Tensors and differential forms

Ross Grogan-Kaylor

Acknowledgements

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

- 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 will have to do a thorough reading of

Every reader should take advantage of Chapter 2, the review chapter on linear algebra. 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 mostly in the realm of linear algebra (not calculus, as one might initially suspect).

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 *Linear Algebra Done Wrong*[1] by Sergei Treil (this is available for free online) or *Linear and Geometric Algebra* by Alan Macdonald for concrete examples. Another text you might consult is Steven Roman's *Advanced Linear Algebra*. Be warned: I have found no linear algebra book that *truly* 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). To be fair, *Linear Algebra Done Wrong* does come very close to some of the best possible explanations. In any case, be wary when reading about these concepts in other books.

---

[1]Funnily enough, Treil's *Linear Algebra Done Wrong* is better in many ways than Axler's *Linear Algebra Done Right*. The former explains why matrix-vector products are the way they are; the later does not, and gives an opaque definition of the matrix of a linear function relative to bases.

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 $\widehat{W} = \{\hat{\mathbf{w}}_i\}_{i=1}^n$ 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^{**}.

- $E^{**} = \{Ч_i\}_{i=1}^n$ 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 3.22. 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 **??**.

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 3.47). 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_j^i is the Kronecker delta function defined by $\delta_j^i = 1$ when $i = j$ and $\delta_j^i = 0$ 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

Logic, proofs, and functions

Notation for definitions

In this book, we use := whenever an equality is enforced to be true by definition (e.g. "define a function f by $f(x) := x^2$"), and use = whenever an equality follows from further reasoning (e.g. since $f(x) := x^2$ we have $f(x) = (x^2)^2 = x^4$).

1.1 Propositional logic

A *proposition* is a statement that is either true or false; an example is "giraffes have ears".

We can denote propositions with letters, and say things such as "let P be a proposition". When a proposition P is true, we write $P = T$. The = symbol is used here to denote *logical equality*, and the T denotes "truth". Similarly, we write $P = F$ when P is false.

Logical operators

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.

Here are some examples: we have $\big((3 > 4) \text{ and (every rectangle is a square)}\big) = F$ and $\big((5 > -3) \text{ or } (2 > 100)\big) = T$.

DeMorgan's laws

You can check by using truth tables for propositions P and Q, we have the following two logical identities, called *DeMorgan's laws*:

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

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

1.2 Predicate logic

A *predicate* is a statement involving one or more "variables" such that the statement is a proposition when the variables take on concrete values. An example of a predicate with a single variable is "giraffes have x many ears".

Just as we could with propositions, we can denote predicates with letters. For example, we can define a predicate Q by $Q(x) = $ (the sky is blue at time x). When we substitute in concrete meaning for the entities in a predicate, we obtain a proposition; for example, $Q(8)$ is the proposition "giraffes have 8 ears".

Quantifiers

In predicate logic, the symbols $\forall$ (read as "for all") and $\exists$ (read as "there exists") are used to produce propositions from single-input predicates. The symbol $\forall$ is known as the *universal quantifier*, and the symbol $\exists$ is known as the *existential quantifier*.

$$\forall x \ P(x) \qquad \text{is the proposition which states "for all } x, P(x) = T\text{"}$$
$$\exists x \ \ Q(x) \qquad \text{is the proposition which states "there exists an } x \text{ such that } Q(x) = T\text{".}$$

We see that $\forall x \ P(x)$ is true exactly when $P(x)$ is true for all x, and that $\exists x \ Q(x)$ 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 proposition $(\forall x \ x > 3)$ is false, while the proposition $(\exists x \ x > 3)$ is true[1].

Nested quantifers

In the last section, we saw that applying a quantifier-variable pair (such as $\forall x$) to a single-input predicate produced a proposition. We now consider what happens when we apply a quantifier-variable pair to a two-input predicate $P(x, y)$. There are four possible quantifier-variable pairs we can apply to $P(x, y)$:

$$\forall x \ P(x, y)$$
$$\exists x \ P(x, y)$$
$$\forall y \ P(x, y)$$
$$\exists y \ P(x, y).$$

Now we claim that while $P(x, y)$ depends on both x and y, the expression $\forall x \ P(x, y)$ depends only on y. The idea is that since the $\forall x$ specifies an iteration over all x, the only variable in $\forall x \ P(x, y)$ that is not being modified by a quantifier is y. (This intuition is similar to the principle that a sum such as $\sum_i a_{ij}$ depends on j but not i, since i is being "summed over" but j is not). Thus, we can define a single-input predicate Q_1 by $Q_1(y) := \forall x \ P(x, y)$.

Similar statements can be made for the other lines, so we can define single-input predicates Q_1, Q_2, Q_3, Q_4 as follows:

$$Q_1(y) := \forall x \ P(x, y)$$
$$Q_2(y) := \exists x \ P(x, y)$$
$$Q_3(x) := \forall y \ P(x, y)$$
$$Q_4(x) := \exists y \ P(x, y).$$

Since we can apply quantifier-variable pairs to the single-input predicates $Q_1(y), Q_2(y), Q_3(y), Q_4(y)$, we obtain the following propositions:

$$\forall x \ (\forall y \ P(x, y))$$
$$\forall x \ (\exists y \ P(x, y))$$
$$\exists x \ (\forall y \ P(x, y))$$
$$\exists x \ (\exists y \ P(x, y)).$$

These propositions involve so-called *nested quantifiers*. In practice, we omit the inner parentheses and use syntax such as $\forall x \ \exists y \ P(x, y)$ rather than syntax such as $\forall x \ (\exists y \ P(x, y))$.

---

[1]Technically, these propositions aren't really sensible since we haven't specified that the x's involved can only be numbers and not other entities.

Quantificational identities

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:

$$\forall x \ \forall y \ P(x,y) = \forall y \ \forall x \ P(x,y)$$
$$\exists x \ \exists y \ Q(x,y) = \exists y \ \exists x \ Q(x,y).$$

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

$$\forall x, y \ P(x,y) := \forall x \ \forall y \ P(x,y)$$
$$\exists x, y \ P(x,y) := \exists x \ \exists y \ P(x,y).$$

Negating quantifications

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

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

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

Nested quantifiers can be negated with this rule, as well. To negate a nested quantifier, just treat the inner quantifier-predicate pair as a single-input predicate. For example,

$$\text{not} \ (\forall x \ \exists y \ P(x,y)) = \exists x \ \text{not} \ (\exists y \ P(x,y)) = \exists x \ \forall y \ \text{not} \ P(x,y).$$

This algorithm works for any number of nested quantifiers (i.e. we could use it to compute not $(\forall x \ \exists y \ \exists z \ \forall w \ P(x,y,z,w)))$ because if we have n quantifiers, then the predicate involving the innermost $n-1$ quantifiers is always a single-input predicate.

Essentially, all of math is expressed using predicate logic

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 $\exists$, predicates, and logical operators derived from *and*, *or*, and *not*

- the collection of all propositions ("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 in terms of previously defined notions, then thinking about these objects in intuitive terms while still keeping the rigorous definitions in mind, and then 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.

Set construction and membership

Set construction

The notation

$$\{x \mid P(x)\}$$

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

Sometimes, authors prefer a colon : over a vertical bar $\mid$ and use the notation

$$\{x : P(x)\}$$

instead of the above.

Set membership

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".

Shorthand for set construction

We define the following shorthand:

$$\{x \in S \mid P(x)\} := \{x \mid x \in S \text{ and } P(x)\}.$$

You would read "$\{x \in S \mid P(x)\}$" out loud as "x in S such that $P(x)$".

---

[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".

Implications with sets

We now define the *implication* operator $\implies$, which is an operation that accepts two propositions as input and produces a proposition as output. 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 on propositions P and Q 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 after a "for all" quantifier*. "For all" quantifications that involve $\implies$ in the following way are what correspond to the English language meaning of "implies":

$$\forall x \; P(x) \implies Q(x).$$

We might express the above in English as "P implies Q". 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, $P(x)$ implies $Q(x)$". 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 of entities for which P is satisfied by using $\implies$; it is only possible for the predicate $(P(x) \implies Q(x))$ to be false when $P(x)$ is true.

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 $P(x) := (x \text{ is a square})$ and $Q(x) := (x \text{ is a rectangle})$. 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.

Bounded quantifiers

Motivated by the above, we define

$$\forall x \in S \; Q(x) := \forall x \; x \in S \implies Q(x)$$
$$\exists x \in S \; Q(x) := \exists x \; x \in S \text{ and } Q(x).$$

This new syntax provides an easy way to use quantifiers that effectively only consider objects that satisfy the predicate $(x \in S)$. For example, the above claim that all squares are rectangles would be written as

$$\forall x \in (\text{set of all squares}) \; x \text{ is a rectangle}.$$

The notations $\forall x \in S$ and $\exists x \in S$ are *bounded quantifiers* because they effectively restrict, i.e. bound, the scope of quantification.

Negation of an implication

We have the following negation rule for implication:

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

This implies the following more intuitive fact:

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

The wordy explanation is that if it is not the case that $P(x)$ implies $Q(x)$ for all x, then there must be some x satisfying $P(x)$ that does not also satisfy $Q(x)$.

Some "thematic inverses" to the above facts are the following:

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

Negating bounded quantifiers

The negations of bounded quantifications are what you expect:

$$\text{not}\Big(\forall x \in S \ P(x)\Big) = \exists x \in S \text{ not } P(x)$$

$$\text{not}\Big(\exists x \in S \ P(x)\Big) = \forall x \in S \text{ not } P(x).$$

Use the facts $\text{not}(P \implies Q) = (P \text{ and } (\text{not } Q))$ and $\text{not}(P \text{ and } Q) = (P \implies (\text{not } Q))$ from the "Negation of an implication" section to prove this.

More about implications

Converses

Given propositions P and Q, consider the implication $P \implies Q$. The *converse* to this implication is the implication $Q \implies P$. Please note that $(P \implies Q) = (Q \implies P)$ is not true for all P, Q! For example, it is true that "whenever it rains, there are clouds", but it is not true that "whenever there are clouds, it rains"!

The contrapositive

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

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

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

You could verify this identity by using truth tables. Here is a nicer proof that uses[3] the facts $(P \text{ or } Q) = (Q \text{ or } P)$ and $Q = \text{not}(\text{not } Q)$:

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

The contrapositive is practically important because it is sometimes easier to prove the contrapositive of an implication than it is to prove the implication itself.

Proof by contradiction

Another way of proving $P \implies Q$ that is similar in spirit to a proof by contrapositive is *proof by contradiction*. In a proof by contradiction, one assumes $(\text{not}(P \implies Q))$ and then derives from this assumption that some additional proposition R and its negation $\text{not}(R)$ are *both* true. Since for any proposition R we have $(R \text{ and } (\text{not } R)) = F$, it intuitively follows that the original assumption $(\text{not}(P \implies Q))$ must have been false; that is, we must have $P \implies Q$.

In summary, a proof of $P \implies Q$ by contradiction is a proof of

$$\text{not}(P \implies Q) \implies (R \text{ and } (\text{not } R)) = F,$$

where R is some proposition.

Since we have $(\text{not}(P \implies Q)) = (P \text{ and } (\text{not } Q))$, we see that a proof of $P \implies Q$ by contradiction is equivalently a proof of

$$(P \text{ and } (\text{not } Q)) \implies (R \text{ and } (\text{not } R)) = F.$$

This is the form of proof by contradiction that is used in practice. Authors typically begin a proof of $P \implies Q$ by contradiction by saying "Assume P. For the sake of obtaining a contradiction, also assume not Q." A contradiction $(R \text{ and } (\text{not } R))$ is then derived, and the author writes the following to finish the proof: "This contradicts the assumption not Q. Therefore Q must be true."

Contradiction vs. contrapositive

We now investigate the difference between proof by contradiction and proof by contrapositive.

When the contradiction $(R \text{ and } (\text{not } R))$ derived in a proof by contradiction is $(P \text{ and } (\text{not } P))$, the overall proof by contradiction is

$$(P \text{ and } (\text{not } Q)) \implies (P \text{ and } (\text{not } P)) = F.$$

The above implies the following:

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

---

[3]We will soon learn more about the assumption $\text{not}(\text{not}(Q)) = Q$.

If we are additionally in a context where we know P to be true, then the above further simplifies to:

$$(\text{not } Q) \implies (\text{not } P),$$

which is the contrapositive. Thus, we see that when P is known to be true and the contradiction obtained is $(P \text{ and } (\text{not } P))$, the proof by contradiction can be rewritten to be a proof by contrapositive. This situation occurs very often in practice.

"If and only if"

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

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

That is, $P \iff Q$ is true exactly when the implication $P \iff Q$ and its converse are both true. 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 $(P \iff Q) = (Q \iff P)$.

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

"Necessary" and "sufficient"

Let P and Q be predicates, 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. Similarly, due to the reasoning "one of ways to make Q happen is to make P happen", P is said to be a *sufficient* condition for Q.

Notice that when $\forall x \ P(x) \iff Q(x)$ iff P and Q are both necessary and sufficient conditions for each other (i.e. P is a necessary and sufficient for Q and Q is a necessary and sufficent condition for P).

Types of proofs

Consider the statement

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

A proof of this statement can be written in a couple of ways. Sometimes, the reasoning can be stated straightforwardly enough that one can write a series of sentences that are "for all" statements:

- "For all $x \in S$ we have <insert some property of x>. Since <insert reasoning here>, then <some other property of x> for all $x \in S$. Thus, for all $x \in S$ we have $P(x)$, as desired.

Other times, the analysis of $x \in S$ is complicated enough that it is too verbose to contain in a sequence of such sentences. In such cases, we use a phrase such as "Let $x \in S$", "Assume $x \in S$", or "Fix $x \in S$" at the beginning of a proof to allude to the syntax "$\forall x \in S$". For example:

- "Let $x \in S$. <Insert very long, elaborate reasoning, which probably takes up a paragraph or so, here>. Therefore $P(x)$, as desired."

The statement

$$\forall x \in S \ P(x) \implies Q(x)$$

can be proved in similar ways. If the reasoning is relatively straightforward, then a proof like the following, in which each sentence is a "for all" statement, can be used:

- "If $P(x)$, then <insert some property of x> for all $x \in S$. Since <insert reasoning here>, then <some other property of x> for all $x \in S$. Therefore $Q(x)$ for all $x \in S$."

Otherwise, when analysis of $x \in S$ with $P(x)$ is too verbose, we allude to the syntax "$\forall x \in S \ P(x) \implies$" by using a phrase such as "Let $x \in S$ and $P(x)$", "Assume $x \in S$ and $P(x)$", or "Fix $x \in S$ and $P(x)$":

- "Let $x \in S$ and assume $P(x)$. <Insert very long, elaborate reasoning, which probably takes up a paragraph or so, here>. Therefore $Q(x)$, as desired.

When reading proofs of the implication $\forall x \in S \ P(x) \implies Q(x)$, note that the author may not explicitly restate every single hypothesis at the beginning of their proof. For instance, the author may not start their proof with either of "If $x \in S$ and $P(x)$" or "Assume $x \in S$ and $P(x)$", opting instead to start it with "If $P(x)$" or "Assume $P(x)$".

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 and write $T \subseteq S$ iff $\forall x \in S \; x \in T$. We say T is a *proper subset* of S iff T is a subset of S and $T \neq S$.

For all sets S, the empty set $\emptyset$ is a subset of S because the proposition $(\forall x \in \emptyset \; x \in S)$ is[4] "true by false hypothesis". The more intuitive fact that $S \subseteq S$ for all sets S is also true.

Notice, the similarity of the subset notation $\subseteq$ for sets to the "less than or equal to" $\leq$ notation for numbers is a helpful reminder of the fact that for any set S we have $S \subseteq S$. A little less intuitive is that we write $T \subsetneq S$ iff T is a proper subset of S; we don't use $T \subset S$ to indicate "proper subset". It's true that the $\subset$ notation would be more analogous to the $<$ notation for numbers, but, unfortunately, some authors write $T \subset S$ to mean $T \subseteq S$.

There is no universal set in ZFC

It seems natural that there would exist a set which contains every set. However, this is not possible in ZFC. Why? Suppose there did exist such a "universal set" U. Then we can consider the subset V of U which contains all non-self-containing sets, $V := \{S \in U \mid S \notin S\}$. Now notice that we have $(V \in V \iff V \notin V)$. This is clearly a contradiction. Thus, there is no "set of all sets" in ZFC.

Common sets

We now define notation for many common infinite sets.

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

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

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 used as an indexing set[5], though we tend to use sets of numbers as indexing sets. The set of elements in S indexed by I is written as

$$\{x_i \in S \mid i \in I\}.$$

---

[4]$(\forall x \in \emptyset \; x \in S) = (\forall x \; x \in \emptyset \implies x \in S) = (\forall x \; F \implies x \in S) = \forall x \; T = T$.

[5]A technical point is that a set I can only index *all* elements of a set S iff there exists an onto function $I \to S$. If there does not exist such a function, I can only index a proper subset of S. We will learn about onto functions later.

Union and intersection

We now define ways to build sets from a collection[6] of sets. Let C be a set which contains sets as elements.

The *union* of the sets in C is defined to be the result of collecting all the elements of all of the sets in C. 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.

$$\bigcup_{S \in C} S := \{x \mid \exists S \in C \ x \in S\} \quad \text{(union)}$$

$$\bigcap_{S \in C} S := \{x \mid \forall S \in C \ x \in S\} \quad \text{(intersection)}.$$

When the collection C is finite and contains $n \in \mathbb{N}$ sets, we often index C by the set $\{1, ..., n\}$, so that $C = \{S_i \mid i \in \{1, ..., n\}\}$, and denote the unions and intersections, respectively, by the following three notations:

$$\bigcup_{i \in \{1, ..., n\}} S_i = \bigcup_{i=1}^{n} S_i = S_1 \cup ... \cup S_n$$

$$\bigcap_{i \in \{1, ..., n\}} S_i = \bigcap_{i=1}^{n} S_i = S_1 \cap ... \cap S_n.$$

When C contains just two sets, $C = \{S, T\}$, we write $S \cup T$ rather than $\bigcup_{S \in C} S$ and $S \cap T$ rather than $\bigcap_{S \in C}$. Applying the above definitions[7], we see that

$$S \cup T = \{x \mid x \in S \text{ or } x \in T\}$$

$$S \cap T = \{x \mid x \in S \text{ and } x \in T\}.$$

Here are some examples: we have $\{a, b, c\} \cup \{b, c, d\} = \{a, b, c, d\}, \emptyset \cap \{a, b\} = \{a, b\}$ and $\{a, b, c\} \cap \{b, c, d\} = \{b, c\}, \{a, b\} \cap \{c, d\} = \emptyset$.

When considered as operations on two sets, we see that $\cup$ and $\cap$ are associative:

$$\forall S \ \forall T \ \forall R \ (S \cup T) \cup R = S \cup (T \cup R)$$

$$\forall S \ \forall T \ \forall R \ (S \cap T) \cap R = S \cap (T \cap R).$$

So, we can "drop the parentheses" and define for sets S, T, and R:

$$S \cup T \cup R := (S \cup T) \cup R = S \cup (T \cup R)$$

$$S \cap T \cap R := (S \cap T) \cap R = S \cap (T \cap R).$$

Set difference and set complement

Set difference

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

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

Most authors denote the set difference $S - T$ with the notation $S \setminus T$.

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 $S^C := U - S$.

---

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

[7]To formally prove the first line of the claimed result, show that $\{x \mid \forall S \in C \ x \in S\}$ and $\{x \mid x \in S \text{ and } x \in T\}$ are subsets of each other. The second line is proved similarly.

Cartesian products

Ordered pairs

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

$$\forall x_1 \ \forall x_2 \ \forall y_1 \ \forall y_2 \ (x_1, y_1) = (x_2, y_2) \iff (x_1 = y_1 \text{ and } x_2 = y_2).$$

One definition that gives us the above characterizing property is $(x, y) := \{x, \{y\}\}$. There are other definitions that work, too. The choice of which particular definition we use for (x, y) isn't really important, and is really just an "implementation detail". What's most important is that the proposition "there exists a mathematical object with the above characterizing property" is true.

