\mathbf{p}}(\mathbb{R}^n)$.

Proof. First, recall from calculus that $\frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$ is linear with respect to $\mathbf{v}$ (see Remark 6.15). From this it follows that $\left\{\frac{\partial}{\partial \hat{\mathfrak{e}}_1}\Big|_{\mathbf{p}}, ..., \frac{\partial}{\partial \hat{\mathfrak{e}}_n}\Big|_{\mathbf{p}}\right\}$ is a basis for the space of directional derivative functions at $\mathbf{p}$. Thus, the space of directional derivative functions at $\mathbf{p}$ is n-dimensional. Since $\mathbb{R}^n_{\mathbf{p}}$ and the space of directional derivative functions at $\mathbf{p}$ have the same dimension, we can show that $\mathbf{v} \mapsto \frac{\partial}{\partial \mathbf{v}}$ is an isomorphism by proving that it has a trivial kernel[2].

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 $E = \{\mathbf{e}^1, ..., \mathbf{e}^n\}$ be a basis for $\mathbb{R}^n$, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be its dual basis for $(\mathbb{R}^n)^*$. Note that since $\phi^{\mathbf{e}_i} : \mathbb{R}^n \to \mathbb{R}$ is defined by $\phi^{\mathbf{e}_i}(\mathbf{p}) = [\mathbf{p}]^i_E$, each $\phi^{\mathbf{e}_i}$ is a C^∞ function on $\mathbb{R}^n$.

At $\mathbf{p} \in \mathbb{R}^n$, the ith coordinate of the zero map is $\frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}(\phi^{\mathbf{e}_i})$ (see Theorem 4.28). 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 6.15), we have

$$0 = \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}(\phi^{\mathbf{e}_i}) = \frac{\partial}{\partial(\sum_j [\mathbf{v}]^j_E \mathbf{e}^j)}\Big|_{\mathbf{p}}(\phi^{\mathbf{e}_i}) = \sum_j [\mathbf{v}]^j_E \Big(\frac{\partial}{\partial \mathbf{e}^j}\Big|_{\mathbf{p}}\Big)(\phi^{\mathbf{e}_i}).$$

Since $\frac{\partial}{\partial \mathbf{e}^j}\Big|_{\mathbf{p}} = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$, the above becomes

$$0 = \sum_j [\mathbf{v}]^j_E \delta^i{}_j = [\mathbf{v}]^i_E.$$

We see $[\mathbf{v}]^i_E = 0$ for all i, so $\mathbf{v} = \mathbf{0}$, which means $\mathbf{v}_{\mathbf{p}} = \mathbf{0}_{\mathbf{p}}$. □

Now, we investigate the vector space of "directional derivative functions" at a point. By "directional derivative function", we mean a function whose input is "the function that gets differentiated"; we are considering functions that might be notated as $f \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}(f)$. While such directional derivative functions can be specified concretely (as was done in the previous sentence), we will find that they can also be described in a coordinate-independent manner, without mention of some $\mathbf{v} \in \mathbb{R}^n$.

Specifically, one can quickly check that if a function satisfies the below condition of being a *derivation*, then it is a directional derivative.

Definition 8.23. [Lee, p. 52] (Derivation at $\mathbf{p} \in \mathbb{R}^n$).

A *derivation at* $\mathbf{p} \in \mathbb{R}^n$ is a linear function $v_{\mathbf{p}} : C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}$ which satisfies a resemblance of the product rule,

[2] We adapt the argument from [Lee, p. 53] (which is intended for a slightly different purpose in that book).

$$v_{\mathbf{p}}(fg) = v_{\mathbf{p}}(f)g(\mathbf{p}) + f(\mathbf{p})v_{\mathbf{p}}(g),$$

for all $f, g \in C^\infty(\mathbb{R}^n \to \mathbb{R})$.

It is relatively quick to check that every "directional derivative function" is a derivation. In the next theorem, we additionally show that every derivation is a directional derivative. First, though, we need the following lemma.

Lemma 8.24. [Lee, p. 53] (Properties of derivations at $\mathbf{p} \in \mathbb{R}^n$).
Consider $\mathbf{p} \in \mathbb{R}^n$ and let $v_{\mathbf{p}}$ be a derivation at $\mathbf{p}$. For any $f, g \in C^\infty(\mathbb{R}^n \to \mathbb{R})$, we have the following:

- 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.

- 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 8.25. (Every directional derivative is a derivation, and every derivation is a directional derivative).
The set of directional derivatives at $\mathbf{p} \in \mathbb{R}^n$ is equal to the set of derivations at $\mathbf{p}$:

$$\Big\{\frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}} \mid \mathbf{v} \in \mathbb{R}^n\Big\} = T_{\mathbf{p}}(\mathbb{R}^n).$$

Proof.
($\subseteq$). Showing that every directional derivative at $\mathbf{p} \in \mathbb{R}^n$ is a derivation at $\mathbf{p}$ follows straightforwardly from the definition of "derivation at $\mathbf{p} \in \mathbb{R}^n$".
($\supseteq$). This direction of the proof is adapted from [Lee, p. 53]. Let $\mathbf{p} \in \mathbb{R}^n$ and let $v_{\mathbf{p}} : C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}$ be a derivation. We must show that $v_{\mathbf{p}}$ is a directional derivative at $\mathbf{p}$, and find a vector $\mathbf{v_p} \in \mathbb{R}^n_{\mathbf{p}}$ for which $v_{\mathbf{p}} = \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$. So, we will find a $\mathbf{v_p} \in \mathbb{R}^n_{\mathbf{p}}$ for which

$$v_{\mathbf{p}}(f) = \Big(\frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}\Big)(f) \text{ for all } f \in C^\infty(\mathbb{R}^n \to \mathbb{R}).$$

We use Taylor's theorem (see [Lee, p. 53]) to write $f \in C^\infty(\mathbb{R}^n \to \mathbb{R})$ 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 $x^i := [\mathbf{x}]^i_{\hat{\mathfrak{E}}}$, $p^i := [\mathbf{p}]^i_{\hat{\mathfrak{E}}}$ are the ith coordinates of $\mathbf{x}$ and $\mathbf{p}$ relative to the standard basis $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ for $\mathbb{R}^n$.
Now we produce the vector $\mathbf{v_p}$. Let $\{\phi^{\hat{\mathfrak{e}}_1}, ..., \phi^{\hat{\mathfrak{e}}_n}\}$ be the induced dual basis for $(\mathbb{R}^n)^*$. We set $\mathbf{v_p} = \sum_{i=1}^n v_{\mathbf{p}}(\phi^{\hat{\mathfrak{e}}_i})\hat{\mathfrak{e}}_i$. (Note that applying the derivation $v_{\mathbf{p}}$ to $\phi^{\hat{\mathfrak{e}}_i}$ makes sense because $\phi^{\hat{\mathfrak{e}}_i}$, being the "$i$th coordinate function on $\mathbb{R}^n$", is a smooth function $\mathbb{R}^n \to \mathbb{R}$). This choice of $\mathbf{v_p}$ makes more sense in hindsight after reading Theorem 8.40. (That theorem will logically depend on this result, though, so we must make this relatively unmotivated choice of $\mathbf{v_p}$!)
Apply $v_{\mathbf{p}}$ to f and use the second bullet point of the previous lemma to obtain

$$v_{\mathbf{p}}(f) = v_{\mathbf{p}}\Big(\sum_{i,j=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}(x^i - p^i)\Big).$$

Since $v_{\mathbf{p}}$ is a linear map and as $\mathbf{p} \mapsto \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}$ is a constant function, we can "pull out the constant" that is $\frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}$:

$$v_{\mathbf{p}}\Big(\sum_{i,j=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}(x^i - p^i)\Big) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} v_{\mathbf{p}}(x^i - p^i) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} (v_{\mathbf{p}}(x^i) - v_{\mathbf{p}}(p^i)).$$

Using the first bullet point of the previous lemma one more time, we have

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} (v_{\mathbf{p}}(x^i) - v_{\mathbf{p}}(p^i)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} v_{\mathbf{p}}(x^i).$$

Recall that 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 $\mathbf{x}$. Thus, $x^i = \phi^{\hat{\mathfrak{e}}_i}$, so the above further simplifies to

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} v_{\mathbf{p}}(x^i) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} v_{\mathbf{p}}(\phi^{\hat{\mathfrak{e}}_i}) = (\nabla \mathbf{F})_{\mathbf{p}} \cdot \mathbf{v}_{\mathbf{p}} = \frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}}.$$

In all, we have $v_{\mathbf{p}}(f) = \frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$. Thus $v_{\mathbf{p}} = \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$, so the tangent vector $v_{\mathbf{p}}$ is indeed a directional derivative. □

We now summarize the results we have discovered about $\mathbb{R}^n_{\mathbf{p}}$, the vector space of directional derivative functions at $\mathbf{p}$, and the vector space of derivations at $\mathbf{p}$.

Theorem 8.26. ($\mathbb{R}^n_{\mathbf{p}} \cong \{$directional derivative functions at $\mathbf{p}\} = \{$derivations at $\mathbf{p} \in \mathbb{R}^n\}$).
For any $\mathbf{p} \in \mathbb{R}^n$, we have

$$\boxed{\mathbb{R}^n_{\mathbf{p}} \overset{\mathbf{v}_{\mathbf{p}} \mapsto \frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}}}{\cong} \{\text{directional derivative functions at } \mathbf{p}\} = \{\text{derivations at } \mathbf{p} \in \mathbb{R}^n\}}$$

Thus, we see that we can identify elements of $\mathbb{R}^n_{\mathbf{p}}$ with the coordinate-independent mathematical objects that are derivations at $\mathbf{p} \in \mathbb{R}^n$. This motivates the following definition.

Proof. This theorem combines the results of Theorem 8.22 and Theorem 8.25 □

Definition 8.27. (Tangent space to $\mathbb{R}^n$).
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 $T_{\mathbf{p}}(\mathbb{R}^n)$. An element of $T_{\mathbf{p}}(\mathbb{R}^n)$ is called a *tangent vector to* $\mathbb{R}^n$ *(at* $\mathbf{p}$*)*.

Theorem 8.28. [Lee, p. 54] (A basis of $T_{\mathbf{p}}(\mathbb{R}^n)$).
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. We proved this in the proof of Theorem 8.22. □

Having come up with a coordinate-independent notion of "tangent vector to $\mathbf{p} \in \mathbb{R}^n$", we can extend that notion to smooth manifolds WWBOC.

Tangent vectors on manifolds

Definition 8.29. [Lee, p. 54] (Derivation at $\mathbf{p} \in M$, tangent space to a manifold).
Let M be a smooth n-manifold WWBOC. A *derivation at* $\mathbf{p} \in M$ is a linear function $v_{\mathbf{p}} : C^\infty(M \to \mathbb{R}) \to \mathbb{R}$ that satisfies the product rule,

$$v_{\mathbf{p}}(fg) = v_{\mathbf{p}}(f)g(\mathbf{p}) + f(\mathbf{p})v_{\mathbf{p}}(g) \text{ for all } f, g \in C^\infty(M \to \mathbb{R}).$$

The set of derivations at $\mathbf{p} \in M$ is called the *tangent space to* M *at* $\mathbf{p}$, and is denoted $T_{\mathbf{p}}(M)$. An element of $T_{\mathbf{p}}(M)$ is called a *tangent vector to* M *(at* $\mathbf{p}$*)*.

Theorem 8.30. [Lee, p. 54] (Properties of derivations at $\mathbf{p} \in M$).

Let M be a smooth n-manifold WWBOC, let $\mathbf{p} \in \mathbb{R}^n$, $v_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^n)$ be a tangent vector at $\mathbf{p}$, and $f, g \in C^\infty(\mathbb{R}^n \to \mathbb{R})$. 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 8.24. $\square$

Definition 8.31. [War, p. 14, 15] (Basis of $T_{\mathbf{p}}(M)$ and its abuse of notation).

Let M be a smooth n-manifold WWBOC, let $(U, \mathbf{x})$ be a smooth chart on M, and let $\mathbf{p} \in M$. In an abuse of notation, we define $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$, $i \in \{1, ..., n\}$, to be the derivations at $\mathbf{p} \in U \subseteq M$ for which

$$\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}(f) := \frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} f_{(U,\mathbf{x})} = \frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}(f \circ \mathbf{x}^{-1}).$$

In the above, $f_{(U,\mathbf{x})}$ denotes the coordinate representation of f relative to the chart $(U, \mathbf{x})$. The $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ to the left of := are what we are defining in our abuse of notation, and the $\frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial}{\partial \hat{\mathfrak{e}}^i}\Big|_{\mathbf{x}(\mathbf{p})}$ on the right hand side are directional derivatives taking in smooth functions on $\mathbb{R}^n$ as their arguments.

We will see that the $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ are derivations at $\mathbf{p} \in M$. Once we have seen this, since we know $\frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$ are derivations at $\mathbf{x}(\mathbf{p}) \in R^n$ (recall Theorem 8.28, it follows that $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}_{i=1}^n$ is a basis of $T_{\mathbf{p}}(M)$.

Proof. We need to show that the $\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}} : C^\infty(M \to \mathbb{R}) \to \mathbb{R}$ 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 $C^\infty(\mathbb{R}^n \to \mathbb{R})$. We show the product rule holds:

$$\begin{aligned}
\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}(fg) &= \frac{\partial(fg \circ \mathbf{x}^{-1})}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\Big|_{\mathbf{x}^{-1}(\mathbf{x}(\mathbf{p}))} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \\
&= \Big(\frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p})) \frac{\partial g}{\partial x^i}\Big|_{\mathbf{p}}\Big) \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \\
&= \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p})) \frac{\partial g}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \\
&= \frac{\partial(f \circ \mathbf{x})}{\partial x^i}\Big|_{\mathbf{p}} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p})) \frac{\partial(g \circ \mathbf{x})}{\partial x^i}\Big|_{\mathbf{p}} \\
&= \Big(\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}\Big)(f)\ g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\ \Big(\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}\Big)(g).
\end{aligned}$$

Here, we used the chain and product rules and then reversed the product and chain rules. $\square$

Differentials of a smooth function on a manifold

Now that we have defined tangent vectors to a manifold, we can define a generalization of the directional derivative: the *differential*. Again, we start in $\mathbb{R}^n$, but then obtain a general framework that presents differentials as functions that send tangent vectors on one smooth manifold to tangent vectors on another smooth manifold.

Definition 8.32. (Differential of a smooth function $\mathbb{R}^n \to \mathbb{R}^m$).