Tuples

Similar to the concept of ordered pair is the concept of "n-tuples". An *n-tuple*, $n \in \mathbb{N}$, is a ordered list of n entities, notated in the form "$(x_1, ..., x_n)$", such that the following characterizing property holds for all indexed sets $\{x_i \mid i \in \{1, ..., n\}\}$ and $\{y_i \mid i \in \{1, ..., n\}\}$:

$$(x_1, ..., x_n) = (y_1, ..., y_n) \iff (\forall i \in \{1, ..., n\} \ x_i = y_i).$$

To prove that there do indeed exist mathematical objects with such a characterizing property, we rely on the concept of an ordered pair. We define the n-tuple $(x_1, ..., x_n)$ to be the ordered pair $((x_1, ..., x_{n-1}), x_n)$, where when $n = 2$ the n-tuple (x_1, x_n) is defined to be the ordered pair (x_1, x_2) and when $n = 1$ the n-tuple (x_1) is defined to be the element x_1. Of course, we could have also used the definition $(x_1, ..., x_n) := (x_1, (x_2, ..., x_n))$; but as before, this is only an "implementation detail", and what's really important is that the proposition "there exists a mathematical object with the above characterizing property" is true.

Cartesian products

The *Cartesian product* of sets S and T is the set $S \times T$ of all ordered pairs whose elements are from S and T:

$$S \times T := \{(s, t) \mid s \in S \text{ and } t \in T\}.$$

The Cartesian product $\times$ is unfortunately not associative: for sets S, T, and R, we have

$$\begin{aligned} (S \times T) \times R &= \{((s, t), r) \mid s \in S \text{ and } t \in T \text{ and } r \in R\} \\ &\neq \{(s, (t, r)) \mid s \in S \text{ and } t \in T \text{ and } r \in R\} = S \times (T \times R). \end{aligned}$$

To circumvent this issue, we define $\times$ to act on n sets $S_1, ..., S_n$ as follows:

$$\underset{i \in \{1, ..., n\}}{\bigtimes} S_i = \bigtimes_{i=1}^{n} S_i = S_1 \times ... \times S_n := \{(s_1, ..., s_n) \mid \forall i \in \{1, ..., n\} \ s_i \in S_i\}.$$

Note that here, $(s_1, ..., s_n)$ is an n-tuple.

nth Cartesian products

Sometimes it's useful to take the Cartesian product of a set with itself multiple times. For this situation, we define, for a set S and $n \in \mathbb{N}$,

$$S^n := \bigtimes_{i=1}^{n} S.$$

One important example is that the identity $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ can be interpreted to state that the plane ($\mathbb{R}^2$) can be thought of as the Cartesian product of the "x-axis" (the first $\mathbb{R}$) with the "y-axis" (the second $\mathbb{R}$).

Abuse of notation

We now define a common abuse of notation related to Cartesian products:

$$(\forall x_1, ..., x_n \in S) := (\forall (x_1, ..., x_n) \in S^n).$$

For example, $\forall x, y \in S$ means $\forall (x, y) \in S^2$.

1.5   Relations

Definition 1.1. (Relation).
Let S and T be sets. A *relation on S and T* is simply a subset of $S \times T$. Given a relation R on S and T, we write xRy to mean $(x, y) \in R$, and say "x is related to y" iff xRy.

Definition 1.2. (Inverse relation).
Given a relation R on set S and T, the *inverse relation* to R is the relation on T and S denoted by R^{-1} and defined by $R^{-1} := \{(y, x) \in T \times S \mid (x, y) \in R\}$. That is, $yR^{-1}x$ iff xRy.

Equivalence relations

Definition 1.3. (Equivalence relation).
An *equivalence relation on a set S* is a relation on S and S that behaves like the equality relation, and thus satisfies the following properties:

- (Reflexivity). $\forall x \in S \ xRx$.

- (Symmetry). $\forall x \in S \ \forall y \in S \ xRy \implies yRx$.

- (Transitivity). $\forall x \in S \ \forall y \in S \ \forall z \in S \ xRy$ and $yRz \implies xRz$.

Definition 1.4. (Equivalence class).

Definition 1.5. (Quotient set).

Functions

Definition 1.6. (Uniqueness).
We say that there exists a *unique* entity x satisfying a predicate P, and write $\exists! x \ P(x)$, iff there is only one entity x for which $P(x)$ is true. Formally, we have $\exists! x \ P(x) := (\exists x \ P(x))$ and $(\forall y \ P(y) \implies y = x)$.

Definition 1.7. (Function).
Let X and Y be sets. A relation f on X and Y is said to be a *function from X to Y*, and is denoted $f : X \to Y$, iff for all $x \in X$ there exists a unique $y \in Y$ such that $(x, y) \in f$.
When f is a function from X to Y, we write $f(x) = y$ to mean $(x, y) \in f$.
Functions are also commonly called *mappings* or *maps*. They are less commonly called *unique assignments*[8].

Remark 1.8. (Well-definedness and uniqueness).
Technically, we should check the well-definedness of every function we define. It is obvious that many functions are well-defined, though, so we don't have to do this a lot of the time.
[have to do it when equivalence classes at play]

Definition 1.9. (Domain, codomain, image).
Let X and Y be sets, and let $f : X \to Y$ be a function. The set X is called the *domain of f* and the set Y is called[9] the *codomain of f*.
The *image of a subset $X_1 \subseteq X$ under f*, denoted $f(X_1)$, is the subset of the codomain Y that is obtained by applying f to all elements of X_1:

---

[8]In mathematics-adjacent literature, it is common for authors to say something like "suppose f uniquely assigns objects from X to objects in Y" instead of saying "suppose f is a function from X to Y".

[9]You may have heard of the *range* of a function in your precalculus class. We do not define the term "range" because it is ambiguous. Some authors use it to mean "image", and others use it to mean "codomain".

$$f(X_1) := \{y \in Y \mid \exists x \in X_1 \; f(x) = y\}.$$

The *image of f* is defined to be the image of the domain under f, $\operatorname{im}(f) := f(X)$.

The *preimage of a subset $Y_1 \subseteq Y$ under f*, denoted $f^{-1}(Y_1)$, is the subset of the domain X whose image is Y_1:

$$f^{-1}(Y_1) = \{x \in X \mid f(x) \in Y_1\}.$$

Definition 1.10. (Function equality).

Let X and Y be sets. We define an equivalence relation $=$ on functions from X to Y so that $f = g \iff (f(x) = g(x) \text{ for all } x \in X)$.

Proof. Checking that this $=$ is an equivalence relation is left to the reader. $\qquad\square$

Definition 1.11. (Function composition).

Let X, Y, Z be sets and consider functions $f : X \to Y$ and $g : Y \to Z$. The *composition of f with g* is the function $g \circ f : X \to Z$ defined by $(g \circ f)(x) := g(f(x))$.

Theorem 1.12. (Function composition is associative).

Let X_1, X_2, X_3, X_4 be sets, and consider functions $f : X_1 \to X_2$, $g : x_2 \to X_3$, and $h : X_3 \to X_4$. We have $(h \circ g) \circ f = h \circ (g \circ f)$.

Proof. We have $((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x)) = h((g \circ f)(x)) = (h \circ (g \circ f))(x)$. Since $((h \circ g) \circ f)(x) = (h \circ (g \circ f))(x)$ for all $x \in X$, then $h \circ (g \circ f) = (h \circ g) \circ f$. $\qquad\square$

Definition 1.13. (Identity function).

Let X be a set. The *identity function on X* is the function id_X defined by $\operatorname{id}_X(x) = x$.

Inverse functions

Definition 1.14. (Invertiblity).

A function is said to be *invertible* iff its inverse relation f^{-1} is a function.

Theorem 1.15. (Invertibility characterization).

If X and Y are sets and $f : X \to Y$ is an invertible function, then the inverse function $f^{-1} : Y \to X$ satisfies $f^{-1}(f(x)) = x$ for all $x \in X$ and $f(f^{-1}(y)) = x$ for all $y \in Y$. That is, $f^{-1} \circ f = \operatorname{id}_X$ and $f \circ f^{-1} = \operatorname{id}_Y$.

Proof. For all $x \in X$, we have $(x, f(x)) \in f$, and, since f^{-1} is the inverse relation of f, we have $(f(x), x) \in f^{-1}$. Thus $f^{-1}(f(x)) = x$ for all $x \in X$.

Similarly, for all $y \in Y$, we have $(y, f^{-1}(y)) \in f^{-1}$, and, since f is the inverse relation of f^{-1}, we have $(f^{-1}(y), y) \in f$. Thus $f(f^{-1}(y)) = y$ for all $y \in Y$. $\qquad\square$

Definition 1.16. (Left and right inverses).

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 $\ell \circ f = \operatorname{id}_X$. A *right-inverse* of f is a function $r : Y \to X$ such that $f \circ r = \operatorname{id}_Y$.

Theorem 1.17. A function is invertible iff it has a left inverse and a right inverse.

Proof.

($\implies$). If f is invertible, then f^{-1} is both a left and right inverse.

($\impliedby$). Suppose f has a left inverse ℓ and a right inverse r. Then $\ell = \ell \circ \operatorname{id}_Y = \ell \circ (f \circ r) = (\ell \circ f) \circ r = \operatorname{id}_X \circ r = r$. If we define $g := l = r$, then g satisfies $f \circ g = f \circ r = \operatorname{id}_Y$ and $g \circ f = \ell \circ f = \operatorname{id}_X$, so g is the inverse of f. This shows the inverse of f exists. $\qquad\square$

'jections

Definition 1.18. (One-to-one functions).

Let X and Y be sets. A function $f : X \to Y$ is said to be *one-to-one* iff for all $x, y \in X$ implies $x \neq y \implies f(x) \neq f(y)$. That is, a function is one-to-one iff are mapped to the same output by the function.

Taking the contrapositive of this implication, we see that f is one-to-one iff for all $x, y \in X$ implies $f(x) = f(y) \implies x = y$.

Theorem 1.19. Let f be a function. Then f is one-to-one $\iff$ f has a left inverse.

Proof.

($\implies$). Assume f is one-to-one.

First, let $y \in f(X)$. Since $y \in f(X)$, there is some $x \in X$ such that $y = f(x)$; since f is one-to-one, there is no $x_1 \in X - \{x\}$ such that $y = f(x_1)$. So, we see that for every $y \in f(X)$ there is a unique x such that $y = f(x)$.

Thus, we may define $\ell : f(X) \to X$ by setting $\ell(y)$ to be the unique x for which $f(x) = y$. Clearly, we have $\ell(f(x)) = x$ for all $x \in X$, so ℓ is a left inverse of f.

($\impliedby$). Assume f has a left inverse ℓ. We need to show the implication $\forall x, y \in X \ f(x) = f(y) \implies x = y$. So, assume $f(x) = f(y)$. Applying the left inverse to both sides of this equation, we obtain $x = y$, as desired. $\square$

Remark 1.20. (One-to-one functions and uniqueness).

Notice that the statement "x is the unique object satisfying $P(x)$" is equivalent to the statement "$P(x)$ and P is one-to-one".

Definition 1.21. (Onto functions).

Let X and Y be sets. A function $f : X \to Y$ is said to be *onto* iff for all $y \in Y$ there exists an $x \in X$ such that $f(x) = y$. That is, a function is onto iff every possible output is mapped to by the function.

Theorem 1.22. Let f be a function. f is onto $\iff$ f has a right inverse.

Proof.

($\implies$). Assume f is onto. Then for all $y \in Y$ there exists[10] an $x_y \in X$ such that $y = f(x)$. Set $X_1 := \cup_{y \in Y} x_y$. Define a right inverse $r : Y \to X_1 \subseteq X$ by $r(y) = x_y$. We have $f(r(x)) = f(x_y) = y$, so r is a right inverse of f.

($\impliedby$). Assume f has a right inverse. We need to show that for all $y \in Y$ there exists an $x_y \in X$ such that $y = f(x)$. This is simple: we have $x_y = r(y)$, since $f(r(y)) = y$. $\square$

Remark 1.23. (Left and right inverses of noninvertible functions are not unique).

Left and right inverses of noninvertible functions are not unique.

(Left inverses). If f is not onto, then additional left inverses of f can be constructed from an existing left inverse ℓ of f, since it is irrelevant how any left inverse acts on elements of Y that are not in the image of f. More specifically, the function $\widetilde{\ell}_g$, defined below, is a left inverse for any function $g : (Y - f(X)) \to X$:

$$\widetilde{\ell}_g(y) := \begin{cases} \ell(y) & y \in f(X) \\ g(y) & y \notin f(X) \end{cases}.$$

(Right inverses). If f is not one-to-one, then for some $y \in Y$, there are at least two $x_y \in X$ such that $y = f(x_y)$. Thus, there are at least two possible values for the set $X_1 := \cup_{y \in Y} x_y$, meaning there are at least two possible values right inverses $r : Y \to X_1 \subseteq X$.

Remark 1.24. (Injections and surjections).

One-to-one functions are also called *injections*, or *injective* functions, and onto functions are called *surjections*, or *surjective* functions.

Definition 1.25. (Bijection).

A function is a *bijection* iff it is an injection and a surjection. That is, a bijection is a function that is both one-to-one and onto.

Theorem 1.26. (Invertible iff bijection).

---

[10]You may wonder at why we write x in the definition of an onto function rather than the x_y here. The answer is that we only write x_y when it useful. In a proof, the notation is useful; in the particular definition we gave, it is not.

Proof. f is invertible iff it has a left inverse and a right inverse. f has a left inverse iff it is one-to-one, and has a right inverse iff it is onto. Therefore f is invertible iff it is one-to-one and onto, i.e., if it is a bijection. $\square$

Remark 1.27. (Restrictions of surjections are bijections).

The following almost-tautological statement can come in very handy when proving that a function is invertible: if $f : X \to Y$ is one-to-one, then $f : X \to f(X)$ is a bijection.

For example, since $(-x)^2 = x^2$ for all $x \in \mathbb{R}$, then $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not one-to-one and is thus not invertible. On the other hand, the function $g : \{x \geq 0 \mid x \in \mathbb{R}\} \to \{x \geq 0 \mid x \in \mathbb{R}\}$ defined by the same formula, $g(x) = x^2$, is a bijection and is thus invertible.

Cardinality of sets

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

Definition 1.28. (Proof by induction).

For the future

- 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 10.23. Somehow, $d\mathbf{x}\left(\frac{\partial}{\partial \widetilde{x}^i_{(V,\mathbf{y})}}\right) = \frac{\partial \mathbf{x}}{\partial \widetilde{x}^i_{(V,\mathbf{y})}}$.

- 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 4.16, 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 10.41 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 point).

Intuitively, a *point* is a location in space.

Definition 2.2. (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).

It is tempting to try to define either *point* in terms of *vector* or *vector* in terms of *point*, but this is actually impossible.

Remark 2.3. (Circularity in intuitive definitions of point and directed line segment).

- (Attempted definition of *vector* in terms of *point*).

  Suppose that we try to define *vector* in terms of *point* as follows: a *vector* is a difference of the form $\mathbf{p}_2 - \mathbf{p}_1$, where $\mathbf{p}_1$ and $\mathbf{p}_2$ are points. This definition doesn't work because any intuitively interpretable definition of "difference of points" is "difference of the corresponding[1] vectors"; which is clearly a circular definition because it involves *vector*.

- (Attempted definition of *point* in terms of *vector*).

  Suppose that we use the above definition of *vector* and attempt to use it to give a definition of *point* that is equivalent to the above definition of *point*. The attempted definition is: a *point* is a sum of the form $(((\mathbf{q} + \mathbf{v}_1) + \mathbf{v}_2) + ...) + \mathbf{v}_n$, where $\mathbf{q}$ is any point, each $\mathbf{v}_i$ is a vector, and where the sum of a point and a vector is the point obtained by starting on the point and "traveling" according to the vector. This definition is circular because it involves *point*.

In addition to the logical circularities, also notice that these attempted definitions suggest that points and vectors are inextricably intertwined: the first attempted definition requires a correspondence between vectors and points, and the second attempted definition requires a notion of adding a point with a vector.

The above result forces us to define *point* and *vector* independently of each other; we have to define *point* in terms of some more primitive mathematical object, and *vector* in terms of another more primitive mathematical object. (We will actually use the same primitive mathematical object for both *point* and *vector*). Notice that even after we have done this, the above detailed intertwinedness and logical circularities will always persist on the intuitive level, for better or for worse.

---

[1]At this point it is unclear if a correspondence between vectors and points even exists. If we assume there is such a correspondence, though, we see the argument is circular.

$\mathbb{R}^n$

We will use elements of $\mathbb{R}^n$ as the primitive mathematical objects that underlie the notion of *point* and *vector*.

Definition 2.4. ($\mathbb{R}^n$).

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 2.5. (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 2.6. (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, we note that 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 $c \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$ 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". The following theorem shows this.

Theorem 2.7. (Properties of $+$ and $\cdot$ in $\mathbb{R}^n$).

1. (Properties of $+$).

   1.1. (Existence of additive identity). There exists $\mathbf{0} \in \mathbb{R}^n$ such that for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{0}+\mathbf{x} = \mathbf{x}$. Specifically, we have

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

   1.2. (Closure under additive inverses). For all $\mathbf{x} \in \mathbb{R}^n$ there exists $-\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$. Specifically, we have

   $$-\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} -x_1 \\ \vdots \\ -x_n \end{pmatrix}.$$

   1.3. (Associativity of $+$). For all $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^n$ we have $(\mathbf{x}_1 + \mathbf{x}_2) + \mathbf{x}_3 = \mathbf{x}_1 + (\mathbf{x}_2 + \mathbf{x}_3)$.

   1.4. (Commutativity of $+$). For all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ we have $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{x}_2 + \mathbf{x}_1$.

2. (Properties of $\cdot$).

   2.1. (Normalization for $\cdot$). For all $\mathbf{x} \in \mathbb{R}^n$ we have $1\mathbf{x} = \mathbf{x}$.

   2.2. (Left-associativity of $\cdot$). For all $\mathbf{x} \in \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$ we have $c_2(c_1\mathbf{x}) = c_2 c_1 \mathbf{x}$.

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

   3.1. For all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ and $c \in \mathbb{R}$ we have $c(\mathbf{x}_1 + \mathbf{x}_2) = c\mathbf{x}_1 + c\mathbf{x}_2$.

   3.2. For all $\mathbf{x} \in \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$ we have $(c_1 + c_2)\mathbf{x} = c_1\mathbf{x} + c_2\mathbf{x}$.

Proof.

1. The claims about the additive identity and additive inverses follow from the fact that 0 is the additive identity in $\mathbb{R}$ and that the additive inverse of $x \in \mathbb{R}$ is denoted by $-x$. The associativity and commutativity of $+$ follow from the associativity and commutativity of addition of real numbers.

2. The normalization property of $\cdot$ follows from the fact that 1 is the multiplicative identity in $\mathbb{R}$. The left-associativity of $\cdot$ follows from the associativity of real numbers.

3. Both compatibility properties follow from the facts that we have $c(x_1 + x_2) = cx_1 + x_2$ for all $c, x_1, x_2 \in \mathbb{R}$ and $(c_1 + c_2)x = c_1 x + c_2 x$ for all $c_1, c_2, x \in \mathbb{R}$. In other words, they follow from the fact that $\cdot$ distributes over $+$ in $\mathbb{R}$ and from the fact that $\cdot$ is commutative in $\mathbb{R}$.

$\square$

Points and vectors in $\mathbb{R}^n$

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

Definition 2.8. (Points and vectors in $\mathbb{R}^n$).
A *point* in $\mathbb{R}^n$ is an element of $\mathbb{R}^n$. A *vector* in $\mathbb{R}^n$ is also an element of $\mathbb{R}^n$. (That was pretty simple, wasn't it?)

When we have $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, we can interpret $\mathbf{x}$ to be either a point or vector. The numbers $x_1, ..., x_n \in \mathbb{R}$ are called the *entries*, or *components*, of $\mathbf{x}$.

If we interpret $\mathbf{x}$ to be a point, then each x_i is the point's displacement on the x_i axis. If we interpret $\mathbf{x}$ to be a vector, then each x_i is the displacement, on the x_i axis, of the end of $\mathbf{x}$ relative to the start of $\mathbf{x}$, where-since vectors are locationless- we additionally allow the start of $\mathbf{x}$ to be any point.

Remark 2.9. (Intertwinedness of points and vectors).
The above definition easily implements the intertwinedness of points and vectors; since every point in $\mathbb{R}^n$ can also be interpreted as a vector in $\mathbb{R}^n$, the correspondence between points and vectors is given by the identity function.

Even though there is always some circularity on an intuitive level between the concepts of *point* and *vector*; the setting of $\mathbb{R}^n$ grounds the following geometric intuitions.

Remark 2.10. (Geometric intuition about points and vectors).
(Geometric intuition about $+$).

- (Point plus vector). Let $\mathbf{p} \in \mathbb{R}^n$ be a point and $\mathbf{v} \in \mathbb{R}^n$ be a vector. We interpret $\mathbf{p} + \mathbf{v}$ to be the point that is obtained by placing the start of $\mathbf{v}$ on $\mathbf{p}$ and then "traveling" according to $\mathbf{v}$.

- (Vector plus vector). Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ be vectors. We interpret $\mathbf{v}_1 + \mathbf{v}_2$ to be the result of starting at the start point of $\mathbf{v}_1$ (which of course can be any point), then traveling according to the vector $\mathbf{v}_1$ to reach the end point of $\mathbf{v}_1$, then applying "locationlessness" to place the start point of $\mathbf{v}_2$ on the end point of $\mathbf{v}_1$, and then traveling to the end point of $\mathbf{v}_2$. That is, the directed line segment that reaches from the "overall" beginning to the "overall" end is $\mathbf{v}_1 + \mathbf{v}_2$.

  - If we use locationlessness of to make the end point of $\mathbf{v}_1$ coincide with the start point of $\mathbf{v}_2$ before we start the above process, then $\mathbf{v}_1 + \mathbf{v}_2$ is represented by the directed line segment that connects the start point of $\mathbf{v}_1$ to the end point of $\mathbf{v}_2$. This interpretation is often called the "tail-to-tip method", since one computes $\mathbf{v}_1 + \mathbf{v}_2$ by forming the vector by drawing a straight line from the "tail" of $\mathbf{v}_1$ to the "tip" of $\mathbf{v}_2$.

- ($\mathbf{0}$ as a point). When $\mathbf{0}$ is considered to be a point, we interpret it to be the "central reference point" of the coordinate system, i.e., the origin.

- ($\mathbf{0}$ as a vector). When $\mathbf{0}$ is interpreted to be a vector, we interpret it to be the locationless directed line segment that has a length of zero.

- (Additive inverses as points). When $\mathbf{p} \in \mathbb{R}^n$ is a point, there is no satisfactory interpretation of $-\mathbf{p}$.

- (Additive inverses as vectors). When $\mathbf{v} \in \mathbb{R}^n$ is a vector, the fact $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ forces us to interpret $-\mathbf{v}$ to be the unique vector whose end is the start of $\mathbf{v}$ if its start is made to coeinide with the end of $\mathbf{v}$.

(Geometric intuition about $\cdot$).

- (Scalar times point). Let $c \in \mathbb{R}$. When $\mathbf{p} \in \mathbb{R}^n$ is a point, there is no satisfactory interpretation of $c\mathbf{p}$.

- (Scalar times vector). Let $c > 0$. When $\mathbf{v} \in \mathbb{R}^n$ is a vector, we interpret $c\mathbf{v}$ to be the vector that is obtained by stretching $\mathbf{v}$ by a factor of c (when $c > 1$) or that is obtained by shrinking $\mathbf{v}$ by a factor of c (when $c \in (0,1)$). When $d < 0$, we interpret $d\mathbf{v}$ to be $-(|d|\mathbf{v})$; that is, we interpret $d\mathbf{v}$ to be the additive inverse of $|d|\mathbf{v}$.

Definition 2.11. ("Vector" means "vector" or "point" depending on context).
In practice, we say "vector in $\mathbb{R}^n$" to mean "element of $\mathbb{R}^n$", and determine whether $\mathbf{v} \in \mathbb{R}^n$ is interpreted to be a point or a locationless directed line segment from context.

Vector-containing sets

At this point, we know that we can make the intuitive notions of "vector" and "point" precise by grounding them in the concrete setting of $\mathbb{R}^n$. This works, of course, but thinking about of vectors in $\mathbb{R}^n$ has a good deal of unnecessary "implementation detail". When thinking about a single vector in $\mathbb{R}^n$, one is secretly thinking about the several numbers that are the vector's components!

Thankfully, the previous theorem (Theorem 2.7) provides an opportunity to escape dealing with components. The theorem states key properties of vectors in $\mathbb{R}^n$ (such as "there exists $\mathbf{0} \in \mathbb{R}^n$ such that for all $\mathbf{x} \in \mathbb{R}^n$ we have $\mathbf{x} + \mathbf{0} = \mathbf{x}$", and "for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ and $c \in \mathbb{R}$ we have $c(\mathbf{x}_1 + \mathbf{x}_2) = c\mathbf{x}_1 + c\mathbf{x}_2$") in terms of the vectors themselves, and *not* in terms of their components!

The properties of this helpful theorem are in general the properties that we expect of vectors. Thus, we can leave behind the baggage of $\mathbb{R}^n$ by effectively defining vectors to be objects that have the properties of the theorem. We do this now.

Definition 2.12. (Vector-containing set over $\mathbb{R}$).

Suppose that T is a set for which there exist functions $+ : T \times T \to T$ and $\cdot : \mathbb{R} \times T \to T$. We say that a set $S \subseteq T$ is a *vector-containing set over* $\mathbb{R}$ iff the operations $+$ and $\cdot$ satisfy the typical[2] 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[3] listed in Theorem 2.7:

1. (Properties of $+$).

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

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

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

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

2. (Properties of $\cdot$).

    2.2. (Normalization for $\cdot$). For all $\mathbf{v} \in T$ we have $1\mathbf{v} = \mathbf{v}$.

    2.1. (Left-associativity of $\cdot$). For all $\mathbf{v} \in T$ and $c_1, c_2 \in \mathbb{R}$ we have $c_2(c_1\mathbf{v}) = c_2 c_1 \mathbf{v}$.

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

    3.1. For all $\mathbf{v}_1, \mathbf{v}_2 \in T$ and $c \in \mathbb{R}$ we have $c(\mathbf{v}_1 + \mathbf{v}_2) = c\mathbf{v}_1 + c\mathbf{v}_2$.

    3.2. For all $\mathbf{v} \in T$ and $c_1, c_2 \in \mathbb{R}$ we have $(c_1 + c_2)\mathbf{v} = c_1\mathbf{v} + c_2\mathbf{v}$.

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*.

As we discussed before, this abstract characterization of a vector-containing set is useful because it decouples our notion of "vector" from the setting of $\mathbb{R}^n$. It allows us to see that many types of objects- which we might have overlooked before- can be interpreted to be vectors.

Remark 2.13. (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, where $+$ and $\cdot$ are as usual (this one isn't intended to be surprising)

- $\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \end{pmatrix} \right\} \subseteq \mathbb{R}^2$, where $+$ and $\cdot$ are the same as the functions $+$ and $\cdot$ on $\mathbb{R}^2$

- $\{\mathbf{v}_1, ..., \mathbf{v}_k\} \subseteq S$, where S is another vector-containing set, and $+$ and $\cdot$ are the same as the functions $+$ and $\cdot$ on S

- 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

---

[2]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!

[3]You may be confused as to why T, rather than S, is involved in all of the above conditions when we are defining what it means for S to be a vector-containing set. This is because we want S to have *most* of the properties that T does, but not all of them. Specifically: even though the above conditions require $\mathbf{0} \in T$ and $(\forall \mathbf{v} \ \mathbf{v} \in T \iff -\mathbf{v} \in T)$, they do not require $\mathbf{0} \in S$ or $(\forall \mathbf{v} \ \mathbf{v} \in S \iff -\mathbf{v} \in S)$.

- 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 a little more abstraction. (Don't worry, though. This abstraction is much easier to grasp than the leap from "vectors in $\mathbb{R}^n$" to "elements of a vector-containing set"). 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*.

Definition 2.14. (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[4] 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 $c_1, c_2 \in K$ we have $c_1 + c_2 \in K$.

    1.2. (Existence of additive identity). There exists $0 \in K$ such that for all $c \in K$ we have $0 + c = c$.

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

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

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

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 $c_1, c_2 \in K$ we have $c_1 \cdot c_2 \in K$.

    2.2. (Existence of multiplicative identity). There exists $1 \in K$ such that for all $c \in K$ we have $1 \cdot c = c$.

    2.3. (Associativity of $\cdot$). For all $c_1, c_2, c_3 \in K$ we have $(c_1 \cdot c_2) \cdot c_3 = c_1 \cdot (c_2 \cdot c_3)$.

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

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

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

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 2.15. (Examples and non-examples of fields).

- The integers $\mathbb{Z}$ are *not* a field because the multiplicative inverse of any $x \neq 1$ is not an integer.

- The rational numbers $\mathbb{Q} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z} \right\}$ is a field.

- $\mathbb{R}$ is a field.

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

Remark 2.16. (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[5].

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

Definition 2.17. (Vector-containing set over a field).
Suppose that K is a field, that T is a set, and that there exist functions $+ : T \times T \to T$ and $\cdot : K \times T \to T$. If $S \subseteq T$ is a set, 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:

---

[4]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".

[5]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.

1. (Properties of $+$).

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

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

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

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

2. (Properties of $\cdot$).

   2.2. (Normalization for $\cdot$). For all $\mathbf{v} \in T$ we have $1\mathbf{v} = \mathbf{v}$.

   2.1. (Left-associativity of $\cdot$). For all $\mathbf{v} \in T$ and $c_1, c_2 \in K$ we have $c_2(c_1\mathbf{v}) = c_2 c_1 \mathbf{v}$.

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

   3.1. For all $\mathbf{v}_1, \mathbf{v}_2 \in T$ and $c \in K$ we have $c(\mathbf{v}_1 + \mathbf{v}_2) = c\mathbf{v}_1 + c\mathbf{v}_2$.

   3.2. For all $\mathbf{v} \in T$ and $c_1, c_2 \in K$ we have $(c_1 + c_2)\mathbf{v} = c_1\mathbf{v} + c_2\mathbf{v}$.

(If you are confused as to why T, rather than S, is involved in all of the above conditions, see the footnote above the conditions in Definition 2.12).

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

Now that our notion of vector-containing set is complete, we can make use of vector-containing sets.

Most often, we interpret vectors to be points. In the setting of the vector-containing set $\mathbb{R}^n$, it is easy to describe a point: you simply list the displacements $x_1, ..., x_n \in \mathbb{R}$ of the point from each of the mutually perpendicular coordinate axes (we discussed this in Remark 2.10). We can't describe a vector from a general vector space this way, though, because the vectors are not necessarily elements of $\mathbb{R}^n$! We can do something

similar, though. Just as $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ is a weighted sum of the vectors $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, ..., \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$, we can

consider weighted sums of vectors from vector-containing sets. The following definition gives a name for vectors that are the result of such weighted sums.

Definition 2.18. (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 $\mathbf{v}_1, ..., \mathbf{v}_k$.

As we continue to investigate vectors, we will primarily be interested in vector-containing sets that are closed under the operation of taking linear combinations. The next two definitions formalize this idea. In the second definition, we see that such vector-containing sets are called *vector spaces*.

Definition 2.19. (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 $\mathbb{R}^3$.

Definition 2.20. (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 $V = \mathrm{span}(S)$.

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".

Remark 2.21. (Vector-containing sets are subsets of vector spaces).

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.

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

Remark 2.22. (Examples of vector spaces).

Here are some examples of vector spaces:

- $\mathbb{R}^n$ over $\mathbb{R}$

- K^n over K, where K is a field

- The set of polynomials with coefficients in a field K 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:

- $\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \end{pmatrix} \right\} \subseteq \mathbb{R}^2$.

- $\{\mathbf{v}_1, ..., \mathbf{v}_k\} \subseteq S$, where S is a vector-containing set

- $\{z \in \mathbb{R} \mid z \geq ax + by\} \subseteq \mathbb{R}^3$, where $a, b \in \mathbb{R}$.

- $\{x_n \in \mathbb{R} \mid x_n \geq c_1 x_1 + ... + c_k x_k\} \subseteq \mathbb{R}^n$, where $c_1, ..., c_k \in \mathbb{R}$, $k \leq n - 1$.

Remark 2.23. ($\emptyset$ is not a vector space).

The empty set $\emptyset$ is not a vector space over any field because it is not equal to the span of any set. (A set is either empty or not. The span of the empty set is $\{0\}$, and the span of a nonempty set is nonempty. Thus no set has span equal to the empty set).

Remark 2.24. (Why is the generality of "vector space" useful?)

In this book, we will often consider an 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 2.25. (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:

There is a vector-containing set S for which $V = \mathrm{span}(S)$.

$$\Longleftrightarrow$$

For all $\mathbf{v}_1, \mathbf{v}_2 \in V$ we have $\mathbf{v}_1 + \mathbf{v}_2 \in V$ ("V is closed under vector addition")

and

for all $c \in K$ and $\mathbf{v} \in V$ we have $c\mathbf{v} \in V$ ("V is closed under vector scaling").

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 2.26. (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 $\mathbf{v} \in V$, $c\mathbf{v} \in V$.

We know that if V is a vector space, then V is a vector-containing set, and thus that there is a set $T \supseteq V$ that contains zero and the additive inverse of every vector in V. We now see that the fact that vector spaces are spanned by vector-containing sets implies these important vectors are members of all vector spaces.

Theorem 2.27. (Vector spaces have zero and additive inverses).
Let V be a vector space. Then for any $\mathbf{v} \in V$,

1. The zero vector is $\mathbf{0} = 0\mathbf{v}$.

2. The additive inverse of $\mathbf{v}$ is $-\mathbf{v} = (-1)\mathbf{v}$.

Because V is closed under vector scaling, both of these vectors are in V.

Proof. We've already explained that if the given formulas for the zero vector and additive inverses are true, then V contains the zero vector and all additive inverses. We now show that the formulas are true.

1. Let $\mathbf{v}, \mathbf{w} \in V$. We want to show that $0\mathbf{v} + \mathbf{w} = \mathbf{w}$. We have $0\mathbf{v} + \mathbf{w} = 0\mathbf{v} + \mathbf{v} + (\mathbf{w} - \mathbf{v}) = (0+1)\mathbf{v} + \mathbf{w} - \mathbf{v} = \mathbf{v} + \mathbf{w} - \mathbf{v} = \mathbf{w}$, as desired.

2. For all $\mathbf{v} \in V$ we have $\mathbf{v} + (-1)\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = (1-1)\mathbf{v} = 0\mathbf{v} = \mathbf{0}$.

$\square$

Vector subspaces

Some vector spaces are subsets of vector spaces. We quickly discuss such spaces.

Definition 2.28. (Vector subspace).
If V and W are vector spaces and $W \subseteq V$, then W is a *vector subspace* of V.

Since every vector subspace is a subset of a vector space, then every vector subspace is a vector-containing set. This fact yields a simple characterization of vector subspaces.

Theorem 2.29. (Vector subspace characterization).
If W is a subset of a vector space V, then W is a vector subspace of V iff W is closed under vector addition and vector scaling. (See Theorem 2.26 for the precise meaning of this).

Remark 2.30. (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 *linearly independent vectors*, which is a generalization of the notion of "non-parallel vectors" to higher dimensions[6].

Definition 2.31. (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 the vectors in S are *linearly independent* from each other iff for all $\mathbf{v} \in S$ we have $\mathbf{v} \notin \text{span}(S - \{\mathbf{v}\})$.

Intuitively, the vectors in S are linearly independent if every vector inside S is required to produce $\text{span}(S)$. Indeed, we prove in Lemma 2.35 that removing a vector from a linearly independent set produces a set whose span is a proper subset of what it was before.

Vectors are said to be *linearly dependent* on each other iff they are not linearly independent from each other.

Remark 2.32. ($\emptyset$ is linearly independent). The empty set $\emptyset$ is linearly independent by false hypothesis.

Theorem 2.33. 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 $(\forall \mathbf{v} \in S\ \mathbf{v} \notin \text{span}(S - \{\mathbf{v}\}))$ is violated by $\mathbf{v} = \mathbf{0}$, since $\mathbf{0}$ is in the span of every set. $\square$

Theorem 2.34. (Condition for linear independence).

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

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

Proof. Set $S := \{\mathbf{v}_1, ..., \mathbf{v}_k\}$.

($\implies$). Assume S is linearly dependent. Then S is not linearly independent, so there is some i for which $\mathbf{v}_i \in \text{span}(S - \{\mathbf{v}_i\}) = \text{span}(\{\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \mathbf{v}_k\})$. This makes $\mathbf{v}_i$ a linear combination of $\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \mathbf{v}_k$, $\mathbf{v}_i = c_1 \mathbf{v}_1 + ... + \cancel{c_i \mathbf{v}_i} + ... + c_k \mathbf{v}_k$, where $c_j \in K$ for $j \in [1, k]$. Subtracting $\mathbf{v}_i$ from both sides of the previous equation, we have $(c_1 \mathbf{v}_1 + ... + c_{i-1} \mathbf{v}_{i-1}) - \mathbf{v}_i + (c_{i+1} \mathbf{v}_{i+1} + ... + c_k \mathbf{v}_k) = \mathbf{0}$, i.e., we have $d_1 \mathbf{v}_1 + ... + d_k \mathbf{v}_k = \mathbf{0}$, where $d_1 = c_1, ..., d_{i-1} = c_{i-1}, d_i = -1, d_{i+1} = c_{i+1}, ..., d_k = c_k$. Thus, as desired, there are $d_1, ..., d_k \in K$ not all zero such that $d_1 \mathbf{v}_1 + ... + d_k \mathbf{v}_k = \mathbf{0}$.

($\impliedby$). Assume there are $c_1, ..., c_k \in K$ not all zero such that $c_1 \mathbf{v}_1 + ... + c_k \mathbf{v}_k = \mathbf{0}$. Then $c_i \mathbf{v}_i = c_1 \mathbf{v}_1 + ... + \cancel{c_i \mathbf{v}_i} + ... + c_k \mathbf{v}_k$. Since $c_i \neq 0$, we can divide by c_i to obtain $\mathbf{v}_i = -\frac{c_1}{c_i} \mathbf{v}_1 - ... - \frac{c_i}{\cancel{c_i}} \cancel{\mathbf{v}_i} - ... - \frac{c_k}{c_i} \mathbf{v}_k$. Thus $\mathbf{v}_i \in \text{span}(\{\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \mathbf{v}_k\}) = \text{span}(S - \{\mathbf{v}_i\})$; $\mathbf{v}_i$ violates the linear independence condition, so S is linearly dependent. $\square$

Lemma 2.35. (Adding and removing vectors from linearly independent and dependent sets).

1. Appending a vector to a linearly independent set produces...

    1.1. another linearly independent set, if the vector isn't in the span of the original set.

    1.2. a linearly dependent set, if the vector is in the span of the original set.

2. Removing a vector from a linearly independent set produces another linearly independent set.

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. $\square$

---

[6]In fact, we will see that linear independence helps us define what *dimension* even means!

Basis and dimension

In this subsection, we 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". (We will come to see that this minimality is equivalent to linear independence). The *dimension* of a *finite-dimensional* vector space will be defined to be the number of vectors in any basis for it.

Definition 2.36. (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 2.37. (Basis for a finite-dimensional vector space).

Let V and W be finite-dimensional vector spaces. We say that a subset $E \subseteq W$ is a *basis* for V iff

1. E spans V.

2. E is minimal: if F is another set of vectors spanning V, then $|F| \geq |E|$.

Remark 2.38. (Bases are automatically subsets of the vector spaces they span).

Notice that if E spans V, then it is automatically true that $E \subseteq V$, since for any $\mathbf{v} \in E$ we have $\mathbf{v} \in \text{span}(E) = V$.

Definition 2.39. (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.

Theorem 2.40. Every finite-dimensional vector space has a basis.

Proof. Let V be a finite-dimensional vector space. Let $C = \{S \mid \text{span}(S) = V \text{ and } S \text{ is finite}\}$. By the well-ordering principle, we can select E to be the element of minimal size from C. This E is a minimal spanning set for V and thus a basis for V. $\qquad\square$

Remark 2.41. (The 0-dimensional vector space).

The empty set $\emptyset$ spans $\{\mathbf{0}\}$ because [...]. It is also a minimal spanning set because there are no sets "smaller" than $\emptyset$. Therefore $\emptyset$ 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 two lemmas enable us to obtain a more useful characterization of bases.