Consider a smooth function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$. We define the *differential* $d\mathbf{F}_\mathbf{p} : T_\mathbf{p}(\mathbb{R}^n) \to T_{\mathbf{F}(\mathbf{p})}(\mathbb{R}^m)$ *of* $\mathbf{F}$ *at* $\mathbf{p}$ to be the function $T_\mathbf{p}(\mathbb{R}^n) \to T_{\mathbf{F}(\mathbf{p})}(\mathbb{R}^m)$ that is induced by the total differential $\mathbb{R}^n \to \mathbb{R}^m$ and the isomorphisms $\mathbb{R}^n \cong T_\mathbf{p}(\mathbb{R}^n), \mathbb{R}^m \cong T_\mathbf{p}(\mathbb{R}^m)$.

Since the matrix of the total differential relative to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ is the Jacobian matrix[3]

$$\frac{\partial \mathbf{F}}{\partial \mathbf{x}}\Big|_\mathbf{p} = \Big(\frac{\partial F^i}{\partial x^j}\Big|_\mathbf{p}\Big),$$

the matrix of $d\mathbf{F}_\mathbf{p}$ relative to the bases $\{\frac{\partial}{\partial x^i}\Big|_\mathbf{p}\}$ for $T_\mathbf{p}(\mathbb{R}^n)$ and $\{\frac{\partial}{\partial y^i}\Big|_{\mathbf{F}(\mathbf{p})}\}$ for $T_{\mathbf{F}(\mathbf{p})}(\mathbb{R}^m)$ must be the same matrix.

To determine a general formula for $d\mathbf{F}_\mathbf{p}$, we will consider how $d\mathbf{F}_\mathbf{p}$ acts on a basis tangent vector $\frac{\partial}{\partial x^i}\Big|_\mathbf{p}$ and extend whatever result we get with the linearity of $d\mathbf{F}_\mathbf{p}$. Recalling from Theorem 2.110 that if $\mathbf{f} : V \to W$ is a linear function between finite-dimensional vector spaces with bases $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$, then we have $\mathbf{f}(\mathbf{e}_i) = \sum_{j=1}^n ([\mathbf{f}(E)]_F)_{ji}\mathbf{f}_j$, we see

$$d\mathbf{F}_\mathbf{p}\Big(\frac{\partial}{\partial x^i}\Big|_\mathbf{p}\Big) = \sum_{j=1}^m a_i^j \frac{\partial}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})},$$

where

$$a_j^i = \Big(\Big[d\mathbf{F}_\mathbf{p}\Big(\Big\{\frac{\partial}{\partial x^i}\Big|_\mathbf{p}\Big\}\Big)\Big]_{\big\{\frac{\partial}{\partial y^j}\big|_{\mathbf{F}(\mathbf{p})}\big\}}\Big)^i_j$$

is the $\frac{i}{j}$ entry of the matrix of $d\mathbf{F}_\mathbf{p}$ relative to the bases $\Big\{\frac{\partial}{\partial x^i}\Big|_\mathbf{p}\Big\}$ and $\Big\{\frac{\partial}{\partial y^i}\Big|_{\mathbf{F}(\mathbf{p})}\Big\}$. Since we imposed $a^i_j = \frac{\partial F^i}{\partial x^j}\Big|_\mathbf{p}$, then

$$\sum_{j=1}^m a_i^j \frac{\partial}{\partial y^j} = \sum_{j=1}^m \frac{\partial F^j}{\partial x^i}\Big|_\mathbf{p} \frac{\partial}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})}.$$

So far, we have shown

$$d\mathbf{F}_\mathbf{p}\Big(\frac{\partial}{\partial x^i}\Big|_\mathbf{p}\Big) = \sum_{j=1}^m \frac{\partial F^j}{\partial x^i}\Big|_\mathbf{p} \frac{\partial}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})}.$$

At the beginning, we claimed that $d\mathbf{F}_\mathbf{p} : \mathbf{T}_p(\mathbb{R}^n) \to \mathbf{T}_{\mathbf{F}(\mathbf{p})}(\mathbb{R}^m)$; that is, we claimed that $d\mathbf{F}_\mathbf{p}$ would map a tangent vector, which is a derivation at a point, to another tangent vector. To determine exactly what derivation we get by computing $d\mathbf{F}_\mathbf{p}\Big(\frac{\partial}{\partial x^i}\Big)$, we must apply this new derivation to a smooth function $f \in C^\infty(V \subseteq \mathbb{R}^m \to \mathbb{R})$:

$$\begin{aligned} d\mathbf{F}_\mathbf{p}\Big(\frac{\partial}{\partial x^i}\Big)(f) &= \Big(\sum_{j=1}^m \frac{\partial F^j}{\partial x^i}\Big|_\mathbf{p} \frac{\partial}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})}\Big)(f) = \sum_{j=1}^m \Big[\Big(\frac{\partial F^j}{\partial x^i}\Big|_\mathbf{p} \frac{\partial}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})}\Big)(f)\Big] \\ &= \sum_{j=1}^m \frac{\partial F^j}{\partial x^i}\Big|_\mathbf{p} \frac{\partial f}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})} = \sum_{j=1}^m \frac{\partial f}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})} \frac{\partial F^j}{\partial x^i}\Big|_\mathbf{p} = \frac{\partial}{\partial x^i}\Big|_\mathbf{p}(f \circ \mathbf{F}). \end{aligned}$$

Note, this last $\frac{\partial}{\partial x^i}\Big|_\mathbf{p}$ is the usual[4] directional derivative, $\frac{\partial}{\partial x^i}\Big|_\mathbf{p} = \frac{\partial}{\partial \hat{\mathbf{e}}^i} : C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}^n$.

[3] F^i denotes the ith component function of $\mathbf{F}$).

[4] The previous $\frac{\partial}{\partial x^i}\Big|_\mathbf{p}$'s and $\frac{\partial}{\partial y^i}\Big|_{\mathbf{F}(\mathbf{p})}$'s are technically "usual" directional derivatives as well, since are in $\mathbb{R}^n$ and $\mathbb{R}^m$, but it is best to not avoid this interpretation due to Definition 8.31.

The above shows

$$d\mathbf{F}_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}(f \circ \mathbf{F}).$$

Since $d\mathbf{F}_{\mathbf{p}}$ is linear and $\Big\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big\}$ 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 8.33. (Differential of a smooth function $\mathbf{F} : M \to N$).
Let M and N be smooth n- and m- dimensional manifolds WWBOC. We define the *differential* $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{\mathbf{F}(\mathbf{p})}(N)$ *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 8.31, which showed $\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}$ is a derivation at $\mathbf{x}(\mathbf{p}) \in \mathbb{R}^n$. ☐

Theorem 8.34. [Lee, p. 55] (Properties of differentials). Let M, N, and P be smooth manifolds WWBOC, let $\mathbf{F} : M \to N$ and $\mathbf{G} : N \to P$ be smooth functions, and let $\mathbf{p} \in M$. We have the following:

- (Chain rule). $d(\mathbf{G} \circ \mathbf{F})_{\mathbf{p}} = d\mathbf{G}_{\mathbf{F}(\mathbf{p})} \circ d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{(\mathbf{G}\circ\mathbf{F})(\mathbf{p})}(P)$.
- (Differential of the identity). $d(\mathbf{I}_M)_{\mathbf{p}} = \mathbf{I}_M$, where $\mathbf{I}_M : T_{\mathbf{p}}(M) \to T_{\mathbf{p}}(M)$ is the identity on M.
- If $\mathbf{F}$ is a diffeomorphism, then $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{\mathbf{F}(\mathbf{p})}(N)$ is a linear isomorphism, and $(d\mathbf{F}_{\mathbf{p}})^{-1} = d(\mathbf{F}^{-1})_{\mathbf{F}(\mathbf{p})}$.

Proof. See [Lee] for the proof. ☐

Theorem 8.35. [Lee, p. 281] (Differential of a smooth function $M \to \mathbb{R}$).
What happens when we take the differential of a smooth function $f : M \to \mathbb{R}$? Well, by definition of the differential of a smooth function $M \to N$, we have

$$df_{\mathbf{p}}(v_{\mathbf{p}})(g) = v_{\mathbf{p}}(g \circ f).$$

We have $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{f(\mathbf{p})}(\mathbb{R})$. 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 $\widetilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ for which the following diagram commutes:

$$\begin{tikzcd}
T_{\mathbf{p}}(M) \arrow[r, "v_{\mathbf{p}} \mapsto df_{\mathbf{p}}(v_{\mathbf{p}})"] \arrow[dr, "\widetilde{d}f_{\mathbf{p}}"'] & T_{f(\mathbf{p})}(\mathbb{R}) \arrow[d, "w \mapsto \sum_{i=1}^n w(x^i)\hat{\mathfrak{e}}_i"] \\
& \mathbb{R}
\end{tikzcd}$$

We think of $\widetilde{d}$ as the differential that is induced by the identification $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$.

From the diagram, we see that the map $\widetilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ acts on a tangent vector $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by

$$\widetilde{d}f_{\mathbf{p}}(v_{\mathbf{p}}) = \sum_{i=1}^{1} df_{\mathbf{p}}(v_{\mathbf{p}})(x^i)\hat{\mathfrak{e}}_i = df_{\mathbf{p}}(v_{\mathbf{p}})(x^1)\hat{\mathfrak{e}}_1 = v_{\mathbf{p}}(x^1 \circ f)\hat{\mathfrak{e}}_1.$$

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 $x^1 : \mathbb{R}^1 \to \mathbb{R}^1$ becomes the identity on $\mathbb{R}$ and $\hat{\mathfrak{e}}_1$ becomes the scalar $1 \in \mathbb{R}$. These final identifications[5] give

$$\widetilde{d}f_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f).$$

In practice, we write $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ to mean $\widetilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$. (This includes "$dx^i$"; whenever x^i is a coordinate function, we write dx^i to mean $\widetilde{d}x^i$, where $\widetilde{d}$ is the differential obtained by identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$). So, the above characterizing condition of the differential of $f : M \to \mathbb{R}$ is restated as

$$\boxed{df_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)}$$

Theorem 8.36. [Lee, p. 281] (Differential of a smooth function $M \to \mathbb{R}$ in coordinates).

Let M be a smooth n-manifold WWBOC or corners, and let $(U, \mathbf{x})$ be a smooth chart on M, with x^i being the ith coordinate function of $\mathbf{x}$. Then the differential of a smooth function $f : M \to \mathbb{R}$ 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 ith coordinate function x^i of $\mathbf{x}$ after identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ (see the previous theorem).

Proof. Notice that since $(\mathbf{v} \mapsto df_{\mathbf{p}}(\mathbf{v})) = df_{\mathbf{p}}$ is an element of $T_{\mathbf{p}}(M)^*$. (We will explore this fact more later, in Section 8.5).

Because $df_{\mathbf{p}} \in T_{\mathbf{p}}(M)^*$, 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 $C_i : U \to \mathbb{R}$.

To determine the C_i, recall that Theorem 4.28 stated that if V is a finite-dimensional vector space, $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is a basis for V, and $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ 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

$$C_i(\mathbf{p}) = df_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big) = \Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}.$$

Thus $df_{\mathbf{p}}$ is

$$df_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \lambda^i_{\mathbf{p}}.$$

[5] To be very formal, we could write the above as $\widetilde{\widetilde{d}}f_{\mathbf{p}}(v_{\mathbf{p}})$ to indicate that $\widetilde{\widetilde{d}}$ is the differential obtained from $\widetilde{d}$ by identifying $\mathbb{R}^1 \cong \mathbb{R}$, but this identification is so trivial that we do not do this.

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 ith coordinate function of $\mathbf{x}$, and where d of a smooth function $M \to \mathbb{R}$ obtained by identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ (so d is same differential that we've been using in this proof; it is the differential $\widetilde{d}$ of the previous theorem). To prove the claim, we notice that since $\frac{\partial x^i}{\partial x^j} = \delta^i{}_j$ we have

$$dx^i|_{\mathbf{p}} = \sum_{j=1}^{n} \delta^i{}_j \lambda|^j_{\mathbf{p}} = \sum_{j=1}^{n} \delta^i{}_j \lambda|^j_{\mathbf{p}} = \lambda|^i_{\mathbf{p}}.$$

□

Remark 8.37. (Differential of a smooth function $\mathbb{R}^n \to \mathbb{R}$ is the directional derivative).

When $M = \mathbb{R}^n$, we can use the first line in the boxed equation of the previous theorem 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 $df_{\mathbf{p}} : T_{\mathbf{p}}(\mathbb{R}^n) \to \mathbb{R}$. Notice that we have $T_{f(\mathbf{p})}(\mathbb{R}^n) \cong \mathbb{R}^n$ because $T_{\mathbf{p}}(\mathbb{R}^n)$ and $\mathbb{R}^n_{\mathbf{p}}$ are both n-dimensional, so we get from $df_{\mathbf{p}} : T_{\mathbf{p}}(\mathbb{R}^n) \to \mathbb{R}$ another induced function $\widetilde{d}f_{\mathbf{p}} : \mathbb{R}^n_{\mathbf{p}} \to \mathbb{R}$, where $\widetilde{d}$ 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, $\widetilde{d}f_{\mathbf{p}}$ is a linear combination of $\widetilde{d}x^i|_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^n)$ with the same weights:

$$\widetilde{d}f_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \widetilde{d}x^i$$

In the proof of the previous theorem, we saw that $dx^i|_{\mathbf{p}}$ is the ith coordinate function on $T_{\mathbf{p}}(M)$. Since $T_{\mathbf{p}}(\mathbb{R}^n) \cong \mathbb{R}^n_{\mathbf{p}}$, and since dx^i, being an induced dual basis vector, is the ith coordinate function on $T_{\mathbf{p}}(M)$, it follows that $\widetilde{d}x^i$ must be the ith coordinate function on $\mathbb{R}^n_{\mathbf{p}}$. This implies that $\widetilde{d}f_{\mathbf{p}}$ acts on $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$ by

$$\widetilde{d}f_{\mathbf{p}}(\mathbf{v}_{\mathbf{p}}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} \widetilde{d}x^i(\mathbf{v}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} [\mathbf{v}_{\mathbf{p}}]^i_{\hat{\mathfrak{E}}} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} = \frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}},$$

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 8.32) 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 Definition 6.13 that the Jacobian is used to express the directional derivative of a smooth function $\mathbb{R}^n \to \mathbb{R}^m$. In this case we have $m = 1$).

8.5 Tangent vectors and tangent covectors with coordinates

Definition 8.38. [Lee, p. 275] (Cotangent space to a manifold).