Lemma 2.42. (*n* linearly independent vectors taken from the same *n*-dimensional vector space span that vector space).

Let V be a finite-dimensional vector space. If a finite set $S \subseteq V$ of vectors is linearly independent and $|S| = \dim(V)$, then S spans V.

Proof. Suppose $S = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is linearly independent and that $n = \dim(V)$. We immediately see that $V = \{\mathbf{0}\}$ and that $\text{span}(S) = \text{span}(\emptyset) = \{\mathbf{0}\} = V$ if $n = \dim(V) = 0$, so assume $n > 0$.

Let K be the field over which V is taken to be a vector space. Since V is n-dimensional, we can take a basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ of V. Since $\text{span}(E) = V$, then there are $c_1, ..., c_n \in K$ such that $\mathbf{v}_1 = c_1 \mathbf{e}_1 + ... + c_n \mathbf{e}_n$. Since S is linearly independent, then $\mathbf{v}_1 \neq \mathbf{0}$, and some c_i must be nonzero. Since we can divide by c_i, we can solve for $\mathbf{e}_1$ in terms of $c_1, ..., c_n, \mathbf{v}_1, \mathbf{e}_2, ..., \mathbf{e}_n$, and conclude $\mathbf{e}_1 \in \text{span}(\{\mathbf{v}_1, \mathbf{e}_2, ..., \mathbf{e}_n\})$.

Now let $\mathbf{v} \in V$. Again because $\text{span}(E) = V$, there are $d_1, ..., d_n \in K$ such that $\mathbf{v} = d_1 \mathbf{e}_1 + ... + d_n \mathbf{e}_n$. Since $\mathbf{e}_1 \in \text{span}(\{\mathbf{v}_1, \mathbf{e}_2, ..., \mathbf{e}_n\})$, we can express $\mathbf{e}_1$ as a linear combination of $\mathbf{v}_1, \mathbf{e}_2, ..., \mathbf{e}_n$; doing this, we see $\mathbf{v}$ is a linear combination of $\mathbf{v}_1, \mathbf{e}_2, ..., \mathbf{e}_n$. Thus $\mathbf{v} \in \text{span}(\{\mathbf{v}_1, \mathbf{e}_2, ..., \mathbf{e}_n\})$. Since $\mathbf{v}$ is arbitrary, then $V \subseteq \text{span}(\{\mathbf{v}_1, \mathbf{e}_1, ..., \mathbf{e}_n\})$, which implies $V = \text{span}(\{\mathbf{v}_1, \mathbf{e}_1, ..., \mathbf{e}_n\})$.

Repeating this process inductively, we can conclude that $V = \text{span}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{e}_3, ..., \mathbf{e}_n\})$, and ultimately that $V = \text{span}(\{\mathbf{v}_1, ..., \mathbf{v}_n\})$. $\qquad\square$

Lemma 2.43. (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 from V is linearly dependent.

2. A set of less than n vectors from V does not span V.

Proof. First we show the logical equivalence.

($(1) \implies (2)$). Suppose for contradiction that a set S of vectors with $|S| < n$ spans V. Remove vectors from S until a linearly independent set T is obtained. By item 4 of Lemma 2.35, removing vectors from linearly dependent sets does not change their span, so T spans V. Thus, T is a linearly independent spanning set with $|T| < n$. Since $n > |T|$, it follows from (1) that T is linearly dependent; contradiction.

($(2) \implies (1)$). Use a similar argument that invokes item 1.1 of Lemma 2.35.

Now we show (1) is true. Consider a set $S = \{\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{v}_{n+1}, ..., \mathbf{v}_k\} \subseteq V$ of more than n vectors from V. By the previous lemma, we know $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ spans V. Thus for each $i \in \{n+1, ..., k\}$, we have $\mathbf{v}_i \in \operatorname{span}(\{\mathbf{v}_1, ..., \mathbf{v}_n\}) \subseteq \operatorname{span}(S - \{\mathbf{v}_i\})$. Since there is a $\mathbf{v} \in S$ such that $\mathbf{v} \in \operatorname{span}(S - \{\mathbf{v}\})$, then S is linearly dependent. $\qquad\square$

Theorem 2.44. (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 E spans V and E is linearly independent.

Proof.

We need to 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, then by item 4 from Lemma 2.35 we can remove a vector from E without changing $\operatorname{span}(E) = V$, so, E is not minimal; contradiction.

($\impliedby$). Suppose that E spans V and that E is linearly independent. Since E spans V, item 2 from the previous lemma implies $|E| \geq \dim(V)$; since E is linearly independent, item 1 from previous lemma implies $|E| \leq \dim(V)$. Therefore $|E| = \dim(V)$. Now we can see the assertion "if F is a spanning set of V then $|F| \geq |E|$" is true. If F is a spanning set of V, then item 2 of the previous lemma implies $|F| \geq \dim(V) = |E|$.

(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). We discussed how the first statement is true in Remark 2.41. The second statement, which is equivalent to ($\forall \mathbf{v} \in \emptyset \ \mathbf{v} \notin \operatorname{span}(\emptyset - \{\mathbf{v}\})$), is true by false hypothesis, since[7] there is no $\mathbf{v}$ for which "$\mathbf{v} \in \emptyset$" is a true statement. $\qquad\square$

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 2.45. (Basis).

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

Remark 2.46. (No need for the well-ordering principle!).

With this new definition of "basis", it is now possible to prove that every finite-dimensional vector space has a basis without using the well-ordering principle, and thus without the axiom of choice[8]. This is because linear independence is in some sense a "local" property (to test if a set S is linearly independent, one only needs to know S), while the minimal spanning condition of before is a "global" property (to test if a set S is a minimal spanning set, one needs to compare against all other spanning sets). Here's an outline of the proof: given a spanning set S of a vector space, a linearly independent spanning set (a basis) can be obtained by removing vectors from S.

Definition 2.47. (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. $\qquad\square$

---

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

[8]The axiom of choice is equivalent to the well-ordering principle.

Remark 2.48. (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 2.49. (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.

2.2   Linear functions

At the beginning of the chapter, 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 2.50. (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 $\mathbf{v}_1, \mathbf{v}_2 \in V$:

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

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 2.51. ("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 2.52. (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. $\qquad\square$

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

Theorem 2.53. (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. $\qquad\square$

Theorem 2.54. (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 $\mathbf{v}$.

This means that only lines that pass through the origin can be images of linear functions.

Theorem 2.55. (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 $\mathbf{g}(\mathbf{v}) := \sum_{i=1}^{n}([\mathbf{v}]_E)_i \mathbf{w}_i$.

We now check that $\mathbf{g}(\mathbf{e}_i) = \mathbf{w}_i$. We have $\mathbf{g}(\mathbf{e}_i) = \sum_{j=1}^{n}([\mathbf{e}_i]_E)_j \mathbf{w}_j = \sum_{j=1}^{n} \delta_{ij} \mathbf{w}_j = \mathbf{w}_i$, as desired.

Lastly, we have to check that $\mathbf{g}$ is linear. This is straightforward and left as an exercise. $\qquad\square$

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

Remark 2.56. (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, convince yourself that each above function $\mathbf{f}$ satisfies $\mathbf{f}(\mathbf{v} + \mathbf{w}) = \mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w})$ and $\mathbf{f}(c\mathbf{v}) = c\mathbf{v}$. (For example, if $\mathbf{f}$ is a rotation about the origin, then $\mathbf{f}(\mathbf{v} + \mathbf{w}) = \mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w})$ is intuitively true because rotating the sum $\mathbf{v} + \mathbf{w}$ intuitively produces the same result as summing the rotated vectors $\mathbf{f}(\mathbf{v})$ and $\mathbf{f}(\mathbf{w})$).

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

Theorem 2.57. 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 2.54 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, $\mathbf{f}(\mathbf{v})$. $\qquad\square$

Now that we are somewhat familiar with linear functions, we present a more abstract result, which formalizes the way in which we can think of linear functions as themselves being vectors.

Theorem 2.58. (Vector space of linear functions).
Let V and W be vector spaces over a field K. The set $\mathcal{L}(V \to W)$ of linear functions $V \to W$, taken together with $+ : V \to V \to V$ being function addition and $\cdot : V \to K \to V$ being function scalar multiplication, is a vector space over K.

Proof. According to Theorem 2.26, we need to show that $\mathcal{L}(V \to W)$ is a vector-containing set and that $\mathcal{L}(V \to W)$ is closed under vector addition and vector scaling. We leave showing that $\mathcal{L}(V \to W)$ is a vector-containing set as an exercise, and show the other two conditions.

- (Closure under vector addition). Assume $\mathbf{f}, \mathbf{g} \in \mathcal{L}(V \to W)$, and consider $\mathbf{f} + \mathbf{g}$. We need to show $(\mathbf{f} + \mathbf{g}) \in \mathcal{L}(V \to W)$; that is, we need to show (1) that $(\mathbf{f} + \mathbf{g})(\mathbf{v} + \mathbf{w}) = (\mathbf{f} + \mathbf{g})(\mathbf{v}) + (\mathbf{f} + \mathbf{g})(\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$ and (2) that $(\mathbf{f} + \mathbf{g})(c\mathbf{v}) = c(\mathbf{f} + \mathbf{g})(\mathbf{v})$ for all $\mathbf{v} \in V$ and $c \in K$.

  1. We have $(\mathbf{f} + \mathbf{g})(\mathbf{v} + \mathbf{w}) = (\mathbf{f} + \mathbf{g})(\mathbf{v}) + (\mathbf{f} + \mathbf{g})(\mathbf{w}) = \mathbf{f}(\mathbf{v}) + \mathbf{g}(\mathbf{v}) + \mathbf{f}(\mathbf{w}) + \mathbf{g}(\mathbf{w}) = \mathbf{f}(\mathbf{v}) + \mathbf{g}(\mathbf{v}) + \mathbf{f}(\mathbf{w}) + \mathbf{g}(\mathbf{w}) = (\mathbf{f} + \mathbf{g})(\mathbf{v}) + (\mathbf{f} + \mathbf{g})(\mathbf{w})$.
  2. We have $(\mathbf{f} + \mathbf{g})(c\mathbf{v}) = \mathbf{f}(c\mathbf{v}) + \mathbf{g}(c\mathbf{v}) = c\mathbf{f}(\mathbf{v}) + c\mathbf{g}(\mathbf{v}) = c(\mathbf{f}(\mathbf{v}) + \mathbf{g}(\mathbf{v})) = c(\mathbf{f} + \mathbf{g})(\mathbf{v})$.

- (Closure under vector scaling). Assume $\mathbf{f} \in \mathcal{L}(V \to W)$ and $c \in K$. We need to show $c\mathbf{f} \in \mathcal{L}(V \to W)$; that is, we need to show (1) that $(c\mathbf{f})(\mathbf{v} + \mathbf{w}) = (c\mathbf{f})(\mathbf{v}) + (c\mathbf{f})(\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$ and (2) that $(c\mathbf{f})(d\mathbf{v}) = d(c\mathbf{f})(\mathbf{v})$ for all $d \in K$ and $\mathbf{v} \in V$.

  1. We have $(c\mathbf{f})(\mathbf{v} + \mathbf{w}) = c\mathbf{f}(\mathbf{v} + \mathbf{w}) = c(\mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w})) = c\mathbf{f}(\mathbf{v}) + c\mathbf{f}(\mathbf{w}) = (c\mathbf{f})(\mathbf{v}) + (c\mathbf{f})(\mathbf{w})$.
  2. We have $(c\mathbf{f})(d\mathbf{v}) = c\mathbf{f}(d\mathbf{v}) = cd\mathbf{f}(\mathbf{v}) = c(d\mathbf{f}(\mathbf{v})) = c(d\mathbf{f})(\mathbf{v})$.

$\qquad\square$

Kernel and image of linear functions

This section focuses on two important vector subspaces that come with every linear function.

Definition 2.59. (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 $\mathbf{f}$:

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

Theorem 2.60. (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. According to Theorem 2.29, we need to show that each claimed subspace is closed under vector addition and vector scaling.

We first do so for the kernel, $\mathbf{f}^{-1}(\{\mathbf{0}\})$.

- (Closure under vector addition). Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{f}^{-1}(\{\mathbf{0}\})$. Then $\mathbf{f}(\mathbf{v}_1) = \mathbf{0}$ and $\mathbf{f}(\mathbf{v}_2) = \mathbf{0}$. We have $\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0}$, so $(\mathbf{v}_1 + \mathbf{v}_2) \in \mathbf{f}^{-1}(\{\mathbf{0}\})$.

- (Closure under vector scaling). Let $\mathbf{v} \in \mathbf{f}^{-1}(\{\mathbf{0}\})$. Then $\mathbf{f}(\mathbf{v}) = \mathbf{0}$. We have $\mathbf{f}(c\mathbf{v}) = c\mathbf{f}(\mathbf{v}) = c \cdot \mathbf{0} = \mathbf{0}$, so $c\mathbf{v} \in \mathbf{f}^{-1}(\{\mathbf{0}\})$.

Now we do so for the image, $\mathbf{f}(V)$.

- (Closure under vector addition). Let $\mathbf{w}_1, \mathbf{w}_2 \in \mathbf{f}(V)$. Then there exist $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $\mathbf{w}_1 = \mathbf{f}(\mathbf{v}_1)$ and $\mathbf{w}_2 = \mathbf{f}(\mathbf{v}_2)$. We have $\mathbf{w}_1 + \mathbf{w}_2 = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2)$, where $(\mathbf{v}_1 + \mathbf{v}_2) \in V$, so $(\mathbf{w}_2 + \mathbf{w}_2) \in \mathbf{f}(V)$.

- (Closure under vector scaling). Let $\mathbf{w} \in \mathbf{f}(V)$. Then there exists $\mathbf{v} \in V$ such that $\mathbf{f}(\mathbf{v}) = \mathbf{w}$. We have $c\mathbf{w} = c\mathbf{f}(\mathbf{v}) = \mathbf{f}(c\mathbf{v})$, where $c\mathbf{v} \in V$, so $c\mathbf{w} \in \mathbf{f}(V)$.

$\square$

Definition 2.61. (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 2.62. (One-to-one linear functions are the linear functions that 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.

($\implies$). $\mathbf{f}$ 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.

($\impliedby$). $\mathbf{f}$ 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. $\square$

Remark 2.63. 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 2.64. (One-to-one linear functions are the linear functions that 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,

$$(\mathbf{v}_1, ..., \mathbf{v}_k \text{ are linearly independent}) \implies (\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k) \text{ are linearly independent})$$

if and only if

$$\mathbf{f} \text{ is one-to-one}$$

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 "$c_1\mathbf{f}(\mathbf{v}_1) + ... + c_k\mathbf{f}(\mathbf{v}_k) = \mathbf{0}$ implies the c_i's are all 0". Thus $\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k)$ are linearly independent, as claimed.

($\Longrightarrow$). Suppose that if $\mathbf{v}_1, ..., \mathbf{v}_k$ are linearly independent, then $\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k)$ are linearly independent. We need to show $\mathbf{f}$ is one-to-one; it suffices to show that $\mathbf{f}$ has a trivial kernel. Let $\mathbf{v} \in \mathbf{f}^{-1}(\{\mathbf{0}\})$, so $\mathbf{f}(\mathbf{v}) = \mathbf{0}$. We want to show $\mathbf{v} = \mathbf{0}$.

Since V is finite-dimensional, there is a basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ for V. Expressing $\mathbf{v}$ relative to E, we have $\mathbf{f}(\mathbf{v}) = \mathbf{f}\left(\sum_{i=1}^{k}([\mathbf{v}]_E)_i\mathbf{e}_i \right) = \sum_{i=1}^{k}([\mathbf{v}]_E)_i\mathbf{f}(\mathbf{e}_i) = \mathbf{0}$. 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 $\sum_{i=1}^{k}([\mathbf{v}]_E)_i\mathbf{f}(\mathbf{e}_i) = \mathbf{0}$. Thus $([\mathbf{v}]_E)_i = 0$ for all i, i.e., $\mathbf{v} = \mathbf{0}$. $\square$

Theorem 2.65. (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 2.43, 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. $\square$

Linear isomorphisms

Definition 2.66. (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[9], 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 2.67. (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. $\qquad\square$

Theorem 2.68. (A linear function between 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 the 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 condition is equivalent to $\dim(V) > \dim(\mathbf{f}(V))$. Since we assumed $\dim(V) = \dim(W)$, this most recent condition is the same as $\dim(W) > \dim(\mathbf{f}(V))$. One can check that if Y_1 and Y_2 are subspaces of the same finite-dimensional vector space, then $\dim(Y_1) > \dim(Y_2)$ iff $Y_1 \supsetneq Y_2$. Using $Y_1 = W$ and $Y_2 = \mathbf{f}(V)$, we obtain the equivalent condition $W \supsetneq \mathbf{f}(V)$. This is equivalent to $\mathbf{f}$ not being onto. $\qquad\square$

Theorem 2.69. (Finite-dimensional vector spaces are isomorphic iff they have the same dimension).

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

Proof.

($\implies$). If $\mathbf{f} : V \to W$ is a linear isomorphism, then it has a trivial kernel. The main dimension theorem then implies that $\dim(W) = \dim(V) - \dim(\mathbf{f}^{-1}(\{\mathbf{0}\})) = \dim(V) - 0 = \dim(V)$.

($\impliedby$). 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 2.72 we will define the linear function $[\cdot]_E : V \to K^n$; Theorem 2.74 will show that $[\cdot]_E$ is a linear isomorphism. $\qquad\square$

Theorem 2.70. (Linear isomorphisms send bases to bases).

Let V and W be finite-dimensional vector spaces. If E is a basis of V and $\mathbf{f} : V \to W$ is a linear isomorphism, then $\mathbf{f}(E)$ is a basis for W.

Proof. Linear isomorphisms are one-to-one, so they preserve the linear independence of vectors. Thus $\mathbf{f}(E)$ is a linearly independent set of $\dim(V)$ many vectors. Since $\mathbf{f}$ is an isomorphism, then $\dim(V) = \dim(W)$, so $\mathbf{f}(E)$ is a linearly independent set of $\dim(W)$ many vectors. Recall from Lemma 2.42 that a linearly independent set of n vectors from an n-dimensional vector space span that space. This makes $\mathbf{f}(E)$ is a basis for W. $\qquad\square$

Definition 2.71. (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.

---

[9]Recall from Theorem 1.26 that any (not necessarily linear) function is invertible iff it is a bijection.

Coordinatization of vectors

Definition 2.72. (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.

We often use the notation $[\cdot]_E$ to denote the function $V \to K^{\dim(V)}$ sending $\mathbf{v} \mapsto [\mathbf{v}]_E$.

The above definition does not make it obvious that the function $[\cdot]_E$ is well-defined. We show this now.

Theorem 2.73. ($[\cdot]_E$ is well-defined).

Let V be a finite-dimensional vector space. The function $[\cdot]_E$ is well-defined for all bases E of V.

Proof. Consider $\mathbf{v}, \mathbf{w} \in V$ that are equal, $\mathbf{v} = \mathbf{w}$. We need to show that $[\mathbf{v}]_E = [\mathbf{w}]_E$.

Set $n := \dim(V)$. We have $\mathbf{v} = \sum_{i=1}^{n} ([\mathbf{v}]_E)_i \mathbf{e}_i$ and $\mathbf{w} = \sum_{i=1}^{n} ([\mathbf{w}]_E)_i \mathbf{e}_i$. Since $\mathbf{v} = \mathbf{w}$, we can subtract one equation from the other to obtain $\mathbf{0} = \sum_{i=1}^{n} \left(([\mathbf{v}]_E)_i - ([\mathbf{w}]_E)_i \right) \mathbf{e}_i$. Since E is linearly independent, then we must have $([\mathbf{v}]_E)_i - ([\mathbf{w}]_E)_i = 0 \iff ([\mathbf{v}]_E)_i = ([\mathbf{w}]_E)_i$ for all i. Thus $[\mathbf{v}]_E = [\mathbf{w}]_E$. $\qquad\square$

Of more practical importance than the above is the following fact.

Theorem 2.74. ($[\cdot]_E$ is a linear isomorphism).

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 $[c\mathbf{v}]_E = c[\mathbf{v}]_E$.