Let M be an smooth n-manifold WWBOC, 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, $T^*_{\mathbf{p}}(M) := T_{\mathbf{p}}(M)^*$.

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 8.39. (Induced bases in a chart).

Let M be a smooth n-manifold WWBOC and let $(U, \mathbf{x})$ be a smooth chart on M about $\mathbf{p} \in M$. Consider the set $C^\infty(U \subseteq M \to \mathbb{R})$ of smooth real-valued functions defined on $U \subseteq M$ as a vector space over $\mathbb{R}$. It is a helpful *mnemonic* to pretend that $\{x^i|_{\mathbf{p}}\}_{i=1}^n$ is a basis for $C^\infty(U \subseteq M \to \mathbb{R})$, where x^i is the ith coordinate function of $\mathbf{x}$, and where we've denoted $x^i|_{\mathbf{p}} := x^i(\mathbf{p})$. (*When interpreted even with a little common sense, the mnemonic is clearly nonsensical*, because $C^\infty(U \subseteq M \to \mathbb{R})$ is an infinite-dimensional vector space, so $\{x^i|_{\mathbf{p}}\}_{i=1}^n$, which is a finite set, cannot be a basis for it). This is because, if we accept this *mnemonic*, then

1. The dual basis for $T_{\mathbf{p}}(M) = C^\infty(U \subseteq M \to \mathbb{R})^*$ induced by $\{x^i\}_{i=1}^n$ is $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}$.
2. The dual basis for $T^*_{\mathbf{p}}(M) = T_{\mathbf{p}}(M)^* = C^\infty(U \subseteq M \to \mathbb{R})^{**}$ induced by $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}$ is $dx^i|_{\mathbf{p}}$, where x^i is the ith 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 8.35 to see happens to d when we identify $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$).

Proof.

1. We have $\left(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right)(x^j) = \delta^i{}_j$.
2. This was 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 8.35.

□

If we continue to accept the mnemnoic of the previous theorem, we obtain the following theorem, which describes how to compute the coordinates of tangent vectors and cotangent vectors as a simple consequence of linear algebra.

Theorem 8.40. (Coordinates of tangent vectors and cotangent vectors).

Theorem 4.28 stated that if V is a finite-dimensional vector space over K, $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is a basis for V, and $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ is the basis for V^* induced by E, then

$$([\mathbf{v}]_E)^i = \phi^{\mathbf{e}_i}(\mathbf{v}) = \phi_{\mathbf{v}}(\Phi_{\mathbf{e}_i})$$
$$([\phi]_{E^*})_i = \phi(\mathbf{e}_i),$$

where $\Phi_{\mathbf{v}} \in V^{**}$ is the linear function $V^* \to K$ 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 WWBOC and let $(U, \mathbf{x})$ be a smooth chart on M about $\mathbf{p} \in M$. Then the ith 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 ith coordinate of a tangent covector $\phi_{\mathbf{p}} \in T^*_{\mathbf{p}}(M)$ are

$$\boxed{\begin{aligned} &([v_{\mathbf{p}}]_{\left\{\frac{\partial}{\partial x^j}\big|_{\mathbf{p}}\right\}_{j=1}^n})^i = v_{\mathbf{p}}(x^i) = \phi_{x^i}(v_{\mathbf{p}}) \\ &([\phi_{\mathbf{p}}]_{\left\{dx^j|_{\mathbf{p}}\right\}_{j=1}^n})_i = \phi_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big) \end{aligned}}$$

In the second equation of the first line, ϕ_{x^i} is the element of $C^\infty(U \subseteq M \to \mathbb{R})^{**} = T^*_{\mathbf{p}}(M)$ that is identified with the ith coordinate function $x^i \in C^\infty(U \subseteq M \to \mathbb{R})$ of $\mathbf{x}$. Recall Theorem 3.30 to see that $\phi_f : C^\infty(U \subseteq M \to \mathbb{R})^* = T_{\mathbf{p}}(M) \to \mathbb{R}$, where $f \in C^\infty(U \subseteq M \to \mathbb{R})$, 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 8.41. (Covariance and contravariance with tangent vectors and tangent covectors).

Recall the convention of "covariance and contravariance" from Definition 3.20. The first half of this convention stipulates that the coordinates of a covector are to be indexed by lower subscripts, while covectors themselves are to be are indexed by upper subscripts. The second half states that the coordinates of a vector (which is "contravariant") are to be indexed by lower subscripts, while vectors themselves are to be indexed by lower subscripts.

We now call attention to the fact that our uses of upper and lower indices for a basis coordinate function x^i, a basis tangent vector $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$, and a basis cotangent vector $dx^i|_{\mathbf{p}}$ can be retroactively interpreted as obeying these conventions.

We can argue that an upper index is used in x^i because x^i is a coordinate of a smooth chart $\mathbf{x} : U \to \mathbb{R}^k$, which is "close enough" to a vector, since its output is $\mathbb{R}^k$. (This argument ignores the fact that the primary vectors of interest when studying manifolds are tangent vectors, not elements of $\mathbb{R}^k$).

As for the notation $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$, we would somehow like to be able to interpret i as being a lower index, since $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ is a tangent vector, which is contravariant. Our justification will be that even though i is typographically an upper index, i appears in the denominator, so it is "low".

With $dx^i|_{\mathbf{p}}$, all is well and good- an upper index is used on a tangent covector- so no retroactive argument is necessary.

Remark 8.42. (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 8.31). 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 of that definition (on the right hand side, the x^i in the "denominator" is used in the notation $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} := \frac{\partial}{\partial \hat{\mathfrak{e}}^i}\Big|_{\mathbf{p}}$ for directional derivatives which act on smooth functions defined on $\mathbb{R}^n$).

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 ith 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 8.43. (Change of coordinates for tangent vectors in terms of basis vectors of $T_{\mathbf{p}}(M)$).

Theorem 2.110 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 ith coordinate functions of $\mathbf{x}$ and $\mathbf{y}$, respectively. Applying the above theorem to the vector space $T_{\mathbf{p}}(M)$ and its bases $E = \left\{\frac{\partial}{\partial x^j}\Big|_{\mathbf{p}}\right\}_{j=1}^n$ and $F = \left\{\frac{\partial}{\partial y^j}\Big|_{\mathbf{p}}\right\}_{j=1}^n$, we have

$$[\mathbf{F}]_E = \left(\left[\frac{\partial}{\partial y^1}\Big|_{\mathbf{p}}\right]_F \quad \cdots \quad \left[\frac{\partial}{\partial y^n}\Big|_{\mathbf{p}}\right]_E\right) = \begin{pmatrix} \frac{\partial x^1}{\partial y^1}\Big|_{\mathbf{p}} & \cdots & \frac{\partial x^1}{\partial y^n}\Big|_{\mathbf{p}} \\ \vdots & & \vdots \\ \frac{\partial x^n}{\partial y^1}\Big|_{\mathbf{p}} & \cdots & \frac{\partial x^n}{\partial y^n}\Big|_{\mathbf{p}} \end{pmatrix} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}},$$

where $\frac{\partial x^i}{\partial y^j}\Big|_{\mathbf{p}} = \left(\frac{\partial}{\partial y^j}\Big|_{\mathbf{p}}\right)(x^i)$, and where x^i is the ith coordinate function of $\mathbf{x}$. The matrix $\frac{\partial \mathbf{x}}{\partial \mathbf{y}}$ is the Jacobian matrix described in Definition 6.13.

Applying the fact $\mathbf{f}_i = \sum_{j=1}^n ([\mathbf{F}]_E)_{ji} \mathbf{e}_j$ from above, we have

$$\boxed{\frac{\partial}{\partial y^i}\Big|_{\mathbf{p}} = \sum_{j=1}^n \frac{\partial x^j}{\partial y^i}\Big|_{\mathbf{p}} \frac{\partial}{\partial x^j}\Big|_{\mathbf{p}}}$$

This change of basis equation strongly resembles the chain rule, and indeed simplifies to the chain rule when $M = \mathbb{R}^n$. When $M \neq \mathbb{R}^n$, 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 8.44. [Lee, p. 65] (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 disjoint[6] union.

Theorem 8.45. [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. Here's a rough idea of the proof. Consider the special case of $M = \mathbb{R}^n$:

$$T(\mathbb{R}^n) = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} T_{\mathbf{p}}(\mathbb{R}^n) \cong \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} \mathbb{R}^n_{\mathbf{p}} = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} (\mathbb{R}^n \times \{\mathbf{p}\}) = \mathbb{R}^n \times \mathbb{R}^n.$$

□

Definition 8.46. [Lee, p. 276] (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. Just as was the case with the tangent bundle, the cotangent bundle has a natural topology and smooth structure that make it into a $2n$-dimensional smooth manifold.

Definition 8.47. [Lee, p. 316, 317] (Tensor bundle).

Let M be a smooth n-manifold. The (p,q) *tensor bundle* of M is the set $T^p_q(T(M)) := \bigsqcup_{\mathbf{p} \in M} T^p_q(T_{\mathbf{p}}(M))$. The tensor bundle has a natural topology and smooth structure that make it into a smooth manifold.

Remark 8.48. [Lee, p. 316, 317] (Tangent bundle and cotangent bundle special cases of the tensor bundle).

We have $T^p_0(T(M)) = \bigsqcup_{\mathbf{p} \in M} (T_{\mathbf{p}}(M))^{\otimes p}$ and $T^0_q(T(M)) = \bigsqcup_{\mathbf{p} \in M} (T^*_{\mathbf{p}}(M))^{\otimes p}$. So in particular, $T^1_0(T(M)) = T(M)$, $T^0_1(T(M)) = T^*(M)$.

Definition 8.49. [Lee, p. 174, 276, 316, 317] (Vector fields, covector fields, and (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 $A \subseteq M \to T(M)$, a *covector field on* A is a continuous function $A \subseteq M \to T^*(M)$, and a (p,q) *tensor field on* A is a continuous function $A \subseteq M \to T^p_q(T(M))$. *Smooth vector fields*, *smooth covector fields*, and *smooth tensor fields* are smooth such functions.

Frames and coframes

Definition 8.50. [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 $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ 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 $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ 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.

[6] Note that a disjoint union is *not* a "pairwise disjoint union". We say $A = \bigsqcup_\alpha A_\alpha$ iff $A = \bigsqcup_\alpha A_\alpha$ and $\cap_\alpha A_\alpha = \emptyset$.

A *local coframe on* $(U, \mathbf{x})$ is a set of covector fields $\{\boldsymbol{\mathcal{E}}^1, ..., \boldsymbol{\mathcal{E}}^n\}$ on U, where, at each $\mathbf{p} \in U \subseteq M$, the $\boldsymbol{\mathcal{E}}^i$'s are linearly independent and span $T^*_{\mathbf{p}}(M)$. A local coframe $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ 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 8.51. [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 smooth[7] 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 *ith local coordinate vector field (on* $(U, \mathbf{x})$*)*.

Recall that $\{dx^i_{\mathbf{p}}\}_{i=1}^n$ is the dual basis of $T^*_{\mathbf{p}}(M)$ induced by the basis $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}_{i=1}^n$ for $T_{\mathbf{p}}(M)$. The smooth[8] 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 *ith local coordinate covector field (on* $(U, \mathbf{x})$*)*.

Definition 8.52. [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 smooth[9] local frame $\{\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}\}$, where $\frac{\partial}{\partial x^i}$ denotes the ith local coordinate vector field, is called a *local coordinate frame (on* $(U, \mathbf{x})$*)*.

The smooth[10] coframe $\{dx_1, ..., dx_n\}$, where dx_i denotes the ith local coordinate covector field, is called the *local coordinate coframe (on* $(U, \mathbf{x})$*)*.

Definition 8.53. [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 $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ on $(U, \mathbf{x})$ and local coframe $\{\boldsymbol{\mathcal{E}}^1, ..., \boldsymbol{\mathcal{E}}^n\}$ on $(U, \mathbf{x})$ are said to be *dual* to each other iff, at each $\mathbf{p} \in U$, the basis $\{\boldsymbol{\mathcal{E}}^1|_{\mathbf{p}}, ..., \boldsymbol{\mathcal{E}}^n|_{\mathbf{p}}\}$ is the dual basis induced for $T^*_{\mathbf{p}}(M)$ induced by the basis $\{\mathbf{E}_1|_{\mathbf{p}}, ..., \mathbf{E}_n|_{\mathbf{p}}\}$ for $T_{\mathbf{p}}(M)$.

As an example, $\frac{\partial}{\partial x^i}$ and dx^i are dual to each other.

[7]The ith coordinate vector field is smooth because its component functions are constants.

[8]The ith coordiante covector field is smooth because its component functions are constants.

[9]The local coordinate frame is smooth because the ith local coordinate vector field is smooth.

[10]It 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.

8.7 Oriented manifolds and their oriented boundaries

In Section 5.3, we defined what it means to orient a finite-dimensional vector space. The key result we motivated was Definition 5.54, which said: if a basis E of a finite-dimensional vector space V is fixed, than another ordered basis F of V is said to be *positively oriented (relative to E)* iff $\det([\mathbf{F}]_E) > 0$ and *negatively oriented (relative to E)* otherwise. Given this definition, we showed in Theorem 5.56 that a choice of element of $\Lambda^{\dim(V)}(V)$, determines an orientation on V. We now extend the idea of orientation to manifolds.

Definition 8.54. [Lee, p. 380] (Oriented manifolds).

Let M be an n-manifold WWBOC. We define the notion of orientation on a manifold in the following steps:

- Since a choice of element of the top exterior power of a finite-dimensional vector space determines an orientation for that vector space, we define a *pointwise orientation form* for M to be any nonvanishing differential n-form on M that is an element of $\widetilde{\Omega}^n(M)$. Such a differential form is an element of $\widetilde{\Lambda}^k(T_{\mathbf{p}}(M)) \cong \Lambda^k(T_{\mathbf{p}}(M))$ at each $\mathbf{p} \in M$, and therefore determines an orientation on each tangent space.
- If ω is a pointwise orientation form on M and $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ is a local frame for the tangent bundle $T(M)$, then we say that $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ is *positively oriented* iff $\{\mathbf{E}_1|_{\mathbf{p}}, ..., \mathbf{E}_n|_{\mathbf{p}}\}$ is positively oriented relative to ω at each $\mathbf{p} \in M$, that $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ is *negatively oriented* iff $\{\mathbf{E}_1|_{\mathbf{p}}, ..., \mathbf{E}_n|_{\mathbf{p}}\}$ is negatively oriented relative to ω at each $\mathbf{p} \in M$.
- A pointwise orientation form ω for M is said to be *continuous* iff every point of M is in the domain of an oriented local frame, where the orientation of the oriented local frame is given by ω.
- An *orientation form* for M is a continuous pointwise orientation form for M. We say M is *orientable*, or that M is an *oriented manifold*, iff there exists an orientation form for M.

Definition 8.55. [Lee, p. 381, 382] (Orientation of a smooth chart on an oriented manifold).

A smooth chart $(U, \mathbf{x})$ on an oriented smooth n-manifold WWBOC is said to be *positively oriented* iff the coordiante frame $\{\frac{\partial}{\partial x^i}\}_{i=1}^n$ is positively oriented, and is said to be *negatively oriented* otherwise.

Boundary orientation

We now present the natural way in which the boundary of a manifold inherits orientation from the rest of the manifold.

Lemma 8.56. If M is an n-manifold, then the boundary ∂M is an $(n-1)$-manifold.

To describe the inherited orientation on a boundary, we need to define what it means for tangent vectors to be "inward-pointing" or "outward-pointing".

Definition 8.57. [Lee, p. 118] (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 a 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.

Lemma 8.58. [Lee, p. 200, problem 8-4] (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.

Now we can describe the induced orientation on the boundary.

Theorem 8.59. [Lee, p. 385, 386] (Induced orientation form on the boundary).

Let ω_M be an orientation form on M. The previous lemma shows that there exists a vector field $\mathbf{N}$ nowhere tangent to ∂M. Thus, there is an induced orientation form $\omega_{\partial M}$ on the boundary (due to interior multiplication)[11] defined by $\omega_{\partial M}(v_1|_{\mathbf{p}}, ..., v_{n-1}|_{\mathbf{p}}) = \omega_M(\mathbf{N}_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., v_{n-1}|_{\mathbf{p}})$. The orientation on ∂M induced by $\omega_{\partial M}$ does not depend on the vector field $\mathbf{N}$ that is nowhere tangent to ∂M.

Proof. We need to show (1) that $\omega_{\partial M}$ is indeed an orientation form on ∂M and (2) that the orientation induced by $\omega_{\partial M}$ is independent of the choice of the nowhere tangent vector field $\mathbf{N}$.

1. We need to show that $\omega_{\partial M}$ never vanishes. Note that if $\{e_1|_{\mathbf{p}}, ..., e_{n-1}|_{\mathbf{p}}\}$ is a basis for $T_{\mathbf{p}}(\partial M)$, then, since $\mathbf{N}$ is nowhere tangent to $T_{\mathbf{p}}(\partial M)$, the set $\{\mathbf{N}_{\mathbf{p}}, e_1|_{\mathbf{p}}, ..., e_{n-1}|_{\mathbf{p}}\}$ is a basis for $T_{\mathbf{p}}(M)$. Because of this, and as ω_M is nonvanishing on M, then $\omega_{\partial M}$ is also nonvanishing. See [Lee, p. 385] for the more precise details.
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 $E = \{\mathbf{N}_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., v_{n-1}|_{\mathbf{p}}\}$ and $F = \{\mathbf{N}'|_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., v_{n-1}|_{\mathbf{p}}\}$ 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 nth component of $\mathbf{N}_{\mathbf{p}}$ relative to E and the nth component of $\mathbf{N}'|_{\mathbf{p}}$ relative to F are both negative; denote these nth 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 only nonzero entry is the nth entry, which is nth entry is $\frac{(\mathbf{N}'|_{\mathbf{p}})_n}{(\mathbf{N}_{\mathbf{p}})_n}$. For $i > 1$, the ith column of the change of basis matrix is $\hat{\mathfrak{e}}_i$. The change of basis matrix is therefore upper triangular, so its determinant is the product of the diagonal entries, i.e., the determinant is $\frac{(\mathbf{N}'|_{\mathbf{p}})_n}{(\mathbf{N}_{\mathbf{p}})_n} > 0$.

□

Orientation of the boundary of a k-parallelapiped

As k-paralellapiped can be be given the structure of an oriented smooth submanifold of $\mathbb{R}^n$ with corners.

Definition 8.60. (Notation for n-parallelapipeds).

Let M be a smooth manifold WWBOC, and let $(U, \mathbf{x})$ be a smooth chart on M. Given $v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$, we define $P_{\mathbf{p}}(v_1, ..., v_n)$ to be the n-parallelapiped anchored at $\mathbf{x}(\mathbf{p}) \in \mathbb{R}^n_{\mathbf{p}}$ that is spanned by $\mathbf{x}(v_1|_{\mathbf{p}}), ..., \mathbf{x}(v_n|_{\mathbf{p}}) \in \mathbb{R}^n_{\mathbf{p}}$. Additionally, if $v \in T_{\mathbf{p}}(M)$ is identified with $\mathbf{v} \in \mathbb{R}^n_{\mathbf{p}}$ under the isomorphism $\left.\frac{\partial}{\partial x^i}\right|_{\mathbf{p}} \mapsto \hat{\mathfrak{e}}_i$, we will use the slight abuse of notation $P_{\mathbf{p}+v}(v_1, ..., v_n)$ to denote the n-parallelapiped anchored at $\mathbf{x}(\mathbf{p}) + \mathbf{v} \in \mathbb{R}^n_{\mathbf{p}}$ that is spanned by $\mathbf{x}(v_1|_{\mathbf{p}}), ..., \mathbf{x}(v_n|_{\mathbf{p}}) \in \mathbb{R}^n_{\mathbf{p}}$.

Theorem 8.61. [HH, p. 542 - 544] (Oriented boundary of an n-parallelapiped).

Let M be an n-dimensional manifold, let $\mathbf{p} \in M$, and consider vectors $v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. Let $P_{\mathbf{p}}(v_1, ..., v_n)$ denote the n-parallelapiped anchored at $\mathbf{p}$ that is spanned by $v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}}$. The orientation of the boundary $\partial P_{\mathbf{p}}(v_1, ..., v_n)$ is given by

$$\boxed{\partial P_{\mathbf{p}}(v_1, ..., v_n) = \sum_{i=1}^{k} (-1)^{i-1} \Big(P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n) - P_{\mathbf{p}}(v_1, ..., \cancel{v_i}, ..., v_n) \Big)}$$

In the sum, $+$ and $-$ signs are used to indicate whether a n-parallelapiped is positively or negatively oriented relative to the orientation of $P_{\mathbf{p}}(v_1, ..., v_n)$.

Proof. Since $P_{\mathbf{p}}(v_1, ..., v_n)$ is an n-parallelapiped, then $\partial P_{\mathbf{p}}(v_1, ..., v_n)$ is a $2n$-parallelapiped. Each face of $\partial P_{\mathbf{p}}(v_1, ..., v_n)$ is of the form $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$ or $P_{\mathbf{p}}(v_1, ..., \cancel{v_i}, ..., v_n)$. We show that the faces of the form $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$ have the same orientation as $P_{\mathbf{p}}(v_1, ..., v_n)$, and that the faces of the form $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$ have the opposite orientation as does $P_{\mathbf{p}}(v_1, ..., v_n)$.

Let the orientation on $P_{\mathbf{p}}(v_1, ..., v_n)$ be given by an orientation form ω_P. By Theorem 8.59, the induced orientation form $\omega_{\partial P}$ on the boundary is defined by $\omega_{\partial P} = \omega_P(v_i|_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., \cancel{v_i|_{\mathbf{p}}}, ..., v_n|_{\mathbf{p}})$. Since $v_i|_{\mathbf{p}}$ is outward-pointing on the face $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$, the orientation of $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$ is the same as the orientation of $P_{\mathbf{p}}(v_1, ..., v_n)$. Conversely, $v_i|_{\mathbf{p}}$ is inward-pointing on the face $P_{\mathbf{p}}(v_1, ..., \cancel{v_i}, ..., v_n)$, so the orientation of $P_{\mathbf{p}}(v_1, ..., \cancel{v_i}, ..., v_n)$ is same as the orientation given by $\omega_P(-v_i|_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., \cancel{v_i|_{\mathbf{p}}}, ..., v_n|_{\mathbf{p}}) = -\omega_{\partial P}$. □

[11]We could have put $\mathbf{N}_{\mathbf{p}}$ in any of ω_M'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]. See p. 358 and Corollary 14.21 on p. 362.

9

Differential forms on manifolds

At long last, we have reached the chapter where we define *differential forms* and how to integrate them over manifolds. In this chapter, we also define an operation on differential forms called the *exterior derivative*, which generalizes the differential of the previous chapter. We will see that the div, grad and curl of vector calculus can be expressed in terms of the exterior derivative and the *Hodge-dual* operator. The *generalized Stokes' theorem*, an elegant generalization of the fundamental theorem of calculus, is the key result of this chapter, and equates the integral of an exterior derivative over a manifold to the integral of the "exterior antiderivative" over the manifold's boundary.

9.1 Differential forms

Definition 9.1. [Lee, p. 360] (Differential k-form).

Let M be a smooth n-manifold WWBOC. We define $\Lambda^k(T^*(M)) := \bigsqcup_{\mathbf{p} \in M} \Lambda^k(T^*_{\mathbf{p}}(M))$, so that $\Lambda^k(T^*(M))$ is the the subset of $T^0_k(T(M))$ of antisymmetric tensors. The vector space $\Lambda^k(T^*(M))$ can be shown to be $\binom{n}{k}$ dimensional, just as is the case for $\Lambda^k(V)$ when V is an n-dimensional vector space.

A *differential k-form on M* is a continuous function $M \to \Lambda^k(T^*(M))$. So, you might say that a differential k-form is a "antisymmetric $(0,k)$ 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 ith component function of $\mathbf{x}$, it follows by definition that a differential k-form ω on U can be expressed as

$$\omega = \sum_{1 \leq i_1 < ... < i_k \leq n} f_{i_1 ... i_k} dx^{i_1} \wedge ... \wedge dx^{i_k},$$

where each $f_{i_1 ... i_k} : U \to \mathbb{R}$.

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. A differential 0-form on M is a continuous function $M \to \mathbb{R}$.

Differential forms treated as functions

We will occasionally need to think of differential k-forms as objects that, when evaluated at $\mathbf{p} \in M$, can act on k vectors in $T_{\mathbf{p}}(M)$.

Derivation 9.5. (Differential forms treated as functions).

Let M be a smooth manifold WWBOC. Since we are able to think of elements of $\Lambda^k(V^*)$ as functions by identifying them with elements of $\widetilde{\Lambda}^k(V^*)$, we can think of the evaluation at a point $\mathbf{p} \in M$ of a differential k-forms on M, which is an element of $\Lambda^k(T^*(M)) = \bigsqcup_{\mathbf{p} \in M} \Lambda^k(T^*_{\mathbf{p}}(M))$, as a function by identifying it with an element of $\widetilde{\Lambda}^k(T^*_{\mathbf{p}}(M))$, where we define $\widetilde{\Lambda}^k(T^*_{\mathbf{p}}(M)) := \bigsqcup_{\mathbf{p}\in M} \widetilde{\Lambda}^k(T^*_{\mathbf{p}}(M))$. To formally establish this, we define $\widetilde{\Omega}^k(M) := \{\text{continuous functions: } M \to \widetilde{\Lambda}^k(T^*(M))\}$ to be the set of differential k-forms that, when evaluated at a point, are functions- namely, multilinear alternating functions accepting k vectors from $T_{\mathbf{p}}(M)$.

9.2 Integration of differential forms on manifolds

In this section, we vaguely follow Chapter 16 of [Lee] (but take some cues from [HH], and fewer from [GP74]), and show how to integrate differential forms over manifolds. We will see that differential forms are the "natural" objects to integrate over manifolds because the pullback of a differential form interplays perfectly with the change of variables theorem from multivariable calculus.

On a more technical note, we will consider only compactly supported differential forms so that the integrals we consider are analogous to "proper" (as opposed to "improper") integrals of functions $f : \mathbb{R}^n \to \mathbb{R}$.

To begin, we define the pullback of a differential k-form by a diffeomorphism and show that it plays nicely with the change of variables theorem.

Pullback by a diffeomorphism and change of variables

Definition 9.6. [GP74, p. 163-165] (Pullback of a differential k-form by a diffeomorphism).

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 $\mathbf{F} : U \subseteq M \to V \subseteq N$ 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 the evaluation of ω at a point $\mathbf{q} \in V \subseteq N$ yields an element of $\Lambda^k T^*_{\mathbf{q}}(N)$, we can pull back differential k-forms on N to differential k-forms on M by using the pullback map on exterior powers.

We define the *pullback* $\Omega^k \mathbf{F}^* : \Omega^k(V) \to \Omega^k(U)$ *of the differential k-form ω on V* to be

$$\boxed{\Big((\Omega^k \mathbf{F}^*)(\omega)\Big)_{\mathbf{p}} := \begin{cases} \Big(\Lambda^k(d\mathbf{F}_{\mathbf{p}})^*\Big)(\omega_{\mathbf{F}(\mathbf{p})}) & \omega \text{ is a differential } k\text{-form on } V,\ k \geq 1 \\ (f \circ \mathbf{F})(\mathbf{p}) & \omega = f \text{ is a differential 0-form on } V \text{ (i.e. } \omega \text{ is a function } V \to \mathbb{R}) \end{cases}}$$

Here's a brief explanation of the notation. The map $(d\mathbf{F}_{\mathbf{p}})^* : T^*_{\mathbf{F}(\mathbf{p})}(V) \to T^*_{\mathbf{p}}(U)$ is the dual of the differential $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(U) \to T_{\mathbf{F}(\mathbf{p})}(V)$, which is a linear map, and $\Lambda^k(d\mathbf{F}_{\mathbf{p}})^*$ is the pullback from the kth exterior power $\Lambda^k(T^*_{\mathbf{F}(\mathbf{p})}(V))$ to the kth exterior power $\Lambda^k(T^*_{\mathbf{p}}(U))$. (For the exact meaning of $\Lambda^k(d\mathbf{F}_{\mathbf{p}})^*$, recall Definition **??**).

Theorem 9.7. (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 $\mathbf{F} : U \subseteq M \to N$.