Let $n := \dim(V)$. We have

$$[\mathbf{v}_1 + \mathbf{v}_2]_E = \Big[\sum_{i=1}^{n} ([\mathbf{v}_1]_E)_i \mathbf{e}_i + \sum_{i=1}^{n} ([\mathbf{v}_2]_E)_i \mathbf{e}_i \Big]_E$$

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

$$= [\mathbf{v}_1]_E + [\mathbf{v}_2]_E.$$

Now we show $[c\mathbf{v}]_E = c[\mathbf{v}]_E$. We have

$$[c\mathbf{v}]_E = \Big[c \Big(\sum_{i=1}^{n} ([\mathbf{v}]_E)_i \mathbf{e}_i \Big) \Big] = \Big[\sum_{i} c([\mathbf{v}]_E)_i \mathbf{e}_i \Big]_E = \begin{pmatrix} c([\mathbf{v}]_E)_1 \\ \vdots \\ c([\mathbf{v}]_E)_n \end{pmatrix} = c \begin{pmatrix} ([\mathbf{v}]_E)_1 \\ \vdots \\ ([\mathbf{v}]_E)_n \end{pmatrix} = c[\mathbf{v}]_E.$$

Since $[\cdot]_E$ sends basis vectors to basis vectors, it preserves linear independence, and is therefore one-to-one (recall Theorem 2.64). Since $[\cdot]_E$ is a one-to-one linear function between vector spaces of the same dimension, it must also be onto. Thus $[\cdot]_E$ is a linear isomorphism.

$\qquad\square$

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 2.75. (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:

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

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 from $K^{\dim(V)}$ that is expressed relative to the basis E for V and producing an output in $K^{\dim(W)}$ 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[10] 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 2.76. (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

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

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

The above equalities can be written more compactly as

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

---

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

(Recall that the jth component of $\hat{\mathbf{\mathfrak{e}}}_i$ is 1 when $i = j$ and 0 otherwise, and that $\hat{\mathfrak{E}} = \{\hat{\mathbf{\mathfrak{e}}}_1, ..., \hat{\mathbf{\mathfrak{e}}}_n\}$ 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 $\mathbf{f}(\hat{\mathbf{\mathfrak{e}}}_1), ..., \mathbf{f}(\hat{\mathbf{\mathfrak{e}}}_n)$. That is, if we know $\mathbf{f}(\hat{\mathbf{\mathfrak{e}}}_1), ..., \mathbf{f}(\hat{\mathbf{\mathfrak{e}}}_n)$, then we can figure out what $\mathbf{f}$ is!

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

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

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

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

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[11] 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 2.77. (Matrix-vector product).

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

From the previous derivation, we know 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

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

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[12] 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:

---

[11] 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 2.55 implies that $\mathbf{F}$ is onto. Showing that $\mathbf{F}$ is one-to-one is a simple exercise.

[12] 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 2.55. 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 2.78. (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 2.76 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 2.55 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 $\mathbf{f}$. $\qquad\square$

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

Definition 2.79. (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 2.80. (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

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

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* $\{\hat{\mathfrak{e}}_1, \dots, \hat{\mathfrak{e}}_n\}$ containing $\hat{\mathfrak{e}}_1, \dots, \hat{\mathfrak{e}}_n$. (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 2.81. (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).

Composition of linear functions

In the previous section, we discovered how the action of a linear function on a vector corresponded to an operation involving the linear function's matrix and the vector. In this section, we discover an operation involving two linear functions' matrices. The operation arises from the fact that a composition of functions represented by matrices produces yet another function represented by a matrix.

Theorem 2.82. (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. We have $(\mathbf{g} \circ \mathbf{f})(\mathbf{v} + \mathbf{w}) = \mathbf{g}(\mathbf{f}(\mathbf{v} + \mathbf{w})) = \mathbf{g}(\mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w})) = \mathbf{g}(\mathbf{f}(\mathbf{v})) + \mathbf{g}(\mathbf{f}(\mathbf{w})) = (\mathbf{g} \circ \mathbf{f})(\mathbf{v}) + (\mathbf{g} \circ \mathbf{f})(\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$ and $(\mathbf{g} \circ \mathbf{f})(c\mathbf{v}) = \mathbf{g}(\mathbf{f}(c\mathbf{v})) = \mathbf{g}(c\mathbf{f}(\mathbf{v})) = c\mathbf{g}(\mathbf{f}(\mathbf{v})) = c(\mathbf{g} \circ \mathbf{f})(\mathbf{v})$ for all $\mathbf{v} \in V$ and $c \in K$. □

Now, we answer the question: "how is the standard matrix of $\mathbf{g} \circ \mathbf{f}$ related to the standard matrices of $\mathbf{g}$ and $\mathbf{f}$?".

Derivation 2.83. (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, $(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}})$:

$$
\begin{aligned}
(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}}) &= \left((\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{e}}_1) \quad \ldots \quad (\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{e}}_n) \right) \\
&= \left(\mathbf{g}(\mathbf{f}(\hat{\mathfrak{e}}_1)) \quad \ldots \quad \mathbf{g}(\mathbf{f}(\hat{\mathfrak{e}}_n)) \right) \\
&= \left(\mathbf{g}(\hat{\mathfrak{F}})\mathbf{f}(\hat{\mathfrak{e}}_1) \quad \ldots \quad \mathbf{g}(\hat{\mathfrak{F}})\mathbf{f}(\hat{\mathfrak{e}}_n) \right).
\end{aligned}
$$

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

$$
(\mathbf{g} \circ \mathbf{f})(\hat{\mathfrak{E}}) = \left(\mathbf{g}(\hat{\mathfrak{F}})(\mathbf{f}(\hat{\mathfrak{E}}))_1 \quad \ldots \quad \mathbf{g}(\hat{\mathfrak{F}})(\mathbf{f}(\hat{\mathfrak{E}}))_n \right).
$$

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:

$$
\mathbf{P}(\mathbf{g}(\hat{\mathfrak{F}}), \mathbf{f}(\hat{\mathfrak{E}})) := \left(\mathbf{g}(\hat{\mathfrak{F}})(\mathbf{f}(\hat{\mathfrak{E}}))_1 \quad \ldots \quad \mathbf{g}(\hat{\mathfrak{F}})(\mathbf{f}(\hat{\mathfrak{E}}))_n \right).
$$

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 $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{pmatrix} \in K^{m \times n}$ and $\mathbf{B} \in K^{m \times p}$:

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

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 $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{pmatrix}$ and a $p \times m$ matrix $\mathbf{B}$, we define the $p \times n$ matrix $\mathbf{BA}$ as follows:

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

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 2.84. (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 $\mathbf{A}$ and $\mathbf{B}$ 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.

Additionally, when $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $p \times m$, then $\mathbf{A}$ represents a linear function $K^n \to K^m$ and $\mathbf{B}$ represents a linear function $K^m \to K^p$, so $\mathbf{BA}$ represents a linear function $K^n \to K^p$. Thus when $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $p \times m$, we know $\mathbf{BA}$ must be $p \times n$.

Remark 2.85. Expanding out the columns of $\mathbf{BA}$, here is what the matrix-matrix product $\mathbf{BA}$ looks like when written out more explicitly:

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

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

Don't try to make too much sense of this now. Theorem 2.93 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 2.86. (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 2.87. (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.

Matrix-vector products

Now we return to the relationship between linear functions and matrices. This relationship hinges on the matrix-vector product, which we revisit. We first restate the definition of the matrix-vector product from Derivation 2.77 for convenience.

Definition 2.88. (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:

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

Now we present several new theorems.

Theorem 2.89. (*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.

By the previous definition, the matrix-vector product $\mathbf{Av}$ is equal to the following:

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

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

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

Here $\cdot : K^n \times K^n \to K$ denotes the *dot product*[13] of vectors in K^n, defined by

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

Notice how since the dot product must take two column vectors as input, the above boxed formula uses the transposed ith row of $\mathbf{A}$, which is a column vector, instead of the "unmodified" ith row of $\mathbf{A}$.

Theorem 2.90. (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. We define $\ker(\mathbf{A}) := \ker(\mathbf{v} \mapsto \mathbf{A}\mathbf{v})$ and $\text{im}(\mathbf{A}) := \text{im}(\mathbf{v} \mapsto \mathbf{A}\mathbf{v})$.

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

Theorem 2.91. (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.

Matrix-matrix products

We also revisit the matrix-matrix product for convenience. Similarly to as was with matrix-vector products, we restate the definition of matrix-matrix product from Definition **??** for convenience.

Definition 2.92. (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* $\mathbf{B}\mathbf{A} \in K^{p \times n}$ to be the following:

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

Now, we present several new concepts regarding matrix-matrix products.

Theorem 2.93. (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 $\mathbf{B}\mathbf{A}$ can be computed by using the definition of the matrix-matrix product (Theorem 2.83) together with the fact that the ith entry of the matrix-vector product $\mathbf{A}\mathbf{v}$ is $(i\text{th row of } \mathbf{A}) \cdot \mathbf{v}$, where $\cdot$ is the dot product. We have

---

[13]We discuss the dot product in-depth later in this chapter.

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

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

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

Similarly to in Theorem 2.89, we have to transpose the ith row of $\mathbf{B}$ so that it is a column vector.

Remark 2.94. (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 2.95. (Inverse matrix).

Let $\mathbf{A}$ be an $m \times n$ matrix with entries in a field K. We say that $\mathbf{A}$ is *invertible* iff there exists a matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^{-1}$.

Theorem 2.96. (Inverse matrix).

Let $\mathbf{A}$ be an invertible $m \times n$ matrix with entries in a field K, and consider the linear function $\mathbf{f} : K^m \to K^n$ defined by $\mathbf{f}(\mathbf{v}) = \mathbf{Av}$. Because $\mathbf{f}^{-1} \circ \mathbf{f} = \mathbf{I} = \mathbf{f} \circ \mathbf{f}^{-1}$, it follows that $\mathbf{A} = \mathbf{f}(\hat{\mathfrak{E}})$ is invertible iff $\mathbf{f}$ is, and that when $\mathbf{A}^{-1}$ exists, it is equal to the standard matrix of $\mathbf{f}^{-1}$. In other words, if $\mathbf{f} : K^n \to K^n$ is an invertible linear function, then $\mathbf{f}(\hat{\mathfrak{E}})^{-1} := \mathbf{f}^{-1}(\hat{\mathfrak{E}})$.

Since it is only possible for $\mathbf{f}$ to be invertible when $n = m$, it follows that $\mathbf{A}$ must be a $n \times n$ matrix, or *square matrix*, in order to be invertible.

Remark 2.97. (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 natural investigations (such as Derivation 2.76). (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!

Matrix transposes

The topic of matrix tranposes does not become theoretically significant until Chapter 4. We present some surface-level details now, though, because they do appear occasionally before Chapter 4.

Definition 2.98. (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 $n \times m$ matrix whose ij entry is the ji entry of $\mathbf{A}$: $\mathbf{A}^\top := (a_{ji})$.

We also use the notation $a_{ij}^\top$ to denote the ij entry of $(a_{ij})^\top$.

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 $\mathbf{A}$.

Remark 2.99. (Concise column vectors).

Since

$$(v_1, ..., v_n)^\top = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \text{ for all } v_1, ..., v_n \in K^n$$

we often find it convenient to use $(v_1, ..., v_n)^\top$ to save vertical writing space.

Theorem 2.100. (Properties of the transpose).

1. $\mathbf{I}^\top = \mathbf{I}$.

2. $(\mathbf{A}^\top)^\top = \mathbf{A}$ for all matrices $\mathbf{A}$.

3. $\mathbf{A} \mapsto \mathbf{A}^\top$ is linear for all matrices $\mathbf{A}$.

4. $(\mathbf{BA})^\top = \mathbf{A}^\top \mathbf{B}^\top$ for all matrices $\mathbf{A}, \mathbf{B}$ compatible with each other for matrix multiplication.

5. $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$ for all invertible matrices $\mathbf{A}$.

Proof.

1. This is true because $\delta_{ij} = \delta_{ji}$.

2. $((a_{ij})^\top)^\top = (a_{ij}^\top)^\top = (a_{ji})^\top = (a_{ji}^\top) = (a_{ij})$.

3. We must show that $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$ for all matrices $\mathbf{A}$ and $\mathbf{B}$ and that $(c\mathbf{A})^\top = c\mathbf{A}^\top$ for all matrices $\mathbf{A}$ and scalars c.

   We have $((a_{ij}) + (b_{ij}))^\top = (a_{ij} + b_{ij})^\top = (a_{ji} + b_{ji}) = (a_{ji}) + (b_{ji}) = (a_{ij})^\top + (b_{ij})^\top$, so the first property holds.

   We also have $(c(a_{ij}))^\top = (ca_{ij})^\top = (ca_{ji}) = c(a_{ji}) = c(a_{ij})^\top$, so the second property holds as well.

4. $((b_{ij})(a_{ij}))^\top = (\sum_k b_{ik} a_{kj})^\top = (\sum_k b_{jk} a_{ki}) = (\sum_k a_{ik}^\top b_{kj}^\top) = (a_{ij}^\top)(b_{ij}^\top) = (a_{ij})^\top (b_{ij})^\top$.

5. We need to show that $\mathbf{A}^\top (\mathbf{A}^{-1})^\top = \mathbf{I}$. By property (2), this equation is equivalent to $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, which is true.

$\square$

Definition 2.101. (Inverse transpose).

Let $\mathbf{A}$ be a matrix. Because the matrix inversion and transposition operations commute, it is unambiguous to define $\mathbf{A}^{-\top} := (\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$.

Matrices relative to bases

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

Lemma 2.102. $([\cdot]_E^{-1}(\hat{\mathfrak{E}}) = E)$.
    Let V be a vector space and let E be a basis of V. We have $[\cdot]_E^{-1}(\hat{\mathfrak{E}}) = E$.

Proof. Since $[\mathbf{e}_i]_E = \hat{\mathfrak{e}}_i$, then $[\cdot]_E^{-1}(\hat{\mathfrak{e}}_i) = \mathbf{e}_i$, and so $[\cdot]_E^{-1}(\hat{\mathfrak{E}}) = E$. $\qquad\square$

Notice that E is not a matrix in the above lemma! It is a list of vectors that are not necessarily column vectors. We investigate the special case when E is indeed a list of column vectors (i.e. a list of vectors from K^n) in Theorem 2.107.

Derivation 2.103. (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[14] 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 $\mathbf{f}_{E,F}(\hat{\mathfrak{E}}) = ([\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1})(\hat{\mathfrak{E}}) = \left[\mathbf{f}\left([\cdot]_E^{-1}(\hat{\mathfrak{E}})\right)\right]_F$. By the previous lemma, we have $[\cdot]_E^{-1}(\hat{\mathfrak{E}}) = E$, 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:

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

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

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

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 2.104. (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 $\hat{\mathfrak{E}}$.

---

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

Vectors relative to bases

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 2.105. (Change of basis for vectors).

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

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 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 2.106. (Alternate proof of change of basis for vectors).

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

The following is an important special case[15] of the change of basis theorem for vectors.

Theorem 2.107. (Change of basis from standard basis for vectors).

Let V be a vector space and let E be a basis of V. Then from the definition of $[\cdot]_E$ (recall Definition 2.72) it immediately follows that

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

When $V = K^n$, then the list E becomes a matrix of column vectors $\mathbf{E}$, and the standard matrices of $[\cdot]_E$ and $[\cdot]_E^{-1}$ exist. Using the result of Lemma 2.102, we see they are:

$$[\cdot]_E^{-1}(\hat{\mathfrak{C}}) = \mathbf{E} \text{ and } [\cdot]_E(\hat{\mathfrak{C}}) = \mathbf{E}^{-1}$$

Importantly, we see that $\mathbf{E}^{-1}$ exists. We can now use $\mathbf{E}^{-1}$ to solve the equation $\mathbf{v} = \mathbf{E}[\mathbf{v}]_E$ for $[\mathbf{v}]_E$:

$$\boxed{[\mathbf{v}]_E = \mathbf{E}^{-1}\mathbf{v} \text{ for all } \mathbf{v} \in K^n}$$

Theorem 2.108. ($\mathbf{I}_{E,F}^{-1} = \mathbf{I}_{F,E}$ when $V = K^n$).

Let V be a finite-dimensional vector space with bases E and F. The identity function $\mathbf{I} : V \to V$ on V satisfies $\mathbf{I}_{E,F}^{-1} = \mathbf{I}_{F,E}$. As a corollary, when $V = K^n$, then E and F are both lists of column vectors, so 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 $\mathbf{I}_{E,F} = [\cdot]_F \circ [\cdot]_E^{-1}$ and $\mathbf{I}_{F,E} = [\cdot]_E \circ [\cdot]_F^{-1}$. We clearly have $\mathbf{I}_{E,F}^{-1} = \mathbf{I}_{F,E}$. $\square$

Theorem 2.109. (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 2.110. (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 2.105. This approach for proving Theorem 2.105 was not used because it involves quite a bit of unmemorable matrix algebra. It's good to know that these theorems are equivalent, though.

---

[15]Use the substitutions $E := \hat{\mathfrak{C}}$ and $F := E$ in the change of basis theorem for vectors to obtain $[\mathbf{v}]_E = [\hat{\mathfrak{C}}]_E[\mathbf{v}]_{\hat{\mathfrak{C}}} = \mathbf{E}^{-1}\mathbf{v}$.

Matrix similarity

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 2.111. (Matrices of linear functions $K^n \to K^n$ relative to bases).

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

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

An important special case of this theorem is that the matrix of $\mathbf{f}$ relative to E and E is $\mathbf{E}^{-1}\mathbf{f}(\hat{\mathfrak{E}})\mathbf{E}$.

Proof. The matrix $[\mathbf{f}(E)]_F$ of $\mathbf{f}$ relative to E and F is the standard matrix $\mathbf{f}_{E,F}(\hat{\mathfrak{E}})$ of $\mathbf{f}_{E,F} = [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}$, which is $\mathbf{f}_{E,F}(\hat{\mathfrak{E}}) = [\cdot]_F(\hat{\mathfrak{E}}) \, \mathbf{f}(\hat{\mathfrak{E}}) \, [\cdot]_E^{-1}(\hat{\mathfrak{E}})$. Since F is a list of column vectors, then the corollary of Theorem 2.108 applies, and $[\cdot]_F(\hat{\mathfrak{E}}) = [\hat{\mathfrak{E}}]_F$ is equal to $[\mathbf{F}]_{\hat{\mathfrak{E}}}^{-1} = \mathbf{F}^{-1}$. Also, because E is a list of column vectors, we can apply Theorem 2.107, which says $[\cdot]_E^{-1}(\hat{\mathfrak{E}}) = \mathbf{E}$. So the most recent expression is equal to $\mathbf{F}^{-1}\mathbf{f}(\hat{\mathfrak{E}})\mathbf{E}$, as claimed. $\qquad\square$

Remark 2.112. (Matrix similarity).

Most linear algebra texts do not define the matrix of a linear function $V \to W$ relative to bases E and F, and only define the "matrix of a linear function $\mathbf{f} : K^n \to K^n$ relative to the basis E of K^n" to be $\mathbf{E}^{-1}\mathbf{f}(\hat{\mathfrak{E}})\mathbf{E}$.

In such texts, one of two definitions of *matrix similarity* is given:

1. Matrices $\mathbf{A}$ and $\mathbf{B}$ are *similar* iff there exists a matrix $\mathbf{E}$ such that $\mathbf{B} = \mathbf{E}^{-1}\mathbf{A}\mathbf{E}$.

2. Matrices $\mathbf{A}$ and $\mathbf{B}$ are *similar* iff there exists a matrix $\mathbf{E}$ such that $\mathbf{B} = \mathbf{E}\mathbf{A}\mathbf{E}^{-1}$.

The definitions are equivalent because matrices of the form $\mathbf{E}^{-1}\mathbf{A}\mathbf{E}$ are also of the form $\mathbf{E}\mathbf{A}\mathbf{E}^{-1}$.

It is much better to use the first definition than the second one, though, since the placement of the matrix inverse in the second definition is different than it is in the formula $[\mathbf{f}(E)]_E = \mathbf{E}^{-1}\mathbf{f}(\hat{\mathfrak{E}})\mathbf{E}$. Using the second definition could lead one to think that "$[\mathbf{f}(E)]_E = \mathbf{E}\mathbf{f}(E)\mathbf{E}^{-1}$", which is not true!

Linear isomorphism between linear functions and matrices

We've alluded to the correspondence between linear functions and matrices several times. Now that we know of matrices relative to bases, we can formalize our knowledge of this relationship.