1. If ω is a differential 1-form on N, then $\Omega^k \mathbf{F}^*(\omega)_{\mathbf{p}} = (d\mathbf{F}_{\mathbf{p}})^*(\omega)$, where $(d\mathbf{F}_{\mathbf{p}})^*$ is the dual of the linear map $d\mathbf{F}_{\mathbf{p}}$.
2. If ω and η are differential k-forms, $k \leq n$, on N, then

$$\Omega^k \mathbf{F}^*(\omega \wedge \eta) = \Omega^k \mathbf{F}^*(\omega) \wedge \Omega^k \mathbf{F}^*(\eta).$$

Proof. All items follow straightforwardly from the definition of $\Omega^k \mathbf{F}^*$. □

Theorem 9.8. (The differential commutes with the pullback of a diffeomorphism).

Let M and N be smooth manifolds WWBOC, let $U \subseteq M$ be open, and consider a smooth function $\mathbf{F} : U \subseteq M \to N$. Additionally, let $\mathbf{p} \in U \subseteq M$, and let $d_{\mathbf{p}}$ denote the map $f \overset{d_{\mathbf{p}}}{\mapsto} df_{\mathbf{p}}$. Then for any smooth function $f : \mathbb{R}^n \to U \subseteq M$, 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 $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ or the differential in which this identification is not performed, and $T_{f(\mathbf{p})}(\mathbb{R})$ is "left alone" (recall Theorem 8.35 and Definition 8.33, respectively).

Proof. For this proof, we will assume that d is the differential in which $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ 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 8.35). (After we have proven the theorem for this interpretation of d, the theorem holds for when d is obtained by identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ precisely because we have an isomorphism $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$).

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 8.33), $(d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}}))(f) = v_{\mathbf{p}}(f \circ \mathbf{F})$. Since f is a 0-form, $f \circ \mathbf{F} = \Omega^k\mathbf{F}^*(f)$ by definition of $\Omega^k\mathbf{F}^*$. So $v_{\mathbf{p}}(f \circ \mathbf{F}) = v_{\mathbf{p}}(\Omega^k\mathbf{F}^*(f))$. Recall that because d is the differential in which $T_{f(\mathbf{p})}(\mathbb{R})$ is "left alone" (see the remarks at the beginning of this proof), we have $v_{\mathbf{p}}(g) = dg_{\mathbf{p}}(v_{\mathbf{p}})$. We set $g = \Omega^k\mathbf{F}^*(f)$ to obtain $v_{\mathbf{p}}(\Omega^k\mathbf{F}^*(f)) = d(\Omega^k\mathbf{F}^*(f))_{\mathbf{p}}(v_{\mathbf{p}})$. ◻

Lemma 9.9. (Technical lemma: $\Omega^k\mathbf{F}^*(dy^i) = \widetilde{F}^i_{(V,\mathbf{y})}$).

Let M and N be smooth n-manifolds WWBOC, let $(U, \mathbf{x})$ be a smooth chart on M, and let $(V, \mathbf{y})$ be a smooth chart on N. Let $\mathbf{F} : U \subseteq M \to V \subseteq N$ be a diffeomorphism. Recall from Definition 8.9 that the coordinate representation of $\mathbf{F}$ relative to the chart $(V, \mathbf{y})$ is $\widetilde{\mathbf{F}}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{F}$, so $y^i \circ \mathbf{F}$ is $y^i \circ \mathbf{F} = \widetilde{F}^i$, the ith coordinate function of $\widetilde{\mathbf{F}}_{(V,\mathbf{y})} = (\widetilde{F}^1_{(V,\mathbf{y})}, ..., \widetilde{F}^n_{(V,\mathbf{y})})^\top$. But $y^i \circ \mathbf{F} = \Omega^k\mathbf{F}^*(y^i)$, so

$$\widetilde{F}^i_{(V,\mathbf{y})} = \Omega^k\mathbf{F}^*(y^i).$$

Since the differential commutes with the pullback of a diffeomorphism, we have $\Omega^k\mathbf{F}^*(dy^i) = d(\Omega^k\mathbf{F}^*)(y^i)$. By the previous lemma, $\Omega^k\mathbf{F}^*(y^i) = \widetilde{F}^i_{(V,\mathbf{y})}$. Therefore

$$\Omega^k\mathbf{F}^*(dy^i) = \widetilde{F}^i_{(V,\mathbf{y})}.$$

Theorem 9.10. [Lee, p. 361] (Pullback of a top degree differential form).

Let M and N be smooth n-manifolds WWBOC, let $(U, \mathbf{x})$ be a smooth chart on M, and let $(V, \mathbf{y})$ be a smooth chart on N. Let $\mathbf{F} : U \subseteq M \to V \subseteq N$ be a diffeomorphism, and consider a differential n-form $\omega \in \Omega^n(V)$ on $V \subseteq N$. (We say that ω is *top degree* differential form on N, since its degree (the "degree" of a differential k-form is k) is the same as the dimension of N). If $f\ dy^1 \wedge ... \wedge dy^n$ is a differential n-form on V, the pullback of $f\ dy^1 \wedge ... \wedge dy^n$ is

$$\boxed{\Omega^k\mathbf{F}^*(f\ dy^1 \wedge ... \wedge dy^n) = (f \circ \mathbf{F}) \det\Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j}\Big) dx^1 \wedge ... \wedge dx^n}$$

Here $\widetilde{\mathbf{F}}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{F}$ is the coordinate representation of $\mathbf{F}$ taken relative to the chart $(V, \mathbf{y})$, and $\widetilde{F}^i_{(V,\mathbf{y})}$ is the ith component function of $\widetilde{\mathbf{F}}_{(V,\mathbf{y})}$.

Proof. By Theorem 9.2, we can write ω in the chart $(V, \mathbf{y})$ as

$$\omega = f\ dy^1 \wedge ... \wedge dy^n,$$

where f is a continuous function $N \to \mathbb{R}$, and where y^i is the ith coordinate function of $\mathbf{y}$.
First, we simply rewrite the above after applying the fact $\Omega^k\mathbf{F}^*(f) = f \circ \mathbf{F}$:

$$\Big(\Omega^k\mathbf{F}^*(\omega)\Big)_{\mathbf{p}} = (f \circ \mathbf{F})(\mathbf{p}) \Big(\Omega^k\mathbf{F}^*(dy^1)\Big)_{\mathbf{p}} \wedge ... \wedge \Big(\Omega^k\mathbf{F}^*(dy^n)\Big)_{\mathbf{p}}.$$

Since $\Omega^k\mathbf{F}^*(dy^i) = \widetilde{F}^i_{(V,\mathbf{y})}$ (see the previous lemma), we have

$$\Big(\Omega^k\mathbf{F}^*(f\ dy^1 \wedge ... \wedge dy^n)\Big)_{\mathbf{p}} = (f \circ \mathbf{F})(\mathbf{p})\ d\widetilde{F}^1_{(V,\mathbf{y})}|_{\mathbf{p}} \wedge ... \wedge d\widetilde{F}^n_{(V,\mathbf{y})}|_{\mathbf{p}}.$$

Applying Theorem 5.70, we have

$$d\widetilde{F}^1_{(V,\mathbf{y})}|_{\mathbf{p}} \wedge ... \wedge d\widetilde{F}^n_{(V,\mathbf{y})}|_{\mathbf{p}} = \det\Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j}\Big) dx^1 \wedge ... \wedge dx^n.$$

Plug this expression for $d\widetilde{F}^1_{(V,\mathbf{y})}|_{\mathbf{p}} \wedge ... \wedge d\widetilde{F}^n_{(V,\mathbf{y})}|_{\mathbf{p}}$ into the previous equation to obtain the theorem. ◻

Remark 9.11. (Star notation for the pullback of a top degree differential form).

In Remark 5.22, we introduced the practice of using $\mathbf{F}^*$ to denote a pullback. From now on, we denote the pullback $\Omega^k \mathbf{F}^* : T_{\mathbf{F}(\mathbf{p})}(V \subseteq N) \to T_{\mathbf{p}}(U \subseteq M)$ by $\mathbf{F}^* : T_{\mathbf{F}(\mathbf{p})}(V \subseteq N) \to T_{\mathbf{p}}(U \subseteq M)$.

Integrating differential forms

We now use the theorem of the previous section to start constructing the notion of integration of a differential form over manifolds. We need the following two technical definitions before doing so, however.

Definition 9.12. [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 9.13. [Lee, p. 43] (Support of a differential form on a manifold).

Let M be a (not necessarily smooth) manifold and let ω be a differential form on M. The *support* of ω is defined to be the closure of the set of points where ω is nonzero, $\text{supp}(\omega) := \text{cl}(M - \omega^{-1}(\mathbf{0}))$. Iff $\text{supp}(\omega) \subseteq A$, then we say ω *is supported in* A. We say ω is *compactly supported* iff $\text{supp}(\omega)$ is compact.

Now, we use the formula for the pullback of a diffeomorphism derived at the end of the previous section to restate the change of variables theorem.

Derivation 9.14. [GP74, p. 166] [Lee, p. 403] (Change of variables theorem in light of the pullback, part 1).

The change of variables theorem says that if $(D, \mathbf{x})$ and $(E, \mathbf{y})$ are charts on $\mathbb{R}^n$ or $\mathbb{H}^n$, where D and E are open domains of integration, and if $\mathbf{F} : \text{cl}(D) \to \text{cl}(E)$ is a smooth map that restricts to a diffeomorphism from D to E, then, for every continuous function $f : \text{cl}(E) \to \mathbb{R}$, we have

$$\int_E f = \int_D (f \circ \mathbf{F}) \Big| \det \Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j} \Big) \Big|.$$

Notice that, when $\Big| \det \Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j} \Big) \Big| = \det \Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j} \Big)$, i.e., when $\det(d\mathbf{F}) > 0$, it is *almost* true that the pullback of the integrand on the left hand side is equal to the integrand on the right hand side, since $f dy^1 \wedge ... \wedge dy^n$ pulls back to $\det \Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j} \Big) \Big| \, dx^1 \wedge ... \wedge dx^n$. That is, the pullback of the differential form corresponding to the integrand of the left hand side is equal to the differential form corresponding to the integrand of the right hand side (where the "correspondence" spoken of is obtained by concatenating $dy^1 \wedge ... \wedge dy^n$ or $dx^1 \wedge ... \wedge dx^n$ onto the integrand). This observation motivates the next definition.

Definition 9.15. [Lee, p. 402] (Integral of a top degree 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 differential n-form $f \; dx^1 \wedge ... \wedge dx^n$ over D*, to be

$$\boxed{\int_D f \; dx^1 \wedge ... \wedge dx^n := \int_D f}$$

Remark 9.16. (The meaning of $dx_1 ... dx_n$).

We should note that our recent definition of the integral of a differential form gives meaning to the $dx_1 ... dx_n$ that is used as a placeholder in an integral. If we use the convention of writing the placeholder $dx_1 ... dx_n$ after the integrand, so that

$$\int_D f = \int_D f \; dx_1 ... dx_n,$$

then the definition of the integral of a differential n-form on an open domain of integration becomes

$$\int_D f\ dx_1 \wedge ... \wedge dx_n := \int_D f\ dx^1...dx^n.$$

In some sense, the placeholder $dx_1...dx_n$ is "secretly" $dx^1 \wedge ... \wedge dx^n$. 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 9.17. One could discover the definition of differential forms and the wedge product by starting with the change of variables theorem and trying to formalize the notion of "pulling back" by treating the notation $dy^1...dy^n$ as the formal symbol $dy^1 \wedge ... \wedge dy^n$. When approached this way, the involvement of the determinant implies the seeming-multilinearity and seeming-antisymmetry of the wedge product.

Theorem 9.18. [Lee, p. 403] (Change of variables theorem in light of the pullback, part 2).

Consider the hypotheses of the previous derivation, Derivation 9.14. *In the case that* $\mathbf{F}$ *is orientation-preserving or orientation-reversing*, the change of variables theorem can be restated as

$$\int_E f\ dy^1 \wedge ... \wedge dy^n = \begin{cases} \int_D \mathbf{F}^*(f\ dy^1 \wedge ... \wedge dy^n) & \mathbf{F} \text{ is orientation-preserving} \\ -\int_D \mathbf{F}^*(f\ dy^1 \wedge ... \wedge dy^n) & \mathbf{F} \text{ is orientation-reversing} \end{cases}.$$

To reemphasize what we discovered in the previous derivation, we again state the pullback of $f\ dy^1 \wedge ... \wedge dy^n$:

$$\mathbf{F}^*(f\ dy^1 \wedge ... \wedge dy^n) = \det(d\mathbf{F})(f \circ \mathbf{F})dx^1 \wedge ... \wedge dx^n.$$

(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).

This definition us to concisely state the change of variables theorem by involving the integrals and pullbacks of differential forms.

Theorem 9.19. [Lee, p. 404] (Change of variables theorem for top degree 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

$$\boxed{\int_E \omega = \begin{cases} \int_D \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-preserving} \\ -\int_D \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-reversing} \end{cases}}$$

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 $\mathbb{H}^n$.

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 9.20. [Lee, p. 404] (Integral of a top degree differential 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 9.21. [Lee, p. 404] (Change of variables theorem for top degree 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 $\mathbf{F} : U \to V$ is a diffeomorphism and $\omega = f\ dx^1 \wedge ... \wedge dx^n$ is a compactly supported differential n-form on V, then

$$\boxed{\int_V \omega = \begin{cases} \int_U \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-preserving} \\ -\int_U \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-reversing} \end{cases}}$$

Proof. The proof is a matter of using technical properties of diffeomorphisms. See [Lee, p. 404]. □

As promised, we are now ready to define integration on smooth manifolds.

Definition 9.22. [Lee, p. 404] (Integral of a top degree differential form that is compactly supported in a single chart over a smooth 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 8.55 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.

Integrating differential forms that are treated as functions

Derivation 9.23. (Pullback of a differential k-form, interpreted as a pointwise function, by a diffeomorphism).

The pullback of a "regular" differential form defined in Definition 9.6 induces a pullback on the differential forms that, when evaluated a point, are treated as functions. We construct this induced pullback now.

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 $\mathbf{F} : U \subseteq M \to V \subseteq N$ be a smooth function. Consider a differential k-form $\widetilde{\omega} \in \widetilde{\Omega}^k(V)$, $k \leq n$, on $V \subseteq N$. Note that $\widetilde{\omega}_\mathbf{p}$ is an alternating multilinear function for all $\mathbf{p} \in M$.