Theorem 2.113. (Linear isomorphism between linear functions and matrices).

Let V, W, and Y be finite-dimensional vector spaces over a field K, with bases E, F, and G.

The function $\mathbf{F} : \mathcal{L}(V \to Y) \to K^{\dim(Y) \times \dim(V)}$ defined by $\mathbf{F}(\mathbf{f}) := [\mathbf{f}(E)]_G$ is a linear isomorphism, and also satisfies $\mathbf{F}(\mathbf{g} \circ \mathbf{f}) = [\mathbf{g}(F)]_G[\mathbf{f}(E)]_F$ for all linear functions $\mathbf{f} : V \to W$ and $\mathbf{g} : W \to Y$.

The inverse $\mathbf{F}^{-1} : K^{\dim(Y) \times \dim(V)} \to \mathcal{L}(V \to Y)$ is defined by $\mathbf{F}^{-1}(\mathbf{A}) = [\cdot]_G^{-1} \circ (\mathbf{v} \mapsto \mathbf{A}\mathbf{v}) \circ [\cdot]_E$, and satisfies $\mathbf{F}^{-1}(\mathbf{BA}) = \mathbf{g} \circ \mathbf{f}$, where $\mathbf{f} : V \to W$ and $\mathbf{g} : W \to Y$ are linear functions with $[\mathbf{f}(E)]_F = \mathbf{A}$ and $[\mathbf{g}(F)]_G = \mathbf{B}$.

Proof. Notice that $\mathbf{F} = \mathbf{Q} \circ \mathbf{P}$, where $\mathbf{P} : \mathcal{L}(V \to Y) \to \mathcal{L}(K^{\dim(V)} \to K^{\dim(Y)})$ is defined by $\mathbf{P}(\mathbf{f}) := \mathbf{f}_{E,G}$ and where $\mathbf{Q} : \mathcal{L}(K^{\dim(V)} \to K^{\dim(Y)}) \to K^{\dim(Y) \times \dim(V)}$ is defined by $\mathbf{Q}(\mathbf{g}) := \mathbf{g}(\hat{\mathfrak{E}})$. (We have $(\mathbf{Q} \circ \mathbf{P})(\mathbf{f}) = \mathbf{Q}(\mathbf{P}(\mathbf{f})) = \mathbf{Q}(\mathbf{f}_{E,G}) = \mathbf{f}_{E,G}(\hat{\mathfrak{E}}) = [\mathbf{f}(E)]_G$).

- ($\mathbf{F}$ is linear). It's easy to check that $\mathbf{f} \mapsto \mathbf{f}(E)$ is a linear function. Once we know this, then since $\mathbf{f} \mapsto \mathbf{f}(E)$ and $[\cdot]_G$ are both linear functions, we see $\mathbf{f} \mapsto \mathbf{F}(\mathbf{f}) = [\mathbf{f}(E)]_G$ is a composition of linear functions, and is thus linear.

- ($\mathbf{F}^{-1} = [\cdot]_G^{-1} \circ (\mathbf{v} \mapsto \mathbf{A}\mathbf{v}) \circ [\cdot]_E$). To prove this, we first find $\mathbf{Q}^{-1}$ and $\mathbf{P}^{-1}$ and then use the fact that $\mathbf{F}^{-1} = \mathbf{P}^{-1} \circ \mathbf{Q}^{-1}$.

  Since $\mathbf{Q}(\mathbf{g})$ is the standard matrix of $\mathbf{g}$, we claim that $\mathbf{Q}^{-1}(\mathbf{A})$ is the linear function obtained from the matrix $\mathbf{A}$: we claim $\mathbf{Q}^{-1}(\mathbf{A}) = (\mathbf{v} \mapsto \mathbf{A}\mathbf{v})$. We have $\mathbf{Q}^{-1}(\mathbf{Q}(\mathbf{g})) = \mathbf{Q}^{-1}(\mathbf{g}(\hat{\mathfrak{E}})) = (\mathbf{v} \mapsto \mathbf{g}(\hat{\mathfrak{E}})\mathbf{v}) = \mathbf{g}$ and $\mathbf{Q}(\mathbf{Q}^{-1}(\mathbf{A})) = \mathbf{Q}(\mathbf{v} \mapsto \mathbf{A}\mathbf{v}) = (\mathbf{v} \mapsto \mathbf{A}\mathbf{v})(\hat{\mathfrak{E}}) = \mathbf{A}$, so $\mathbf{Q}^{-1}$ is indeed defined by $\mathbf{Q}^{-1}(\mathbf{A}) = (\mathbf{v} \mapsto \mathbf{A}\mathbf{v})$.

  Since we have $\mathbf{P}(\mathbf{f}) = \mathbf{f}_{E,G} = [\cdot]_G \circ \mathbf{f} \circ [\cdot]_E^{-1}$, then $\mathbf{P}^{-1}(\mathbf{g}) = [\cdot]_G^{-1} \circ \mathbf{g} \circ [\cdot]_E$.

  Thus $\mathbf{F}^{-1} = (\mathbf{P}^{-1} \circ \mathbf{Q}^{-1})(\mathbf{A}) = [\cdot]_G^{-1} \circ (\mathbf{v} \mapsto \mathbf{A}\mathbf{v}) \circ [\cdot]_E$.

- ($\mathbf{F}(\mathbf{g} \circ \mathbf{f}) = [\mathbf{g}(F)]_G[\mathbf{f}(E)]_F$). We have $\mathbf{F}(\mathbf{g} \circ \mathbf{f}) = (\mathbf{Q} \circ \mathbf{P})(\mathbf{g} \circ \mathbf{f}) = \mathbf{Q}(\mathbf{P}(\mathbf{g} \circ \mathbf{f})) = \mathbf{Q}((\mathbf{g} \circ \mathbf{f})_{E,G})$. Notice that $(\mathbf{g} \circ \mathbf{f})_{E,G} = [\cdot]_G \circ \mathbf{g} \circ \mathbf{f} \circ [\cdot]_E^{-1} = ([\cdot]_G \circ \mathbf{g} \circ [\cdot]_F^{-1}) \circ ([\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}) = \mathbf{g}_{F,G} \circ \mathbf{f}_{E,F}$. Thus $\mathbf{Q}((\mathbf{g} \circ \mathbf{f})_{E,G}) = \mathbf{Q}(\mathbf{g}_{F,G} \circ \mathbf{f}_{E,F})$. Since $\mathbf{f}_{E,F} : K^{\dim(V)} \to K^{\dim(W)}$ and $\mathbf{g} : K^{\dim(W)} \to K^{\dim(Y)}$, we can apply the result from the end of Derivation 2.83 to conclude that the standard matrix of the composition is equal to the matrix-matrix product of the corresponding standard matrices; that is, $\mathbf{Q}(\mathbf{g}_{F,G} \circ \mathbf{f}_{E,F}) = \mathbf{g}_{F,G}(\hat{\mathfrak{E}})\mathbf{f}_{E,F}(\hat{\mathfrak{E}}) = [\mathbf{g}(F)]_G[\mathbf{f}(E)]_F$.

- ($\mathbf{F}^{-1}(\mathbf{BA}) = \mathbf{g} \circ \mathbf{f}$, where $[\mathbf{f}(E)]_F = \mathbf{A}$ and $[\mathbf{g}(F)]_G = \mathbf{B}$). We start with the proven fact $\mathbf{F}(\mathbf{g} \circ \mathbf{f}) = [\mathbf{g}(F)]_G[\mathbf{f}(E)]_F$. Applying $\mathbf{F}^{-1}$ to both sides, we obtain $\mathbf{g} \circ \mathbf{f} = [\cdot]_G^{-1} \circ (\mathbf{v} \mapsto [\mathbf{g}(F)]_G[\mathbf{f}(E)]_F\mathbf{v}) \circ [\cdot]_E$. Consider the matrix-matrix product that appears in the expression on the right side, $\mathbf{g}_{F,G}(\hat{\mathfrak{E}})\mathbf{f}_{E,F}(\hat{\mathfrak{E}})$. Again applying the result from the end of Derivation 2.83, we see that the matrix-matrix product of standard matrices $\mathbf{g}_{F,G}(\hat{\mathfrak{E}})\mathbf{f}_{E,F}(\hat{\mathfrak{E}})$ is equal to the standard matrix of the composition $\mathbf{g}_{F,G} \circ \mathbf{f}_{E,F}$; that is, $\mathbf{g}_{F,G}(\hat{\mathfrak{E}})\mathbf{f}_{E,F}(\hat{\mathfrak{E}}) = (\mathbf{g}_{F,G} \circ \mathbf{f}_{E,F})(\hat{\mathfrak{E}})$. Furthermore, we have $\mathbf{v} \mapsto (\mathbf{g}_{F,G} \circ \mathbf{f}_{E,F})(\hat{\mathfrak{E}})\mathbf{v} = \mathbf{g}_{F,G} \circ \mathbf{f}_{E,F}$. Thus $\mathbf{g} \circ \mathbf{f} = [\cdot]_G^{-1} \circ (\mathbf{g}_{F,G} \circ \mathbf{f}_{E,F}) \circ [\cdot]_E = [\cdot]_G^{-1} \circ ([\cdot]_G \circ \mathbf{g} \circ [\cdot]_F^{-1}) \circ ([\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}) \circ [\cdot]_E = \mathbf{g} \circ \mathbf{f}$.

  In all, we've shown that $\mathbf{F}^{-1}([\mathbf{g}(F)]_G[\mathbf{f}(E)]_F) = \mathbf{g} \circ \mathbf{f}$ for all linear functions $\mathbf{f} : V \to W$ and $\mathbf{g} : W \to Y$. Recall from Theorem 2.55 that for any set of vectors, there exists a linear function sending basis vectors to those vectors. It follows that for any matrices $\mathbf{A}$ and $\mathbf{B}$ there exist linear functions $\mathbf{f} : V \to W$ and $\mathbf{g} : W \to Y$ such that $\mathbf{A} = [\mathbf{f}(E)]_F$ and $\mathbf{B} = [\mathbf{g}(F)]_G$. We can apply the equation $\mathbf{F}^{-1}([\mathbf{g}(F)]_G[\mathbf{f}(E)]_F) = \mathbf{g} \circ \mathbf{f}$ to these $\mathbf{f}$ and $\mathbf{g}$ to obtain that $\mathbf{F}^{-1}(\mathbf{BA}) = \mathbf{g} \circ \mathbf{f}$, where $[\mathbf{f}(E)]_F = \mathbf{A}$ and $[\mathbf{g}(F)]_G = \mathbf{B}$.

□

The following special case of the above theorem follows immediately.

Theorem 2.114. (Special case of linear isomorphism between linear functions and matrices).

Let K be a field, and let $\hat{\mathfrak{E}}$ be a basis for K^n and $\hat{\mathfrak{F}}$ be a basis for K^m.

The function $\mathbf{F} : \mathcal{L}(K^n \to K^p) \to K^{p \times n}$ defined by $\mathbf{F}(\mathbf{f}) := \mathbf{f}(\hat{\mathfrak{E}})$ is a linear isomorphism, and also satisfies $\mathbf{F}(\mathbf{g} \circ \mathbf{f}) = \mathbf{g}(\hat{\mathfrak{F}})\mathbf{f}(\hat{\mathfrak{E}})$ for all linear functions $\mathbf{f} : K^n \to K^m$ and $\mathbf{g} : K^m \to K^p$.

The inverse $\mathbf{F}^{-1} : K^{p \times n} \to \mathcal{L}(K^n \to K^p)$ is defined by $\mathbf{F}^{-1}(\mathbf{A}) = (\mathbf{v} \mapsto \mathbf{A}\mathbf{v})$, and satisfies $\mathbf{F}^{-1}(\mathbf{BA}) = \mathbf{g} \circ \mathbf{f}$, where $\mathbf{f}(\hat{\mathfrak{E}}) = \mathbf{A}$ and $\mathbf{g}(\hat{\mathfrak{E}}) = \mathbf{B}$.

2.4 Length and angle

In this section, we develop the notions of the *length of a vector* and the *angle between two vectors* for vectors from a finite-dimensional vector space over $\mathbb{R}$.

We approach this by first defining notions of length and angle in $\mathbb{R}^2$, generalizing these notions to $\mathbb{R}^n$, and then generalizing again to finite-dimensional vector spaces over $\mathbb{R}$.

It is easy to understand length in $\mathbb{R}^2$, and easy to generalize length in $\mathbb{R}^2$ to length in $\mathbb{R}^n$. All of the nuance in this section arises in the analysis of angle in $\mathbb{R}^2$. Once angle in $\mathbb{R}^2$ is understood, it too is easily generalized to $\mathbb{R}^n$.

We briefly outline the analysis of angle in $\mathbb{R}^2$.

Immediately after defining the angle between vectors in $\mathbb{R}^2$, we stumble across a fundamental concept that underlies both length and angle- the *dot product* of vectors in $\mathbb{R}^2$. The dot product is a subtle concept that requires careful explanation. Unfortunately, all texts I have seen have some sort of major error in their presentation of this important concept.

Most authors define the dot product using the "algebraic characterization" $\mathbf{v} \cdot \mathbf{w} := \sum_{i=1}^{2}([\mathbf{v}]_{\hat{\mathfrak{e}}})_i([\mathbf{w}]_{\hat{\mathfrak{e}}})_i$, and then use the law of cosines to show that this definition implies the "geometric characterization", $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||\,||\mathbf{w}||\cos(\theta)$, where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$. This approach is incorrect because using the law of cosines gives no intuition[16], and because there is typically no motivation given for the algebraic characterization; it just "pops out of nowhere".

Some authors try to provide more motivation for the dot product by starting with the geometric characterization, $\mathbf{v} \cdot \mathbf{w} := ||\mathbf{v}||\,||\mathbf{w}||\cos(\theta)$, as the definition, explaining that this characterization is a natural idea[17], and then showing that the algebraic characterization follows. This is an improvement over starting abruptly with the algebraic characterization, to be sure, but unfortunately, I have always seen that authors use the law of cosines to show the algebraic characterization follows from the geometric characterization. As mentioned before, the law of cosines gives no intuition.

Both of the typical approaches to (algebraic characterization) $\implies$ (geometric characterization) and (geometric characterization) $\implies$ (algebraic characterization) share a major failing not yet mentioned: they don't engage with a satisfactory definition of angle. Either angle is implicitly- because of the use of the law of cosines- taken to be as given in the cumbersome and imprecise definition from Euclidean geometry, or, when the algebraic characterization of the dot product is used as the definition, a rigorous but unmotivated definition[18] of angle is given, and the equivalence of this definition to the (cumbersome and imprecise) one relied upon by the law of cosines is not explained.

All of this mess is alleviated in this section. We start by investigating a rigorous definition of angle between vectors in $\mathbb{R}^2$. Our investigations of angle cause us to stumble upon the algebraic characterization of the dot product. The geometric characterization immediately follows, though it follows logically- not quite intuitively. To provide the missing intuition, we present proofs of (geometric characterization) $\implies$ (algebraic characterization) and (algebraic characterization) $\implies$ (geometric characterization) that do not rely on the law of cosines. Lastly, we are able to use our knowledge of the dot product to provide a useful formula for computing vector projections, and, most satisfyingly, to prove the law of cosines.

The *cross product* also comes with two common pedagogical problems. The last section of Chapter **??** describes these problems and presents a satisfying explanation of the cross product. (The cross product is addressed in Chapter **??** because understanding cross products requires understanding determinants).

---

[16]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 less intuitive than geometry done via linear algebra (to me, at least).

[17]The motivation goes like this. If a box is pushed with a force $\mathbf{F} \in \mathbb{R}^2$ along a floor through a displacement $\mathbf{x}$, then $||\mathbf{F}||\cos(\theta)$ is the amount of force exerted in the direction of the displacement $\mathbf{x}$, and thus $||\mathbf{x}||(||\mathbf{F}||\cos(\theta)) = ||\mathbf{F}||\,||\mathbf{x}||\cos(\theta)$ is the product of the amount of force aligned with the displacement with the magnitude of the displacement.

[18]The rigorous but unmotivated definition is $\theta(\mathbf{v}, \mathbf{w}) := \arccos(\hat{\mathbf{v}} \cdot \hat{\mathbf{w}})$, where the notation $\hat{\mathbf{v}}$ and $\hat{\mathbf{w}}$ is defined in Definition 2.118.

Length in $\mathbb{R}^n$

Definition 2.115. (Length of a vector in $\mathbb{R}^2$).

Let $\mathbf{v} \in \mathbb{R}^2$. If we interpret $([\mathbf{v}]_{\hat{\mathfrak{e}}})_1$ and $([\mathbf{v}]_{\hat{\mathfrak{e}}})_2$ as the lengths of the base and height, respectively, of a right triangle, then the length of the hypotenuse of this right triangle is $\sqrt{([\mathbf{v}]_{\hat{\mathfrak{e}}})_1^2 + ([\mathbf{v}]_{\hat{\mathfrak{e}}})_2^2}$. Since $\mathbf{v} = ([\mathbf{v}]_{\hat{\mathfrak{e}}})_1 \hat{\mathfrak{e}}_1 + ([\mathbf{v}]_{\hat{\mathfrak{e}}})_2 \hat{\mathfrak{e}}_2$, it makes sense to think of the hypotenuse as being $\mathbf{v}$, and thus, to define the *length* (or *magnitude*) $||\mathbf{v}||$ of $\mathbf{v}$ to be the length of the hypotenuse: $||\mathbf{v}|| := \sqrt{([\mathbf{v}]_{\hat{\mathfrak{e}}})_1^2 + ([\mathbf{v}]_{\hat{\mathfrak{e}}})_2^2}$.

Definition 2.116. (Length of a vector in $\mathbb{R}^n$).

When $n > 2$ and we have a vector $\mathbf{v}_n \in \mathbb{R}^n$, we consider $\mathbf{v}_n$ in the form $\mathbf{v}_n = \mathbf{v}_{n-1} + ([\mathbf{v}]_{\hat{\mathfrak{e}}})_n \hat{\mathfrak{e}}_n$, where $\mathbf{v}_{n-1} \in \mathbb{R}^{n-1}$, and define the *length* (or *magnitude*) $||\mathbf{v}_n||$ of $\mathbf{v}_n$ to be

$$||\mathbf{v}_n|| = \sqrt{||\mathbf{v}_{n-1}||^2 + ([\mathbf{v}]_{\hat{\mathfrak{e}}})_n^2} \text{ when } n > 2.$$

Theorem 2.117. (Length of a vector in $\mathbb{R}^n$).

For all $\mathbf{v} \in \mathbb{R}^n$ and all n, the length of $\mathbf{v}$ is

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

Proof. We prove the theorem by induction.

(Base case). The definition of the length of a vector in $\mathbb{R}^2$ establishes the base case.

(Inductive case). Let $\mathbf{v}_n \in \mathbb{R}^n$. Assume as the inductive hypothesis that the claim is true for n; we need to show the claim is true for $n + 1$.

By the definition the length of a vector in $\mathbb{R}^{n+1}$ (Definition **??**), for $\mathbf{v}_{n+1} \in \mathbb{R}^{n+1}$ we have $||\mathbf{v}_{n+1}|| = \sqrt{||\mathbf{v}_n||^2 + ([\mathbf{v}]_{\hat{\mathfrak{e}}})_{n+1}^2}$. By the inductive hypothesis, $||\mathbf{v}_n|| = \sqrt{\sum_{i=1}^{n} ([\mathbf{v}]_{\hat{\mathfrak{e}}})_i^2}$, so

$$||\mathbf{v}_{n+1}|| = \sqrt{\sqrt{\sum_{i=1}^{n} \left(([\mathbf{v}]_{\hat{\mathfrak{e}}})_i^2\right)}^2 + ([\mathbf{v}]_{\hat{\mathfrak{e}}})_{n+1}^2} = \sqrt{\sum_{i=1}^{n} \left(([\mathbf{v}]_{\hat{\mathfrak{e}}})_i^2\right) + ([\mathbf{v}]_{\hat{\mathfrak{e}}})_{n+1}^2} = \sqrt{\sum_{i=1}^{n+1} ([\mathbf{v}]_{\hat{\mathfrak{e}}})_i^2}.$$

This proves the claim for $n + 1$. $\square$

Definition 2.118. (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 $\mathbf{v} \in \mathbb{R}^n$.