Referring back to Definition 9.6, we see the induced pullback $\widetilde{\Omega}^k\mathbf{F}^*$ is defined by:

$$\Big((\widetilde{\Omega}^k\mathbf{F}^*)(\widetilde{\omega})\Big)_\mathbf{p} := \begin{cases} \Big(\widetilde{\Lambda}^k(d\mathbf{F}_\mathbf{p})^*\Big)(\widetilde{\omega}_{\mathbf{F}(\mathbf{p})}) & \widetilde{\omega} \text{ is a differential } k\text{-form on } V,\ k \geq 1 \\ (f \circ \mathbf{F})(\mathbf{p}) & f \text{ is a differential 0-form on } V \text{ (i.e. } f \text{ is a function } V \to \mathbb{R}) \end{cases}$$

Since $\mathbf{T} \in \Lambda^k(W^*) \xmapsto{\widetilde{\Lambda}^k\mathbf{f}^*} \mathbf{T} \circ \mathbf{f} = \mathbf{f}^*(\mathbf{T}) \in \Lambda^k(V^*)$ (see the equation in Theorem 5.74 right above the boxed equation), the above becomes

$$\begin{cases} (d\mathbf{F}_\mathbf{p})^*(\widetilde{\omega}_{\mathbf{F}(\mathbf{p})}) = \widetilde{\omega}_{\mathbf{F}(\mathbf{p})} \circ d\mathbf{F}_\mathbf{p} & \widetilde{\omega} \text{ is a differential } k\text{-form on } V,\ k \geq 1 \\ (f \circ \mathbf{F})(\mathbf{p}) & f \text{ is a differential 0-form on } V \text{ (i.e. } f \text{ is a function } V \to \mathbb{R}) \end{cases}.$$

(In the first case, $(d\mathbf{F}_{\mathbf{p}})^*$ denotes the dual of the linear map $d\mathbf{F}_{\mathbf{p}}$). Therefore, $\Omega^k\mathbf{F}^*(\widetilde{\omega})$ acts on k vectors $v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by

$$\boxed{\Omega^k\mathbf{F}^*(\widetilde{\omega})(v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}})) = \widetilde{\omega}_{\mathbf{F}(\mathbf{p})}(d\mathbf{F}_{\mathbf{p}}(v_1|_{\mathbf{p}})..., d\mathbf{F}_{\mathbf{p}}(v_n|_{\mathbf{p}}) \quad \widetilde{\omega} \text{ is a differential } k\text{-form on } V,\ k \geq 1}$$

When f is a 0-form, we still have $\Omega^k\mathbf{F}^*(f) = f \circ \mathbf{F}$.

(Note, instead of using the intermediary step of the equation before the above boxed equation, we could have directly applied the boxed equation of Theorem 5.74 to the case $k \geq 1$).

The next theorem describes a way to interpret the integral of a "regular" differential form $\omega \in \Omega^k(M)$ over $U \subseteq M$ as an integral of a corresponding differential form $\widetilde{\omega} \in \widetilde{\Omega}^k(M)$ evaluated on a basis of $T_{\mathbf{p}}(M)$.

Theorem 9.24. [HH, p. 515] (Integral of a top degree differential form that is compactly supported in a single chart over a smooth manifold).

Let M be an oriented smooth n-manifold with or without boundary, let $(V, \mathbf{y})$ be a smooth oriented chart on M, let $\omega = f\ dy^1 \wedge ... \wedge dy^n \in \Omega^k(M)$ be a "regular" differential n-form on M with compact support in V, and let $\widetilde{\omega} = f\ dy^1 \widetilde{\wedge} ... \widetilde{\wedge} dy^n \in \widetilde{\Omega}^k(M)$ be the corresponding differential form that is pointwise a multilinear alternating function. Define a chart $(U, \mathbf{x})$ on $\mathbf{x}(U)$ by $U = \mathbf{y}(V)$ and $\mathbf{x} = \mathbf{y}^{-1}$, and let $\widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}$ denotes the ith coordinate function of $\mathbf{x}$ relative to the chart $(V, \mathbf{y})$. Then

$$\boxed{\int_U \omega = \pm \int_U \widetilde{\omega}\Big(\frac{\partial \mathbf{x}}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial \mathbf{x}}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}}\Big)}$$

Proof. By Definition 9.22, the integral of ω over M is

$$\int_M \omega = \int_M f\ dy^1 \wedge ... \wedge dy^n = \pm \int_{\mathbf{y}(V)} (\mathbf{y}^{-1})^*(f\ dy^1 \wedge ... \wedge dy^n) = \pm \int_U \mathbf{x}^*(f\ dy^1 \wedge ... \wedge dy^n),$$

where the $\pm$ sign depends on the orientation of the chart $(V, \mathbf{y})$. We are again using $(\mathbf{y}^{-1})^* = \mathbf{x}^*$ to denote the pullback $\Omega^k(\mathbf{y}^{-1})^* = \Omega^k\mathbf{x}^*$.

Since $\mathbf{x}^*(dy^i) = \mathbf{x}^i_{(V,\mathbf{y})}$ (recall Lemma 9.9), the above becomes

$$\pm \int_U \mathbf{x}^*(f\ dy^1 \wedge ... \wedge dy^n) = \pm \int_U (f \circ \mathbf{x})\ d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}.$$

Here, $\widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}$ denotes the ith coordinate function of $\mathbf{x}$ relative to the chart $(V, \mathbf{y})$. (Sidenote: $\mathbf{x}$ relative to $(V, \mathbf{y})$ is $\mathbf{x}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{x} = \mathbf{x}^{-1} \circ \mathbf{x} = \mathbf{I}_U$).

Definition 9.15 says that the integral of a differential form over a subset of $\mathbb{R}^n$ or $\mathbb{H}^n$ is computed by "erasing" the $d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}$. Thus, we have

$$\pm \int_U (f \circ \mathbf{x})\ d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})} = \pm \int_V f \circ \mathbf{x}.$$

We now consider the differential form $\mathbf{x}^*(\widetilde{\omega}) = (f \circ \mathbf{x})\ d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})} \in \widetilde{\Omega}^k(U)$ that corresponds to the "regular" differential form[1] $\mathbf{x}^*(\omega) = (f \circ \mathbf{x})\ d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})} \in \Omega^k(U)$. Applying Lemma 5.69 at each point in U, we have

$$d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}\Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}}\Big) = 1.$$

(Here, $\frac{\partial}{\partial \widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}}$ denotes the map $\mathbf{p} \mapsto \frac{\partial}{\partial \widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}}\Big|_{\mathbf{p}}$). Thus

[1] "Regular" here means "is not treated as a function when evaluated at a point".

$$\pm \int_U f \circ \mathbf{x} = \pm \int_U \underbrace{(f \circ \mathbf{x}) d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}}_{\mathbf{x}^*(\widetilde{\omega})} \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}} \Big) = \pm \int_U \mathbf{x}^*(\widetilde{\omega}) \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}} \Big).$$

We have translated our integral into one that is expressed in terms of a differential form acting on tangent vectors. Now, we apply the pullback from the previous theorem that acts on such differential forms to see

$$\pm \int_U \mathbf{x}^*(\widetilde{\omega}) \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}} \Big) = \pm \int_U \widetilde{\omega} \Big(d\mathbf{x} \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}} \Big), ..., d\mathbf{x} \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}} \Big) \Big) = \pm \int_U \widetilde{\omega} \Big(\frac{\partial \mathbf{x}}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial \mathbf{x}}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}} \Big).$$

($d\mathbf{x}$ is shorthand for the map $d\mathbf{x} \mapsto d\mathbf{x_p}$). □

Theorem 9.25. (Integral of a differential top degree differential 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 zero[2] for all $i \neq j$. Then

$$\int_U \omega = \sum_{i=1}^k \int_{U_i} \omega$$

Proof. The proof of this theorem adapts the proof of Proposition 16.8 in [Lee] (which was intended for a slightly different purpose in that book). 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 $\text{supp}((\mathbf{x}^{-1})^*) \subseteq \bigcup_{i=1}^k \mathbf{x}(U_i)$. Additionally, since $\mathbf{x}$ is a smooth map, then $U_i \cap U_j$ having measure zero for all $i \neq j$ implies[3] that $\mathbf{x}(U_i) \cap \mathbf{x}(U_j)$ has measure zero for all $i \neq j$. Thus, using the standard calculus theorem pertaining to "breaking up integrals" (Theorem 6.16), we see $\mathbf{x}(U_i), ..., \mathbf{x}(U_k)$ 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. □

The previous theorem is generalized by the following definition, which finalizes the definition of integration of differential forms.

Definition 9.26. [Lee, p. 408] (Integral of a top degree differential form on a smooth manifold).

Let M be an oriented smooth n-manifold WWBOC 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 an open cover of $\text{supp}(\omega)$, where the pairwise intersections have measure zero.

We define *the integral of* ω *over* M *to be*

$$\boxed{\int_M \omega := \sum_{i=1}^k \int_{U_i} \omega}$$

Note that each integral in the sum on the right hand side is interpreted with the previous definition.

It is necessary to show that this definition doesn't depend on the choice of open cover. See [**?**] for this detail.

[2] Informally, 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.

Theorem 9.27. (Integral of a differential form on a smooth manifold is linear).

Let M be an oriented smooth n-manifold WWBOC, and let $(U, \mathbf{x})$ be a smooth chart on M. Then for all constant functions $c_1, c_2 : U \subseteq M \to \mathbb{R}$ and all compactly supported differential k-forms ω, η on M, we have

$$\int_M c_1\omega + c_2\eta = c_1 \int_M \omega + c_2 \int_M \eta.$$

Proof. It suffices to prove that this theorem holds when ω and η are supported in the same single smooth chart $(U, \mathbf{x})$. After applying Definition 9.22- pull back ω and η and integrate each over $\mathbf{x}(U) \subseteq \mathbb{R}^n$- we see the theorem holds in this special case due to the linearity of the integral of a function $\mathbb{R}^n \to \mathbb{R}$. $\square$

9.3 The exterior derivative

To define a notion of differentiating differential forms, we will interpret differential forms to be objects that, when evaluated a point of a manifold, are multilinear alternating functions accepting tangent vectors as input.

Definition 9.28. [HH, p. 545] (The exterior derivative).

If V is a compact region in $\mathbb{R}^3$ with volume $|V|$, then the divergence div of a smooth vector field $\mathbf{V} : V \subseteq \mathbb{R}^3 \to \mathbb{R}^3$ is defined as flux per "infinitesimal" volume:

$$\text{div}(\mathbf{V}) := \lim_{|V| \to 0} \frac{1}{|V|} \int_{\partial V} \mathbf{V} \cdot \hat{\mathbf{n}} dS.$$

Similarly, if C is a simple closed 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 $\text{curl}(\mathbf{V})$ is defined as work per "infinitesimal" surface area:

$$\text{curl}(\mathbf{V}) \cdot \hat{\mathbf{n}} := \lim_{A(C) \to 0} \frac{1}{A(C)} \int_C \mathbf{V} \cdot d\mathbf{r} = \lim_{A(C) \to 0} \frac{1}{A(C)} \int_C \mathbf{V} \cdot \frac{d\mathbf{r}}{dt} dt.$$

We present a notion of derivative on differential forms that is defined analagously to divergence and curl.

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 (thought of as an element of $\widetilde{\Omega}^{k+1}(M)$) with compact support in U. Let $P_{\mathbf{p}}(\mathbf{v}_1, ..., \mathbf{v}_{k+1})$ denote the subset of $T_{\mathbf{p}}(U)$ spanned by $\mathbf{v}_1|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}} \in T_{\mathbf{p}}(U)$. We say $P_{\mathbf{p}}(\mathbf{v}_1, ..., \mathbf{v}_{k+1})$ is 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

$$\boxed{d\omega_{\mathbf{p}}(\mathbf{v}_1, ..., \mathbf{v}_{k+1}) := \lim_{h \to 0} \frac{1}{h^{k+1}} \int_{\partial P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})} \omega.}$$

In words, the exterior derivative $d\omega$ is evaluated at $\mathbf{p}$ on $k+1$ tangent vectors $\mathbf{v}_1|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ 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 $(k+1)$-parallelpiped, 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 $(k+1)$-paralellapiped, and as the ith 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 faces of the paralellapiped are $(k+1) - 1 = k$ dimensional). 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 $\lim_{h \to 0} \frac{1}{h}$ does not exist. We will see in the proof below that the limit *does* exist.

Theorem 9.29. [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 ith component function of $\mathbf{x}$.

For any smooth function $f : U \to \mathbb{R}$,

1. The exterior derivative (denoted df) is equal to the differential (also denoted df), where the differential of f is the differential after identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ described by Theorem 8.35).
2. $d(f\ dx^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dx^{i_k}) = df \widetilde{\wedge} dx^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dx^{i_k}$.

Notice that (1) implies that the exterior derivative of any constant differential form is the zero differential form 0.

Proof.

1. $df(P_{\mathbf{p}}(\mathbf{v})) = \lim_{h\to 0}\frac{1}{h}\Big(f(\mathbf{x}+h\mathbf{v}) - f(\mathbf{x})\Big)$. Recalling Theorem 6.9, we see that this expression is exactly the directional derivative of f at $\mathbf{p}$ in the direction of $\mathbf{v}$.