Angle in $\mathbb{R}^2$

Definition 2.119. (Angle between vectors in $\mathbb{R}^2$).

Let vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ have the same length $r = ||\mathbf{v}|| = ||\mathbf{w}||$, suppose the initial points of $\mathbf{v}$ and $\mathbf{w}$ coincide, and consider the circle of radius r that has the coinciding initial points of $\mathbf{v}$ and $\mathbf{w}$ as its center. Let $s(\mathbf{x}, \mathbf{v}, \mathbf{w})$ be the length traversed on the path $\mathbf{x} : \mathbb{R} \to \mathbb{R}^2$ from $\mathbf{v}$ to $\mathbf{w}$:

$$s(\mathbf{x}, \mathbf{v}, \mathbf{w}) := \int_{t_1}^{t_2} \left|\left| \frac{d\mathbf{x}(t)}{dt} \right|\right| dt, \text{where } \mathbf{x}(t_1) = \mathbf{v}, \ \mathbf{x}(t_2) = \mathbf{w}, t_1 \leq t_2.$$

Now let $\mathbf{c} : \mathbb{R} \to \mathbb{R}^2$ be the traditional parameterization of the circle of radius r with center $\mathbf{0}$, and $\widetilde{\mathbf{c}}$ be the reverse parameterization:

$$\mathbf{c}(t) := r \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \text{ and } \widetilde{\mathbf{c}}(t) := \mathbf{c}(-t).$$

find and replace "unsigned". $t \in [0, 2\pi)$ **should probably be** $t \in [0, \pi]$

We define the *angle* $\theta(\mathbf{v}, \mathbf{w})$ *between* $\mathbf{v}$ *and* $\mathbf{w}$ to be the ratio of the shortest path between $\mathbf{v}$ and $\mathbf{w}$ on the circle to the radius of the circle:

$$\theta(\mathbf{v}, \mathbf{w}) := \frac{\min(s(\mathbf{c}, \mathbf{v}, \mathbf{w}), s(\widetilde{\mathbf{c}}, \mathbf{v}, \mathbf{w}))}{r}.$$

When $\mathbf{v}$ and $\mathbf{w}$ don't have the same length, we define $\theta(\mathbf{v}, \mathbf{w}) := \theta(r_0 \hat{\mathbf{v}}, r_0 \hat{\mathbf{w}})$, where $r_0 > 0$ is any positive real number.

Remark 2.120. (Well-definedness of angle between two vectors).

Since $\theta(\mathbf{v}, \mathbf{w})$ depends on the radius r of the circle whose center coincides with the initial points of $\mathbf{v}$ and $\mathbf{w}$, it isn't yet clear that $\theta : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is well-defined.

Theorem 2.121. (Properties of θ).

1. $\theta(\mathbf{v}, \mathbf{w}) \geq 0$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$.

2. $\theta(\mathbf{v}, \mathbf{w}) = \theta(\mathbf{w}, \mathbf{v})$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$.

Proof.

1. We have $s(\mathbf{x}, \mathbf{v}, \mathbf{w}) \geq 0$ for all $\mathbf{x} \in \{\mathbf{c}, \widetilde{\mathbf{c}}\}$ because the integrand in $s(\mathbf{x}, \mathbf{v}, \mathbf{w})$ is nonnegative for all $\mathbf{x} \in \{\mathbf{c}, \widetilde{\mathbf{c}}\}$ and $t_1 \leq t_2$.

   Since $\theta(\mathbf{v}, \mathbf{w}) = \frac{s(\mathbf{y}, \mathbf{v}, \mathbf{w})}{r}$ for some $\mathbf{y} \in \{\mathbf{c}, \widetilde{\mathbf{c}}\}$, then because $s(\mathbf{y}, \mathbf{v}, \mathbf{w}) \geq 0$ and $r > 0$, we have $\theta(\mathbf{v}, \mathbf{w}) = \frac{s(\mathbf{y}, \mathbf{v}, \mathbf{w})}{r} \geq 0$.

2. We have $s(\mathbf{x}, \mathbf{v}, \mathbf{w}) = s(\mathbf{x}, \mathbf{w}, \mathbf{v})$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ and $\mathbf{x} \in \{\mathbf{c}, \widetilde{\mathbf{c}}\}$ because the length of a path is equal to the length of the reverse path.

   Since $\theta(\mathbf{v}, \mathbf{w}) = \frac{s(\mathbf{y}, \mathbf{v}, \mathbf{w})}{r}$ for some $\mathbf{y} \in \{\mathbf{c}, \widetilde{\mathbf{c}}\}$, then $\theta(\mathbf{y}, \mathbf{v}, \mathbf{w}) = \frac{s(\mathbf{y}, \mathbf{v}, \mathbf{w})}{r} = \frac{s(\mathbf{y}, \mathbf{w}, \mathbf{v})}{r} = \theta(\mathbf{y}, \mathbf{w}, \mathbf{v})$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$.

$\square$

Definition 2.122. (Perpendicular vectors in $\mathbb{R}^2$). We say that $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ are *perpendicular*, or *orthogonal*, iff $\theta(\mathbf{v}, \mathbf{w}) = \frac{\pi}{2}$.

Theorem 2.123. (Angle formula in $\mathbb{R}^2$).

$$\theta(\mathbf{v}, \mathbf{w}) = \arccos\Big(\sum_{i=1}^{2} ([\hat{\mathbf{v}}]_{\hat{\mathbf{e}}})_i ([\hat{\mathbf{w}}]_{\hat{\mathbf{e}}})_i \Big) \text{ for all } \mathbf{v}, \mathbf{w} \in \mathbb{R}^2.$$

Note, since $\arccos : [-1, 1] \to [0, \pi]$, we have $\theta(\mathbf{v}, \mathbf{w}) \in [0, \pi]$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$.

Proof. We prove the theorem in the case when $||\mathbf{v}|| = ||\mathbf{w}|| = r$; the general case follows from this immediately. For some $\mathbf{x} \in \{\mathbf{c}, \widetilde{\mathbf{c}}\}$, we have

$$\theta(\mathbf{v}, \mathbf{w}) = \frac{s(\mathbf{x}, \mathbf{v}, \mathbf{w})}{r} = \frac{1}{r} \int_{t_1}^{t_2} \Big|\Big| \frac{d\mathbf{x}(t)}{dt} \Big|\Big| dt, \text{ where } \mathbf{x}(t_1) = \mathbf{v}, \mathbf{x}(t_2) = \mathbf{w}, t_1 \leq t_2.$$

We have $||\frac{d\mathbf{x}(t)}{dt}|| = r$ for all $\mathbf{x} \in \{\mathbf{c}, \widetilde{\mathbf{c}}\}$, so $\theta(\mathbf{v}, \mathbf{w}) = t_2 - t_1$. Since, $\mathbf{x}(t_1) = r\hat{\mathbf{v}}$ and $\mathbf{x}(t_2) = r\hat{\mathbf{w}}$ for all $\mathbf{x} \in \{\mathbf{c}, \widetilde{\mathbf{c}}\}$, then

$$r\cos(\pm t_1) = (r[\hat{\mathbf{v}}]_{\hat{\mathbf{e}}})_1 \text{ and } r\cos(\pm t_2) = (r[\hat{\mathbf{w}}]_{\hat{\mathbf{e}}})_1,$$

where the $\pm$ signs are both $+$ when $\mathbf{x} = \mathbf{c}$ and are both $-$ when $\mathbf{x} = \widetilde{\mathbf{c}}$. Thus

$$\theta(\mathbf{v}, \mathbf{w}) = t_2 - t_1 = \pm\arccos\Big(([\hat{\mathbf{w}}]_{\hat{\mathbf{e}}})_1 \Big) \mp \arccos\Big(([\hat{\mathbf{v}}]_{\hat{\mathbf{e}}})_1 \Big),$$

where the $+, -$ case occurs when $\mathbf{x} = \mathbf{c}$ and the $-, +$ case occurs when $\mathbf{x} = \widetilde{\mathbf{c}}$.

The cosine angle difference identity $\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$ implies[19] the identity $\arccos(\gamma) - \arccos(\delta) = \arccos(\gamma\delta + \sqrt{1 - \gamma^2}\sqrt{1 - \delta^2})$. In the $+, -$ case of the above, this identity gives

$$\theta(\mathbf{v}, \mathbf{w}) = \arccos\left(([\hat{\mathbf{w}}]_{\hat{\mathbf{e}}})_1 ([\hat{\mathbf{v}}]_{\hat{\mathbf{e}}})_1 + \sqrt{1 - ([\hat{\mathbf{w}}]_{\hat{\mathbf{e}}})_1^2}\sqrt{1 - ([\hat{\mathbf{v}}]_{\hat{\mathbf{e}}})_1^2} \right)$$

$$= \arccos\left(([\hat{\mathbf{w}}]_{\hat{\mathbf{e}}})_1 ([\hat{\mathbf{v}}]_{\hat{\mathbf{e}}})_1 + ([\hat{\mathbf{w}}]_{\hat{\mathbf{e}}})_2 ([\hat{\mathbf{v}}]_{\hat{\mathbf{e}}})_2 \right).$$

In the $-, +$ case, applying the identity yields an analogous (and ultimately equivalent) result. $\qquad\square$

Theorem 2.124. (Angle between two vectors is well-defined).

The formula for $\theta(\mathbf{v}, \mathbf{w})$ given by the previous theorem does not depend on the radius r of the circle whose center coincides with the initial points of $\mathbf{v}$ and $\mathbf{w}$, so we now know that $\theta(\mathbf{v}, \mathbf{w})$ is well-defined.

Definition 2.125. (The dot product of unit vectors).

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. The expression that appears inside arccos in the formula for $\theta(\mathbf{v}, \mathbf{w})$,

$$\sum_{i=1}^{2} ([\hat{\mathbf{v}}]_{\hat{\mathbf{e}}})_i ([\hat{\mathbf{w}}]_{\hat{\mathbf{e}}})_i,$$

is denoted $\hat{\mathbf{v}} \cdot \hat{\mathbf{w}}$ and is called the *dot product of $\hat{\mathbf{v}}$ and $\hat{\mathbf{w}}$*.

With this new definition, the formula for angle in $\mathbb{R}^2$ suddenly has a new, more compact form.

Theorem 2.126. (Angle formula in $\mathbb{R}^2$).

$$\theta(\mathbf{v}, \mathbf{w}) = \arccos(\hat{\mathbf{v}} \cdot \hat{\mathbf{w}}) \text{ for all } \mathbf{v}, \mathbf{w} \in \mathbb{R}^2.$$

The formula from the previous definition can also be used to define a dot product of non-unit vectors.

Definition 2.127. (The dot product in $\mathbb{R}^2$).

The *dot product $\mathbf{v} \cdot \mathbf{w}$ of $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$* is

$$\mathbf{v} \cdot \mathbf{w} := \sum_{i=1}^{2} ([\mathbf{v}]_{\hat{\mathbf{e}}})_i ([\mathbf{w}]_{\hat{\mathbf{e}}})_i.$$

Theorem 2.128. (Length in $\mathbb{R}^2$ in terms of dot product in $\mathbb{R}^2$).

$$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}} \text{ for all } \mathbf{v} \in \mathbb{R}^2.$$

Proof. We have $\mathbf{v} \cdot \mathbf{v} = ([\mathbf{v}]_{\hat{\mathbf{e}}})_1^2 + ([\mathbf{v}]_{\hat{\mathbf{e}}})_2^2$. Take the square root of of both sides to obtain the result. $\qquad\square$

Theorem 2.129. (The dot product in $\mathbb{R}^2$ is bilinear).

The dot product in $\mathbb{R}^2$ is a bilinear function. That is, both $\mathbf{v} \mapsto \mathbf{v} \cdot \mathbf{w}$ and $\mathbf{w} \mapsto \mathbf{v} \cdot \mathbf{w}$ are linear functions.

Proof. The dot product in $\mathbb{R}^2$ is symmetric, so it suffices to show that it is linear in either argument; it suffices to show that $\mathbf{v} \mapsto \mathbf{v} \cdot \mathbf{w}$ is a linear function. That is, it suffices to show that for all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in \mathbb{R}^2$ we have $(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w}$ and for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ and $c \in \mathbb{R}$ we have $(c\mathbf{v}) \cdot \mathbf{w} = c(\mathbf{v} \cdot \mathbf{w})$.

Note that because $[\cdot]_{\hat{\mathbf{e}}}$ is linear we have $([\mathbf{v}_1 + \mathbf{v}_2]_{\hat{\mathbf{e}}})_i = ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i + ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i$ and $([c\mathbf{v}]_{\hat{\mathbf{e}}})_i = c([\mathbf{v}]_{\hat{\mathbf{e}}})_i$ for all i. Using this fact, we see

$$(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \sum_{i=1}^{2} ([\mathbf{v}_1 + \mathbf{v}_2]_{\hat{\mathbf{e}}})_i ([\mathbf{w}]_{\hat{\mathbf{e}}})_i = \sum_{i=1}^{2} \left(([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i + ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i \right) ([\mathbf{w}]_{\hat{\mathbf{e}}})_i$$

$$= \sum_{i=1}^{2} \left(([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{w}]_{\hat{\mathbf{e}}})_i + ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i ([\mathbf{w}]_{\hat{\mathbf{e}}})_i \right) = \sum_{i=1}^{2} ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_i ([\mathbf{w}]_{\hat{\mathbf{e}}})_i + \sum_{i=1}^{2} ([\mathbf{v}_2]_{\hat{\mathbf{e}}})_i ([\mathbf{w}]_{\hat{\mathbf{e}}})_i$$

$$= \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w}$$

---

[19]To obtain the second identity, write the first identity in terms of γ, δ by making the substitutions $\gamma := \cos(\alpha) \iff \alpha = \arccos(\gamma)$ and $\delta := \cos(\beta) \iff \beta = \arccos(\delta)$, and take arccos of both sides.

and

$$(c\mathbf{v}) \cdot \mathbf{w} = \sum_{i=1}^{2} ([c\mathbf{v}]_{\hat{\mathfrak{e}}})_i ([\mathbf{w}]_{\hat{\mathfrak{e}}})_i = \sum_{i=1}^{2} c([\mathbf{v}]_{\hat{\mathfrak{e}}})_i ([\mathbf{w}]_{\hat{\mathfrak{e}}})_i = c\sum_{i=1}^{2} ([\mathbf{v}]_{\hat{\mathfrak{e}}})_i ([\mathbf{w}]_{\hat{\mathfrak{e}}})_i = c(\mathbf{v} \cdot \mathbf{w}),$$

as claimed. $\qquad\square$

With this expanded definition of the dot product, it's possible to express the dot product of $\mathbf{v}$ and $\mathbf{w}$ in terms of the angle between $\mathbf{v}$ and $\mathbf{w}$.

Theorem 2.130. (Dot product in $\mathbb{R}^2$ in terms of angle in $\mathbb{R}^2$).

$$\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}|| \, ||\mathbf{w}|| \cos(\theta(\mathbf{v}, \mathbf{w})) \text{ for all } \mathbf{v}, \mathbf{w} \in \mathbb{R}^2.$$

Proof. From the previous theorem, we have $\theta(\mathbf{v}, \mathbf{w}) = \arccos\left(\frac{\mathbf{v}}{||\mathbf{v}||} \cdot \frac{\mathbf{w}}{||\mathbf{w}||}\right)$. Since the dot product is bilinear, we have $\frac{\mathbf{v}}{||\mathbf{v}||} \cdot \frac{\mathbf{w}}{||\mathbf{w}||} = \frac{1}{||\mathbf{v}||}\left(\mathbf{v} \cdot \frac{\mathbf{w}}{||\mathbf{w}||}\right) = \frac{1}{||\mathbf{v}|| \, ||\mathbf{w}||}(\mathbf{v} \cdot \mathbf{w}) = \frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}|| \, ||\mathbf{w}||}$. Thus $\theta(\mathbf{v}, \mathbf{w}) = \arccos\left(\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}|| \, ||\mathbf{w}||}\right)$. Take cos of both sides and then multiply both sides by $||\mathbf{v}|| \, ||\mathbf{w}||$ to obtain the result. $\qquad\square$

Remark 2.131. (The dot product of perpendicular vectors in $\mathbb{R}^2$ is zero).
It follows immediately from the previous theorem that $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ are perpendicular iff $\mathbf{v} \cdot \mathbf{w} = 0$.

At this point, we've seen that the "algebraic" characterization of the dot product, $\mathbf{v} \cdot \mathbf{w} := \sum_{i=1}^{2}([\mathbf{v}]_{\hat{\mathfrak{e}}})_i([\mathbf{w}]_{\hat{\mathfrak{e}}})_i$, and the "geometric" characterization of the dot product, $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}|| \, ||\mathbf{w}|| \cos(\theta(\mathbf{v}, \mathbf{w}))$, are true (with the algebraic characterization being true by definition), but don't yet understand their connection. We now discover how each characterization follows from the other.

Prerequisites for understanding the dot product

Vector projection

Lemma 2.132. (Orthogonal basis of $\mathbb{R}^2$).
Let $\mathbf{v} \in \mathbb{R}^2$. If $\mathbf{v}_\perp$ is perpendicular to $\mathbf{v}$, then $\{\mathbf{v}, \mathbf{v}_\perp\}$ is a basis of $\mathbb{R}^2$.

Proof. Recall from Lemma 2.42 that n linearly independent vectors taken from the same n-dimensional vector space span that vector space. Since $\dim(\mathbb{R}^2) = 2$, it suffices to show that $\mathbf{v}, \mathbf{v}_\perp$ are linearly independent.

Suppose for contradiction that $\mathbf{v}, \mathbf{v}_\perp$ are linearly dependent. Then there exist $c, d \in \mathbb{R}$ not both zero such that $c\mathbf{v} + d\mathbf{v}_\perp = \mathbf{0}$. Assume without loss of generality that $c \neq 0$. Then we can divide by c, and we have $\mathbf{v} = -\frac{d}{c}\mathbf{v}_\perp$. Taking the dot product of each side with $\mathbf{v}$, we obtain $\mathbf{v} \cdot \mathbf{v} = -\frac{d}{c}\mathbf{v}_\perp \cdot \mathbf{v}$. Since $\mathbf{v}$ and $\mathbf{v}_\perp$ are perpendicular, we have $\mathbf{v}_\perp \cdot \mathbf{v} = 0$ and thus $\mathbf{v} \cdot \mathbf{v} = 0$, so $\mathbf{v}$ is perpendicular to itself. Since $\mathbf{v}$ was arbitrary, this shows that every vector in $\mathbb{R}^2$ is perpendicular to itself. This is clearly not true; contradiction. $\square$

Definition 2.133. (Vector projection in $\mathbb{R}^2$).
Consider vectors $\mathbf{v}, \mathbf{w}, \mathbf{w}_\perp \in \mathbb{R}^2$, where $\mathbf{w}_\perp$ is perpendicular to $\mathbf{w}$. Since $\{\hat{\mathbf{w}}, \hat{\mathbf{w}}_\perp\}$ is a basis of $\mathbb{R}^2$, there exist unique $v_{||}, v_\perp \in \mathbb{R}$ such that $\mathbf{v} = v_{||}\hat{\mathbf{w}} + v_\perp\hat{\mathbf{w}}_\perp$.
The *vector projection* of $\mathbf{v}$ is $\text{proj}(\mathbf{v} \to \mathbf{w}) := v_{||}\hat{\mathbf{w}}$.
The *signed magnitude of the vector projection* of $\mathbf{v}$ onto $\mathbf{w}$ is $\text{sproj}(\mathbf{v} \to \mathbf{w}) := v_{||}$.

Remark 2.134. Note that, for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$, we have $\text{proj}(\mathbf{v} \to \mathbf{w}) = \text{proj}(\mathbf{v} \to \hat{\mathbf{w}})$ because $\hat{\hat{\mathbf{w}}} = \hat{\mathbf{w}}$.

Theorem 2.135. (Vector projection and its signed magnitude in $\mathbb{R}^2$).
Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. We have the following.

1. $\text{proj}(\mathbf{v} \to \mathbf{w}) = \text{sproj}(\mathbf{v} \to \mathbf{w})\hat{\mathbf{w}}$.

2. $\text{sproj}(\mathbf{v} \to \mathbf{w}) = \pm||\text{proj}(\mathbf{v} \to \mathbf{w})||$.

This second fact is what should be emphasized when thinking about the signed magnitude of a vector projection.