2. The $(k+1)$-parallelapiped $P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})$ has $2(k+1)$ faces. For each $h > 0$, we construct charts $\{(V^h_{\text{f}i}, \mathbf{y}_{\text{f}i})\}_{i=1}^{k+1}$ that parameterize the $k+1$ faces on the "front" of this $(k+1)$-parallelapiped and charts $\{(V^h_{\text{b}i}, \mathbf{y}_{\text{b}i})\}_{i=1}^{k+1}$ that parameterize the remaining $k+1$ faces on its "back".
   Specifically, let $\mathbf{v}_i|_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$ be the image of $v_i|_{\mathbf{p}}$ under the isomorphism $T_{\mathbf{p}}(P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})) \cong \mathbb{R}^n_{\mathbf{p}}$ that sends $\frac{\partial}{\partial y^i} \mapsto \hat{\mathfrak{e}}_i$.We define, for $\mathbf{q} \in [0,h]^k \subseteq \mathbb{R}^k$,

$$
\begin{aligned}
\mathbf{y}_{\text{f}i}^{-1}(\mathbf{q}) &:= q^1\mathbf{v}_1|_{\mathbf{p}} + ... + q^{i-1}\mathbf{v}_{i-1}|_{\mathbf{p}} + \cancel{q^i\mathbf{v}_i|_{\mathbf{p}}} + q^{i+1}\mathbf{v}_{i+1}|_{\mathbf{p}} + ... + q^{k+1}\mathbf{v}_{k+1}|_{\mathbf{p}}\\
\mathbf{y}_{\text{b}i}^{-1}(\mathbf{q}) &:= q^1\mathbf{v}_1|_{\mathbf{p}} + ... + q^{i-1}\mathbf{v}_{i-1}|_{\mathbf{p}} + h\mathbf{v}_i|_{\mathbf{p}} + q^{i+1}\mathbf{v}_{i+1}|_{\mathbf{p}} + ... + q^{k+1}\mathbf{v}_{k+1}|_{\mathbf{p}}.
\end{aligned}
$$

   Following the conventions of Theorem 9.24, we define $U^h_{\text{f}i} = \mathbf{y}(V^h_{\text{f}i})$, $\mathbf{x}_{\text{f}i} = \mathbf{y}^{-1}$ and $U^h_{\text{b}i} = \mathbf{y}(V^h_{\text{b}i})$, $\mathbf{x}_{\text{b}i} = \mathbf{y}^{-1}$ so that $\{(U^h_{\text{f}i}, \mathbf{x}_{\text{f}i})\}_{i=1}^{k+1}$ and $\{(U^h_{\text{b}i}, \mathbf{x}_{\text{b}i})\}_{i=1}^{k+1}$ are charts in $\mathbb{R}^n$ or $\mathbb{H}^n$. Note that $\mathbf{x}_{\text{b}i} = \mathbf{y}_{\text{b}i}^{-1} = \mathbf{y}_{\text{f}i}^{-1} + h\mathbf{v}_i = \mathbf{x}_{\text{f}i} + h\mathbf{v}^i$. Additionally, since we required $\mathbf{q} \in [0,h]^k \subseteq \mathbb{R}^k$, then $U^h_{\text{f}i} \sqcup U^h_{\text{b}i}$ is a cube of side length h in $\mathbb{R}^k$.
   We need to compute

$$\lim_{h\to 0}\frac{1}{h^{k+1}}\int_{\partial P_{\mathbf{p}}(hv_1|_{\mathbf{p}},...,hv_{k+1}|_{\mathbf{p}})} f\ dy^{i_1}\widetilde{\wedge}...\widetilde{\wedge}dy^{i_k}.$$

   We use Theorem 9.24 to treat the integral as the integral of a differential form evaluated on tangent vectors. Also using Theorem 9.25 to break up the domain of integration and Theorem 8.61 to account for the orientation of the boundary $\partial P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})$, the integral inside the limit becomes

$$
\begin{aligned}
\int_{\partial P_{\mathbf{p}}(\mathbf{v}_1,...,\mathbf{v}_{k+1})} f\ d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k} = &-\underbrace{\sum_{i=1}^{k+1}(-1)^{i-1}\int_{U^h_{\text{f}i}}(f\circ\mathbf{x}_{\text{f}i})\ d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k}\Big(\frac{\partial\mathbf{x}_{\text{f}i}}{\partial q^1}, ..., \cancel{\frac{\partial\mathbf{x}_{\text{f}i}}{\partial q^i}}, ..., \frac{\partial\mathbf{x}_{\text{f}i}}{\partial q^{k+1}}\Big)}_{\text{sum of integrals over front faces}}\\
&+\underbrace{\sum_{i=1}^{k+1}(-1)^{i-1}\int_{U^h_{\text{b}i}}(f\circ\mathbf{x}_{\text{b}i})\ d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k}\Big(\frac{\partial\mathbf{x}_{\text{b}i}}{\partial q^1}, ..., \cancel{\frac{\partial\mathbf{x}_{\text{b}i}}{\partial q^i}}, ..., \frac{\partial\mathbf{x}_{\text{b}i}}{\partial q^{k+1}}\Big)}_{\text{sum of integrals over back faces}}.
\end{aligned}
$$

   For $j \neq i$, we have $\frac{\partial\mathbf{x}_{\text{f}i}}{\partial q^j} = \frac{\partial\mathbf{x}_{\text{b}i}}{\partial q^j} = \mathbf{v}_j|_{\mathbf{p}}$. Since the argument of $d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k}$ in each integrand is the same, the above is

$$\sum_{i=1}^{k+1}(-1)^{i-1}\int_{U^h_{\text{f}i}\sqcup U^h_{\text{b}i}}(f\circ\mathbf{x}_{\text{b}i} - f\circ\mathbf{x}_{\text{f}i})\ d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).$$

   Recall from the hypotheses of the theorem that f is assumed to be a continuous function $U \to \mathbb{R}$, where $(\mathbf{x}, U)$ is a smooth chart on M. (Don't confuse $\mathbf{x} : U \to \mathbb{R}^n$ with $\mathbf{x}_{\text{f}i} : U^h_{\text{f}i} \subseteq \mathbb{R}^n \to U$ and $\mathbf{x}_{\text{b}i} : U^h_{\text{b}i} \subseteq \mathbb{R}^n \to U$!). We now write $f : U \to \mathbb{R}$ as a Taylor polynomial centered at $\mathbf{p}_0 \in U$ with remainder:

$$T_f(\mathbf{p}) = T^0_f(\mathbf{p}) + T^1_f(\mathbf{p}) + R_f(\mathbf{p}) = f(\mathbf{p}_0) + df_{\mathbf{p}_0}(\mathbf{p}) + R_f(\mathbf{p}),$$

   where $|R_f(\mathbf{p})| \leq C_1||\mathbf{x}(\mathbf{p})||^2$ for some $C_1 > 0$.
   Plugging this Taylor expansion into the term $f\circ\mathbf{x}_{\text{f}i} - f\circ\mathbf{x}_{\text{b}i}$ from the above integrand, we see

$$
\begin{aligned}
T_f\circ\mathbf{x}_{\text{b}i} - T_f\circ\mathbf{x}_{\text{f}i} &= (T^0_f - T^1_f + R_f)\circ\mathbf{x}_{\text{b}i} - (T^0_f - T^1_f + R_f)\circ\mathbf{x}_{\text{f}i}\\
&= (T^0_f\circ\mathbf{x}_{\text{b}i} - T^0_f\circ\mathbf{x}_{\text{f}i}) + (T^1_f\circ\mathbf{x}_{\text{b}i} - T^1_f\circ\mathbf{x}_{\text{f}i}) + (R_f\circ\mathbf{x}_{\text{b}i} - R_f\circ\mathbf{x}_{\text{f}i}).
\end{aligned}
$$

Since T^0_f is a constant function, the first term is zero: $T^0_f \circ \mathbf{x}_{\mathrm{f}i} - T^0_f \circ \mathbf{x}_{\mathrm{b}i} = 0$. We now plug this term into the integrand to complete the proof. The integral from the above sum of integrals (where the sum of integrals is inside the limit) becomes

$$\begin{aligned}&\int_{U^h_{\mathrm{f}i} \sqcup U^h_{\mathrm{b}i}} (T^1_f \circ \mathbf{x}_{\mathrm{b}i} - T^1_f \circ \mathbf{x}_{\mathrm{f}i}) d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}) \\ &+ \int_{U^h_{\mathrm{f}i} \sqcup U^h_{\mathrm{b}i}} (R_f \circ \mathbf{x}_{\mathrm{b}i} - R_f \circ \mathbf{x}_{\mathrm{f}i}) d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).\end{aligned}$$

Observe that $|R_f(\mathbf{p})| \leq C_1 || \max_{\mathbf{x}(\mathbf{p})} \mathbf{x}(\mathbf{p})||^2 = C_1 h^2$, which implies $||R_f \circ \mathbf{x}_{\mathrm{f}i} - R_f \circ \mathbf{x}_{\mathrm{b}i}||^2 \leq C_2 h^2$ for some $C_2 > 0$. Since the $\mathbf{v}_i|_{\mathbf{p}}$'s in $d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}})$ are constant with respect to the limit, which takes $h \to 0$, then $|d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}})|$ is bounded above with respect to h. Thus, the integrals containing the terms $R_f \circ \mathbf{x}_{\mathrm{b}i} - R_f \circ \mathbf{x}_{\mathrm{f}i}$ disappear in the limit. This means that the limit becomes

$$\begin{aligned}&\lim_{h \to 0} \frac{1}{h^{k+1}} \int_{\partial P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})} f \; dy^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dy^{i_k} \\ &= \lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} \int_{U^h_{\mathrm{f}i} \sqcup U^h_{\mathrm{b}i}} (T^1_f \circ \mathbf{x}_{\mathrm{b}i} - T^1_f \circ \mathbf{x}_{\mathrm{f}i}) d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}})\end{aligned}$$

Observe that the term $T^1_f \circ \mathbf{x}_{\mathrm{b}i} - T^1_f \circ \mathbf{x}_{\mathrm{f}i}$ in the integrand is $T^1_f \circ \mathbf{x}_{\mathrm{b}i} - T^1_f \circ \mathbf{x}_{\mathrm{f}i} = h df_{\mathbf{p}_0}(\mathbf{v}_i)$:

$$\begin{aligned}(T^1_f \circ \mathbf{x}_{\mathrm{b}i} - T^1_f \circ \mathbf{x}_{\mathrm{f}i})(\mathbf{p}) &= df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{b}i}(\mathbf{p})) - df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{f}i}(\mathbf{p})) = df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{b}i}(\mathbf{p})) - df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{f}i}(\mathbf{p})) \\ &= df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{f}i}(\mathbf{p}) + h\mathbf{v}_i) - d_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{f}i}(\mathbf{p})) = h df_{\mathbf{p}_0}(\mathbf{v}_i).\end{aligned}$$

Thus, the limit is

$$\lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} \int_{U^h_{\mathrm{f}i} \sqcup U^h_{\mathrm{b}i}} h df_{\mathbf{p}_0}(\mathbf{v}_i) \; d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).$$

(Here, $df_{\mathbf{p}_0}(\mathbf{v}_i) \in U^h_{\mathrm{fi}} \sqcup U^h_{\mathrm{b}i}$ denotes the constant map $\mathbf{q} \in\mapsto df_{\mathbf{p}_0}(\mathbf{v}_i)$). Since the integral is taken over points in $U^h_{\mathrm{fi}} \sqcup U^h_{\mathrm{b}i}$, then $h df_{\mathbf{p}_0}(\mathbf{v}_i)$ is constant with respect to the integral, and the above is

$$\lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} h df_{\mathbf{p}_0}(\mathbf{v}_i) \int_{U^h_{\mathrm{f}i} \sqcup U^h_{\mathrm{b}i}} d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).$$

$U^h_{\mathrm{fi}} \sqcup U^h_{\mathrm{b}i}$ is a cube in $\mathbb{R}^k$ with side length h, so the limit at last disappears, since $\frac{1}{h^{k+1}}$ is canceled by the h and h^k inside the sum:

$$\begin{aligned}&\lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} h df_{\mathbf{p}_0}(\mathbf{v}_i) h^k \; d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}) \\ &= \sum_{i=1}^{k+1} (-1)^{i-1} df_{\mathbf{p}_0}(\mathbf{v}_i) \; d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}) \\ &= \mathrm{alt}(df \widetilde{\otimes} (d\widetilde{\mathbf{x}}^{i_1} \widetilde{\otimes} ... \widetilde{\otimes} d\widetilde{\mathbf{x}}^{i_k}))(\mathbf{v}_1|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}) \\ &= df \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).\end{aligned}$$

□

Theorem 9.30. [Lee, p. 366] (The exterior derivative commutes with the pullback of a smooth function $M \to N$).

Let M and N be smooth manifolds WWBOC, let $U \subseteq M$ be open, and consider a smooth function $\mathbf{F} : U \subseteq M \to N$. Additionally, let $\mathbf{p} \in U \subseteq M$, and let $d_{\mathbf{p}}$ denote the map $f \overset{d_{\mathbf{p}}}{\mapsto} df_{\mathbf{p}}$. Then for any smooth differential k-form $\omega \in \widetilde{\Omega}^k(N)$ on N, we have

$$\mathbf{F}^* \circ d_{\mathbf{p}} = d_{\mathbf{p}} \circ \mathbf{F}^* \iff \mathbf{F}^*(d_{\mathbf{p}}\omega) = d_{\mathbf{p}}(\mathbf{F}^*(\omega)).$$

Proof. The proof of this relies on extrapolating the result of Theorem 9.8. □

Theorem 9.31. [Lee, p. 364] ($d \circ d = \mathbf{0}$).

Let M be a smooth manifold WWBOC, and let $(U, \mathbf{x})$ be a smooth chart on M. For any smooth differential form ω on U, performing the exterior derivative twice on ω produces the zero differential form. That is, $d \circ d$ is the zero map, $d \circ d = \mathbf{0}$.

Proof. According to [Ebe19], this property follows from the fact that the exterior derivative commutes with the pullback of a diffeomorphism. Here is the rough idea.

A differential form ω on $\mathbb{R}^n$ is said to be *invariant* with respect to a diffeomorphism $\mathbf{F}$ iff $\mathbf{F}^*(\omega) = \omega$. A differential form ω on $\mathbb{R}^n$ that is invariant with respect to all translation diffeomorphisms is said to be *translation-invariant*. Using the fact that the exterior derivative commutes with the pullback of a diffeomorphism, we can show that the exterior derivative of a translation-invariant differential form must also be translation-invariant.

Consider the diffeomorphism $\mathbf{F}_x : \mathbb{R}^n \to \mathbb{R}^n$, where $x \in \mathbb{R}$, defined by $\mathbf{F}_x(\mathbf{q}) = x\mathbf{q}$. Supposedly, it is easy to check that each x acts on translation-invariant differential k-forms with multiplication by x^k. Using this fact, we show that all translation-invariant differential forms ω satisfy $d\omega = 0$. Since any differential form on $\mathbb{R}^n$ can be written as a linear combination of translation-invariant differential forms on $\mathbb{R}^n$, it follows that $d^2\omega = 0$ for any differential form ω on $\mathbb{R}^n$. This result is quickly extended to differential forms on arbitrary smooth manifolds.

We now prove the claim that all translation-invariant differential forms ω satisfy $d\omega = 0$. Suppose that ω is a translation invariant differential form that is acted on by x in the manner described above. Then $x^k\omega = \mathbf{F}_x^*(\omega)$. Taking the exterior derivative, we have $d(x^k\omega) = d(\mathbf{F}_x^*(\omega))$. We take out the constant on the left hand side and use the fact that the exterior derivative commutes with the pullback of a diffeomorphism on the right hand side to get $x^k d\omega = \mathbf{F}_x^*(d\omega)$. Then, since $d\omega$ is a differential $(k+1)$-form, $\mathbf{F}_x^*(d\omega) = x^{k+1}d\omega$. In all, we have $x^k d\omega = x^{k+1}d\omega$, so $d\omega = xd\omega$. Somehow, this implies that $d\omega = 0$. Thus, we have shown that all translation invariant differential forms ω satisfy $d\omega = 0$. □

Theorem 9.32. [HH, p. 652 - 655] (Linearity of the exterior derivative).
Let M be a smooth manifold WWBOC, and let $(U, \mathbf{x})$ be a smooth chart on M.

- $d(f_1\omega + f_2\eta) = f_1 d\omega + f_2 d\eta$ for all $\omega, \eta \in \widetilde{\Omega}^k(U)$
- $d(c\omega) = cd\omega$ for all $\omega \in \widetilde{\Omega}^k(U)$ when $c : U \to \mathbb{R}$ is a constant function.

Proof. By part (1) of Theorem 9.29, the exterior derivative on 0-forms is the differential, which is linear. So the exterior derivative is linear on 0-forms. Extend this result using part (2) of Theorem 9.29, which is $d(f\ dx^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dx^{i_k}) = df \widetilde{\wedge} dx^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dx^{i_k}$. □

Remark 9.33. [Lee, p. 364] (Exterior derivative on $\Omega^k(M)$).
Let $(U, \mathbf{x})$ be a smooth chart on M. The exterior derivative we have presented operates on differential forms from $\widetilde{\Omega}^k(M)$. There is a corresponding induced exterior derivative operation d on $\Omega^k(M)$ ("regular" differential forms) satisfying

1. $d(f\ dx^{i_1} \wedge ... \wedge dx^{i_k}) = df \wedge dx^{i_1} \wedge ... \wedge dx^{i_k}$ for all smooth functions $f : U \to \mathbb{R}$.
2. The induced exterior derivative commutes with the pullback of a smooth function $M \to N$.
3. d is linear.
4. d satisfies a "product rule": $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta$, for all $\omega \in \Omega^k(U)$ and $\eta \in \Omega^\ell(U)$.

It is shown in [Lee, p. 364] that these axioms indeed determine a unique operation.

The exterior derivative and vector calculus

We now present how the exterior derivative, when coupled with the *Hodge-dual* operator (soon to be introduced), generalizes div, grad, and curl from multivariable calculus.

Definition 9.34. (Hodge-dual).
Let V be an n-dimensional vector space. The *Hodge-dual* is the linear map $\Lambda^k(V) \to \Lambda^k(V)$ that, for any ordered basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ of V, satisfies

$$(\mathbf{e}_{i_1} \wedge ... \wedge \mathbf{e}_{i_k}) \wedge *(\mathbf{e}_{i_1} \wedge ... \wedge \mathbf{e}_{i_k}) = \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n,$$

for any strictly increasing index sequence $i_1 < ... < i_k$.
That is, for any permutation $\sigma \in S_n$,

$$*(\mathbf{e}_{\sigma(1)} \wedge ... \wedge \mathbf{e}_{\sigma(k)}) = \mathrm{sgn}(\sigma)(\mathbf{e}_{\sigma(k+1)} \wedge ... \wedge \mathbf{e}_{\sigma(n)}).$$

Theorem 9.35. (Div, grad, and curl via the exterior derivative).
Consider $\mathbb{R}^3$ as a smooth manifold, and let $\sharp$ and $\flat$ be the musical isomorphisms induced by the choice of basis $\left\{ \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \right\}_{i=1}^n$ for $T_{\mathbf{p}}(\mathbb{R}^3)$. This means that

$$\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \in T_{\mathbf{p}}(M) \overset{\flat}{\mapsto} dx^i|_{\mathbf{p}} \in T^*_{\mathbf{p}}(M)$$
$$dx^i|_{\mathbf{p}} \in T^*_{\mathbf{p}}(M) \overset{\sharp}{\mapsto} \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \in T_{\mathbf{p}}(M).$$

(It may be helpful to recall Theorems 8.39 and 8.40).
Additionally, let $\mathbf{F}$ be the isomorphism $T_{\mathbf{p}}(\mathbb{R}^3) \cong \mathbb{R}^3$ sending $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \mapsto \hat{\mathfrak{e}}_i \in \mathbb{R}^3$. Then, if $f : \mathbb{R}^3 \to \mathbb{R}$ is a smooth function and $\mathbf{V} : \mathbb{R}^3 \to \mathbb{R}^3$ is a smooth vector field, we can express div, grad, and curl as:

$$\boxed{\begin{aligned} \nabla f &= \mathbf{F}((df)^\sharp) \\ \mathrm{curl}(\mathbf{V}) &= \mathbf{F}((*d(\mathbf{V}^\flat))^\sharp) \\ \mathrm{div}(\mathbf{V}) &= *d*(\mathbf{V}^\flat) \end{aligned}}$$

Proof.

- From Theorem 8.36, we have $df_{\mathbf{p}} = \sum_{i=1}^3 \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}$, so $(df)^\sharp = \sum_{i=1}^3 \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$. Then $\mathbf{F}((df)^\sharp) = \sum_{i=1}^3 \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \hat{\mathfrak{e}}_i = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}}$.
- Here's an outline of what happens in the computation. $\mathbf{V}^\flat$ is a differential 1-form corresponding to the vector field $\mathbf{V}$, so $d(\mathbf{V}^\flat)$ is a differential 2-form. The 2-wedges in the linear combination for $d(\mathbf{V}^\flat)$ get sent to the "1-wedges" $dx^1|_{\mathbf{p}}, dx^2|_{\mathbf{p}}, dx^3|_{\mathbf{p}}$ (which are really "no wedges") by $*$. Then each $dx^i|_{\mathbf{p}}$ is sent to $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ by $\sharp$. Lastly, each $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ is sent to $\hat{\mathfrak{e}}_i$ by $\mathbf{F}$. We are left with a vector field in $\mathbb{R}^3$ that is the curl of the vector field $\mathbf{V}$.
- For a smooth vector field $\mathbf{V} : \mathbb{R}^2 \to \mathbb{R}^2$, we have $\text{div}(\mathbf{V}) = \text{curl}(\mathbf{R}_{\frac{\pi}{2}}(\mathbf{V}))$, where $\mathbf{R}_{\frac{\pi}{2}}$ is counterclockwise rotation by $\frac{\pi}{2}$. This is because if $\mathbf{V} = (V_1, V_2)$, then $\mathbf{R}_{\frac{\pi}{2}}(\mathbf{V}) = (-V_2, V_1)$. Thus, we can apply the second line of the boxed equation to prove the theorem when $\mathbf{V}$ is a 2-dimensional vector field.

$\square$

Theorem 9.36. [Lee, p. 368] (Vector calculus commutative diagram).

The following commutative diagram summarizes how div, grad, and curl correspond to the exterior derivative action on differential 0-, 1-, 2-, and 3- forms on $\mathbb{R}^3$.

$$\begin{tikzcd}
C^\infty(\mathbb{R}^3 \to \mathbb{R}) \arrow[r, "\text{grad}"] \arrow[d, "\mathbf{I}"'] & \text{vector fields on } \mathbb{R}^3 \arrow[r, "\text{curl}"] & \text{vector fields on } \mathbb{R}^3 \arrow[r, "\text{div}"] & C^\infty(\mathbb{R}^3 \to \mathbb{R}) \arrow[d, "*"] \\
\Omega^0(\mathbb{R}^3) \arrow[r, "d"'] & \Omega^1(\mathbb{R}^3) \arrow[r, "d"'] & \Omega^2(\mathbb{R}^3) \arrow[r, "d"'] & \Omega^3(\mathbb{R}^3)
\end{tikzcd}$$

Note that the fact $d \circ d = \mathbf{0}$ implies $\text{curl} \circ \text{grad} = \mathbf{0}$ and $\text{div} \circ \text{curl} = \mathbf{0}$.

9.4 The generalized Stokes' theorem

Having set up integration on manifolds and having defined the exterior derivative, we are almost ready to present the generalized Stokes' theorem. To prove the generalized Stokes' theorem, we will use a tool called a *partition of unity.* A partition of unity is essentially the "theoretically nice" way to break up the support of a compactly supported differential form. Instead of breaking up the support into distinct "chunks", partitions of unity can be thought of as "smoothly fading the differential form in and out".

Definition 9.37. [Lee, p. 43] (Partitions of unity).

If $\{U_\alpha\}$ is an open cover of M and $\{f_\alpha\}$ is a set of functions $M \to \mathbb{R}$, 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 $\mathbf{p} \in M$
- $(\sum_\alpha f_\alpha(\mathbf{p}) = 1 \text{ for all } \mathbf{p} \in M) \iff (\sum_\alpha f \text{ is the multiplicative identity of } \{\text{functions } M \to \mathbb{R}\})$.

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 WWBOC admits a smooth partition of unity ([Lee, p. 43]). Since every manifold has an open cover, this means that every smooth manifold WWBOC admits a smooth partition of unity.

Theorem 9.38. [Lee, p. 43] (Existence of smooth partitions of unity).

We take for granted the fact from [Lee] that any open cover of any smooth manifold WWBOC has a smooth partition of unity.

Theorem 9.39. (Integral of a differential n-form on a smooth n-manifold via partition of unity).

Let M be an oriented smooth n-manifold WWBOC 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 an open cover of $\text{supp}(\omega)$, with the pairwise intersections having measure zero. 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 applies the definition of the integral over a manifold of a differential form that is compactly supported in a single chart (see Definition 9.22).

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:

$$\boxed{\int_M \omega = \sum_{i=1}^k \int_M f_i\omega}$$

Remark 9.40. (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 9.26), as this definition is an intuitive starting point.

Theorem 9.41. [Lee, p. 407] (Properties of integrals of differential forms).

Theorem 9.42. [HH, p. 561] (The generalized Stokes' theorem on a single smooth chart).

Let M be a smooth n-manifold and consider a smooth oriented chart $(U, \mathbf{x})$ on M. Let ω be a smooth differential $(k-1)$-form that is compactly supported in U. 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 9.26.

Proof. For simplicity, assume $M = \mathbb{R}^n$ or $\mathbb{H}^n$.

Each "approximation" above translates 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 transitive when treated this way.

($\int_U d\omega \approx \sum_i d\omega(P_i)$). Let $\epsilon > 0$. Then, by the definition of the integral via Riemann sums, 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[4] 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 $C_i = P_{\mathbf{p}}(h\hat{\mathfrak{e}}_1, ..., h\hat{\mathfrak{e}}_k)$ for some $\mathbf{p}$, and where $h = 2^{-N}$. By the definition of the exterior derivative (Definition 9.28), 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 $|\int_{C_i} d\omega - \sum_i d\omega(C_i)| < Kh^{k+1} < \epsilon$.

($\sum_i \int_{\partial P_i} \omega = \int_{\partial U} \omega$). We prove this last step with with direct equality rather than by using a converging approximation. Written out more formally, $\sum_i \int_{\partial P_i} \omega$ is $\sum_{C \in D_N(\text{cl}(\mathbb{H}^k))} \int_{\partial C} \omega$. Applying Theorem 8.61, which gives the oriented boundary of a k-parallelapiped, all the internal boundaries in the sum $\sum_{C \in D_N(\text{cl}(\mathbb{H}^k))} \int_{\partial C}$ cancel, since each boundary appears twice with opposite orientations. Thus $\sum_{C \in D_N(\text{cl}(\mathbb{H}^k))} \int_{\partial C} \omega = \sum_{C' \in D_N(\partial \text{cl}(\mathbb{H}^k))} \int_{C'} \omega = \int_{\partial M} \omega$. $\square$

Remark 9.43. The above theorem is the real "heart" of Stokes' theorem. The proof of the full-blown Stokes' theorem uses partitions of unity to extend the previous result to an entire manifold.

Theorem 9.44. [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

$$\int_M d\omega = \int_{\partial M} \omega.$$

[4]Intuitively, the dyadic decomposition of a subset of $\mathbb{R}^k$ of 2^{-N} is a partition of that subset obtained by halving the subset N times. See [HH, p. 356].

Proof. The proof of this theorem is adapted from [HH, p. 661 - 665].

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 have[5] 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 the definition of an integral of a differential form over a manifold the definition of the integral of a differential form over a manifold that is compactly supported in a single chart (Definitions 9.22 and 9.26, respectively) 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 the differential commutes with the pullback of a differential (Theorem 9.30), we have

$$\sum_{i=1}^{k} \int_{\mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(d\omega_i) = \sum_{i=1}^{k} \int_{\mathbf{x}_i(U_i)} d\Big((\mathbf{x}_i^{-1})^*(\omega_i)\Big).$$

We apply the previous theorem- Stokes' theorem on a single smooth chart- to the integral inside the sum to obtain

$$\sum_{i=1}^{k} \int_{\mathbf{x}_i(U_i)} d\Big((\mathbf{x}_i^{-1})^*(\omega_i)\Big) = \sum_{i=1}^{k} \int_{\partial\mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(\omega_i).$$

We use that $\partial\mathbf{x}_i(U_i) = \mathbf{x}_i(\partial U_i)$ and apply the definition of a differential form over a manifold that is compactly supported in a single chart, 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 open cover for $\text{supp}(\omega) \subseteq M$, where the pairwise intersections have measure zero, then $\{\partial U_i\}_{i=1}^k$ is a finite open cover for $\partial\text{supp}(\omega) \subseteq \partial M$, where the pairwise intersections have measure zero. 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 9.39. □

Remark 9.45. ("∂ is the adjoint of d").

If we define a "bilinear function" on differential forms ω and smooth manifolds M by "$\langle \omega, M\rangle := \int_M \omega$", then Stokes' theorem is stated as

$$\text{“}\langle d\omega, M\rangle = \langle \omega, \partial M\rangle\text{”}.$$

Recalling Definition 4.24, it *appears* that the boundary operator ∂ is the "adjoint" of the exterior derivative d. (Quotation marks have been used excessively in this remark because the statements of this remark are only heuristic and not mathematically precise).

[5] This 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 $([\omega]_{\{f_i\}_{i=1}^k})_i$.

Bibliography

[BW97] Javier Bonet and Richard D. Wood. Nonlinear Continuum Mechanics for Finite Element Analysis. Cambridge Univeristy Press, 1 edition, 1997.

[Ebe19] Johannes Ebert. https://mathoverflow.net/questions/21024/what-is-the-exterior-derivative-intuitively, 2019.

[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.

The source that is most frequently used is [Lee]; this is a comprehensive reference for differential geometry and manifolds. Explanations in [Lee] are generally good, but results sometimes are not motivated in the best way, and connections between concepts are sometimes missed. Following [Lee], we have used [HH] the second-most. The book [HH] possibly has the best motivation I have ever seen in a math book. The definition of the exterior derivative and proof of the generalized Stokes' theorem were adapted from [HH].