Proof. (1) follows from the definition of $\text{proj}(\mathbf{v} \to \mathbf{w})$. (2) is obtained by taking the magnitude of each side of (1). $\square$

Orthogonal linear functions

Theorem 2.136. (Preserved quantities and linear functions).
If $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear function, then the following statements are equivalent:

1. ($\mathbf{f}$ preserves length). $||\mathbf{v}|| = ||\mathbf{f}(\mathbf{v})||$ for all $\mathbf{v} \in \mathbb{R}^2$.

2. ($\mathbf{f}$ preserves dot product). $\mathbf{v} \cdot \mathbf{w} = \mathbf{f}(\mathbf{v}) \cdot \mathbf{f}(\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$.

3. ($\mathbf{f}$ preserves length and angle). $||\mathbf{v}|| = ||\mathbf{f}(\mathbf{v})||$ for all $\mathbf{v} \in \mathbb{R}^2$ and $\theta(\mathbf{v}, \mathbf{w}) = \theta(\mathbf{f}(\mathbf{v}), \mathbf{f}(\mathbf{w}))$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$.

Proof.

- ((1) $\implies$ (2)). Suppose $\mathbf{f}$ preserves length. If we showed that $\mathbf{v} \cdot \mathbf{w}$ depends on $||\mathbf{v}||$ and $||\mathbf{w}||$ (on lengths), it would follow that $\mathbf{f}$ preserves dot product. This is what we will do.

  We need to show that $\mathbf{v} \cdot \mathbf{w}$ depends on $||\mathbf{v}||$ and $||\mathbf{w}||$. Note that $||\mathbf{v}||^2 = \mathbf{v} \cdot \mathbf{v}$ for $\mathbf{v} \in \mathbb{R}^n$. So $||\mathbf{v} + \mathbf{w}||^2 = (\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{v} \cdot \mathbf{v} + 2\mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} = ||\mathbf{v}||^2 + 2\mathbf{v} \cdot \mathbf{w} + ||\mathbf{w}||^2$. Solve for $\mathbf{v} \cdot \mathbf{w}$ to see that $\mathbf{v} \cdot \mathbf{w}$ does indeed depend on $||\mathbf{v}||$ and $||\mathbf{w}||$:

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

  ((1) $\impliedby$ (2)). If $\mathbf{f}$ preserves dot product, then $||\mathbf{f}(\mathbf{v})|| = \sqrt{\mathbf{f}(\mathbf{v}) \cdot \mathbf{f}(\mathbf{v})} = \sqrt{\mathbf{v} \cdot \mathbf{v}} = ||\mathbf{v}||$, so $\mathbf{f}$ preserves length.

- ((2) $\implies$ (3)). If $\mathbf{f}$ preserves length, then, since angle is a function of quantities that are preserved by $\mathbf{f}$ (dot product and length), $\theta(\mathbf{v}, \mathbf{w}) = \arccos\left(\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||\,||\mathbf{w}||}\right)$, it too is preserved by $\mathbf{f}$.

  ((2) $\impliedby$ (3)). Suppose $\mathbf{f}$ preserves angle: $\theta(\mathbf{v}, \mathbf{w}) = \theta(\mathbf{f}(\mathbf{v}), \mathbf{f}(\mathbf{w}))$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. Then $\mathbf{f}(\mathbf{v}) \cdot \mathbf{f}(\mathbf{w}) = ||\mathbf{f}(\mathbf{v})||\,||\mathbf{f}(\mathbf{w})|| \cos(\theta(\mathbf{f}(\mathbf{v}), \mathbf{f}(\mathbf{w})))$

$\square$

Definition 2.137. (Orthogonal linear function).

We say that a linear function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is *orthogonal* iff it preserves dot product.

Remark 2.138. (Orthogonal linear function).

"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.

Rotation

Definition 2.139. (Rotation in $\mathbb{R}^2$).

We define the *rotation* $\mathbf{R}_\theta$ *by* $\theta \in [0, 2\pi)$ to be the linear function $\mathbf{R}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ that sends $\hat{\mathfrak{e}}_1$ to $(\cos(\theta), \sin(\theta))^\top$ and $\hat{\mathfrak{e}}_2$ to $(-\sin(\theta), \cos(\theta))$.

Drawing a picture and using trigonometry confirms that this definition agrees with our intuition on rotation.

Theorem 2.140. (Rotation in $\mathbb{R}^2$).

We have

$$\mathbf{R}_\theta(\mathbf{v}) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \mathbf{v} \text{ for all } \mathbf{v} \in \mathbb{R}^2.$$

Theorem 2.141. (A rotation by ϕ changes angle by ϕ). We have $\cos(\theta(\mathbf{v}, \mathbf{R}_\phi(\mathbf{w}))) = \cos(\theta(\mathbf{v}, \mathbf{w}) + \phi)$ for all $\phi \in [0, 2\pi)$.

Proof.

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

$\square$

Proof. Let $\mathbf{c}$ and $\widetilde{\mathbf{c}}$ be the parameterizations of the circle of radius r with center $\mathbf{0}$ from Definition 2.119.

We have

$$\theta(\mathbf{v}, \mathbf{R}_s(\mathbf{w})) = \frac{\min(s(\mathbf{c}, \mathbf{v}, \mathbf{R}_\phi(\mathbf{w})), s(\widetilde{\mathbf{c}}, \mathbf{v}, \mathbf{R}_\phi(\mathbf{w})))}{r} = \begin{cases} \theta(\mathbf{v}, \mathbf{w}) + \phi & \theta(\mathbf{v}, \mathbf{w}) + \phi \leq \pi \\ 2\pi - (\theta(\mathbf{v}, \mathbf{w}) + \phi) & \theta(\mathbf{v}, \mathbf{w}) + \phi > \pi \end{cases}.$$

Take cos of both sides to obtain the result. $\square$

Theorem 2.142. Rotations in $\mathbb{R}^2$ are orthogonal linear functions.

Proof. It suffices to show that rotations preserve length. So, for $\mathbf{v} \in \mathbb{R}^2$, consider $||\mathbf{v}||$ and $||\mathbf{R}_\theta(\mathbf{v})||$.

We have $\mathbf{R}_\phi(\mathbf{v}) = \mathbf{R}_\phi(||\mathbf{v}||\hat{\mathbf{v}}) = ||\mathbf{v}||\mathbf{R}_\phi(\hat{\mathbf{v}})$. Since $\hat{\mathbf{v}} = (\cos(\theta), \sin(\theta))^\top$ for some $\theta \in [0, 2\pi)$, ... $\square$

Theorem 2.143. $\mathbf{R}_\theta(\hat{\mathbf{v}}) = \widehat{\mathbf{R}_\theta(\mathbf{v})}$

Proof. $\square$

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

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. For any $\theta \in [0, 2\pi)$, we have $\mathbf{R}_\theta(\text{proj}(\mathbf{v} \to \mathbf{w})) = \text{proj}(\mathbf{R}_\theta(\mathbf{v}) \to \mathbf{R}_\theta(\mathbf{w}))$.

Proof. We have $\mathbf{v} = v_{||}\hat{\mathbf{v}} + v_\perp\hat{\mathbf{v}}_\perp$, so $\mathbf{R}_\theta(\mathbf{v}) = v_{||}\mathbf{f}(\hat{\mathbf{v}}) + v_\perp\mathbf{R}_\theta\hat{\mathbf{v}}_\perp$. The claim follows if we show (1) that $\mathbf{R}_\theta(\hat{\mathbf{w}}) = \widehat{\mathbf{R}_\theta(\hat{\mathbf{w}})}$ and (2) that $\mathbf{R}_\theta(\hat{\mathbf{w}}_\perp) = \widehat{\mathbf{R}_\theta(\mathbf{w})}_\perp$. (1) is true because rotations are length-preserving. (2) is true because there exists a rotation that sends $\mathbf{w}$ to $\mathbf{w}_\perp$, because rotations commute with each other, and because rotations are length-preserving. $\square$

Geometric characterization $\implies$ algebraic characterization

Lemma 2.145. (Signed projection onto a vector in $\mathbb{R}^2$ is a linear function).
    Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$. The map $\mathbf{v}_1 \mapsto \mathrm{sproj}(\mathbf{v}_1 \to \mathbf{v}_2)$ is linear.

Proof. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in \mathbb{R}^2$, and define $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}$ by $\mathbf{f}(\mathbf{v}) = \mathrm{proj}(\mathbf{v} \to \mathbf{w})$. We need to show that $\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2)$ and that $\mathbf{f}(c\mathbf{v}) = c\mathbf{f}(\mathbf{v})$ for all $c \in \mathbb{R}$.
    We have

$$
\begin{aligned}
\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) &= \mathbf{f}\Big(\big((v_1)_{||}\hat{\mathbf{w}} + (v_1)_{\perp}\hat{\mathbf{w}}_{\perp} \big) + \big((v_2)_{||}\hat{\mathbf{w}} + (v_2)_{\perp}\hat{\mathbf{w}}_{\perp} \big) \Big), \ \text{where } (v_i)_{||} = \mathrm{sproj}(\mathbf{v}_i \to \mathbf{w}) \text{ for } i \in \{1, 2\} \\
&= \mathbf{f}\Big(\big((v_1)_{||} + (v_2)_{||} \big)\hat{\mathbf{w}} + \big((v_1)_{\perp} + (v_2)_{\perp} \big)\hat{\mathbf{w}}_{\perp} \Big) \\
&= \mathrm{sproj}\Big(\big[\big((v_1)_{||} + (v_2)_{||} \big)\hat{\mathbf{w}} + \big((v_1)_{\perp} + (v_2)_{\perp} \big)\hat{\mathbf{w}}_{\perp} \big] \to \mathbf{w} \Big) \\
&= (v_1)_{||} + (v_2)_{||} = \mathrm{sproj}(\mathbf{v}_1 \to \mathbf{w}) + \mathrm{sproj}(\mathbf{v}_2 \to \mathbf{w}) = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2).
\end{aligned}
$$

And we have

$$
\begin{aligned}
\mathbf{f}(c\mathbf{v}) &= \mathbf{f}\Big(c\big(v_{||}\hat{\mathbf{w}} + v_{\perp}\hat{\mathbf{w}}_{\perp} \big) \Big), \ \text{where } v_{||} = \mathrm{sproj}(\mathbf{v} \to \mathbf{w}) \\
&= \mathbf{f}\Big(cv_{||}\hat{\mathbf{w}} + cv_{\perp}\hat{\mathbf{w}}_{\perp} \Big) = \mathrm{sproj}\Big(\big(cv_{||}\hat{\mathbf{w}} + cv_{\perp}\hat{\mathbf{w}}_{\perp} \big) \to \mathbf{w} \Big) \\
&= cv_{||}\hat{\mathbf{w}} = c\,\mathrm{sproj}(\mathbf{v} \to \mathbf{w}) = c\mathbf{f}(\mathbf{v}).
\end{aligned}
$$

$\square$

Theorem 2.146. (The dot product in $\mathbb{R}^2$ is a bilinear function).
    The dot product in $\mathbb{R}^2$ is a bilinear function. That is, both $\mathbf{v} \mapsto \mathbf{v} \cdot \mathbf{w}$ and $\mathbf{w} \mapsto \mathbf{v} \cdot \mathbf{w}$ are linear functions.

Proof. The dot product in $\mathbb{R}^2$ is symmetric, so it suffices to show that it is a linear function in either argument; it suffices to show that $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}$ defined by $\mathbf{f}(\mathbf{v}) = \mathbf{v} \cdot \mathbf{w}$ is a linear function. Well, $\mathbf{f}(\mathbf{v}) = ||\mathbf{w}||\,\mathrm{sproj}(\mathbf{v} \to \mathbf{w})$, and we know that $\mathbf{v} \mapsto \mathrm{sproj}(\mathbf{v} \to \mathbf{w})$ is linear. Since $\mathbf{f}$ is the result of scaling a linear function by $||\mathbf{w}||$, it too is a linear function. $\square$

    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 $\mathbb{R}^2$. Before we derive this alternate expression, though, we will need to note the following basic fact.

Lemma 2.147. Let $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ be the standard basis for $\mathbb{R}^2$. We have

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

for all $i, j \in \{1, 2\}$.

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

Derivation 2.148. (Algebraic dot product in $\mathbb{R}^2$).
    We can now derive the algebraic characterization of the dot product by using its bilinearity (Theorem 2.146) together with the previous lemma.
    We will make use of the following general fact about bilinear functions[20]. 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

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

---

[20]If V_1, V_2, and W are vector spaces over a field K, then $\mathbf{f} : V_1 \times V_2 \to W$ is a *bilinear* function iff $\mathbf{v} \mapsto \mathbf{f}(\mathbf{v}, \mathbf{w})$ and $\mathbf{w} \mapsto \mathbf{f}(\mathbf{v}, \mathbf{w})$ are both linear functions.

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

The dot product is a bilinear function $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, 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

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

Therefore

$$\boxed{\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{2} ([\mathbf{v}]_{\hat{\mathfrak{E}}})_i ([\mathbf{w}]_{\hat{\mathfrak{E}}})_i}$$

Algebraic characterization $\implies$ geometric characterization

Derivation 2.149. (Algebraic characterization $\implies$ geometric characterization).

If

$$\mathbf{v} \cdot \mathbf{w} := \sum_{i=1}^{2} ([\mathbf{v}]_{\hat{\mathfrak{E}}})_i ([\mathbf{w}]_{\hat{\mathfrak{E}}})_i$$

then

$$\mathbf{v} \cdot \mathbf{w} := \mathrm{sproj}(\mathbf{v} \to \mathbf{w}) \, ||\mathbf{w}|| \text{ for all } \mathbf{v}, \mathbf{w} \in \mathbb{R}^2.$$

Proof. Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ and consider $\mathbf{v} \cdot \mathbf{w}$. We can[21] choose $\theta \in [0, 2\pi)$ so that $\mathbf{R}_\theta(\hat{\mathbf{v}}) = \hat{\mathfrak{e}}_1$, i.e., so that $\mathbf{R}_\theta(\mathbf{v}) = ||\mathbf{v}||\hat{\mathfrak{e}}_1$. With this choice of θ, we have

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

We have

$$([\mathbf{R}_\theta(\mathbf{w})]_{\hat{\mathfrak{E}}})_1 = \mathrm{proj}(\mathbf{R}_\theta(\mathbf{w}) \to \hat{\mathfrak{e}}_1) = \mathrm{proj}(\mathbf{R}_\theta(\mathbf{w}) \to \mathbf{R}_\theta(\hat{\mathbf{v}})) = \mathrm{proj}(\mathbf{R}_\theta(\mathbf{w}) \to \widehat{\mathbf{R}_\theta(\mathbf{v})})$$
$$= \mathrm{proj}(\mathbf{R}_\theta(\mathbf{w}) \to \mathbf{R}_\theta(\mathbf{v})) = \mathrm{proj}(\mathbf{w} \to \mathbf{v}),$$

where the last equality is by Lemma **??**.

Therefore $\mathbf{v} \cdot \mathbf{w} = ||\mathrm{proj}(\mathbf{w} \to \mathbf{v})|| \, ||\mathbf{v}||$. Since the dot product is symmetric, we can swap $\mathbf{v}$ and $\mathbf{w}$; this gives the desired result. $\qquad \square$

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

The algebraic characterization of the dot product in $\mathbb{R}^2$ generalizes straightforwardly to $\mathbb{R}^n$.

Definition 2.150. (Dot product in $\mathbb{R}^n$).

We define the *dot product of* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ to be

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

Theorem 2.151. (Length in $\mathbb{R}^n$).

For all $\mathbf{v} \in \mathbb{R}^n$ we have $||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

Proof. We have $\mathbf{v} \cdot \mathbf{v} = \sum_{i=1}^{n} ([\mathbf{v}]_{\hat{\mathfrak{E}}})_i^2$. Take the square root of both sides to obtain the result. $\qquad \square$

---

[21]Use $\theta = \arccos(||\mathbf{v}||)$.

In $\mathbb{R}^2$, we were able to derive that $\theta(\mathbf{v}, \mathbf{w}) = \arccos(\hat{\mathbf{v}} \cdot \hat{\mathbf{w}})$. This result inspires our definition of angle in $\mathbb{R}^n$.

Definition 2.152. (Angle in $\mathbb{R}^n$).
We define the *angle between* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ to be

$$\theta(\mathbf{v}, \mathbf{w}) := \arccos(\hat{\mathbf{v}} \cdot \hat{\mathbf{w}}).$$

Vector projection in terms of dot produt

In the case of $n = 2$, we could have proved the following theorem immediately after defining $\mathbf{v} \cdot \mathbf{w} := \text{sproj}(\mathbf{v} \to \mathbf{w}) \, ||\mathbf{w}||$ for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ in Definition [...]. The usefulness of this next theorem, however, wouldn't have been apparent until knowing that $\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n}([\mathbf{v}]_{\hat{\mathfrak{e}}})_i([\mathbf{w}]_{\hat{\mathfrak{e}}})_i$.

Theorem 2.153. (Vector projection in $\mathbb{R}^n$ terms of dot product in $\mathbb{R}^n$).
If $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, then

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

In the case $n = 2$, if we didn't know the algebraic characterization of the dot product, this theorem would be a bit of a tautology. However, knowing the algebraic characterization of the dot product allows for easy computation of the expressions $\mathbf{v} \cdot \mathbf{w}$ and $\mathbf{w} \cdot \mathbf{w}$ whenever we know the coordinates of $\mathbf{v}$ and $\mathbf{w}$ relative to the standard basis[22] for $\mathbb{R}^2$.

Proof. Since $\mathbf{v} \cdot \mathbf{w} = \text{sproj}(\mathbf{v} \to \mathbf{w}) \, ||\mathbf{w}||$, then $\text{sproj}(\mathbf{v} \to \mathbf{w}) = \frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{w}||}$, and, by the first fact of Theorem 2.135, $\text{proj}(\mathbf{v} \to \mathbf{w}) = \text{sproj}(\mathbf{v} \to \mathbf{w})\hat{\mathbf{w}} = \frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{w}||}\hat{\mathbf{w}}$. $\qquad\square$

The law of cosines

The next theorem presents the law of cosines as it is best understood, which is as a consequence of the equivalence between the geometric and algebraic dot product formulas.

Theorem 2.154. (Law of cosines in $\mathbb{R}^n$).
We have

$$||\mathbf{v} - \mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2 - 2||\mathbf{v}|| \, ||\mathbf{w}|| \cos(\theta(\mathbf{v}, \mathbf{w})).$$

Note that when $\theta(\mathbf{v}, \mathbf{w}) = 0$, we recover the Pythagorean theorem.

Proof. We have $||\mathbf{v} - \mathbf{w}||^2 = (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w}) = \mathbf{v} \cdot \mathbf{v} - 2\mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} = ||\mathbf{v}||^2 + ||\mathbf{w}||^2 - 2||\mathbf{v}|| \, ||\mathbf{w}|| \cos(\theta)$. $\quad\square$

---

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

Inner product spaces

This section generalizes the notion of the dot product, which we only know to be applicable to $\mathbb{R}^n$, to general vector spaces.

Definition 2.155. (Inner product).

Let V be a vector space over a field K. An *inner product* on V is a function $\langle \cdot, \cdot \rangle : V \times V \to K$ with the following properties:

1. (Bilinearity). The functions $\mathbf{v} \mapsto \langle \mathbf{v}, \mathbf{w} \rangle$ and $\mathbf{w} \mapsto \langle \mathbf{v}, \mathbf{w} \rangle$ are linear for all $\mathbf{v}, \mathbf{w} \in V$. In other words $\langle \cdot, \cdot \rangle$ is "linear in each argument".

2. (Symmetry). $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ for all $\mathbf{v}, \mathbf{w} \in V$.

3. (Nondegeneracy). $\langle \mathbf{v}, \mathbf{v} \rangle = 0 \implies \mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in V$.

Inner products are called "inner products" because they generalize the dot product, which satisfies $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^\top \mathbf{w}$ for all $\mathbf{v}, \mathbf{w} \in K^n$; notice how the position of the transpose $\top$ is "inside" $\mathbf{v}$ and $\mathbf{w}$, here. Contrastingly, the "outer product" of vectors $\mathbf{w}, \mathbf{v} \in K^n$ is sometimes defined to be $\mathbf{w}\mathbf{v}^\top$, since the position of the transpose $\top$ is on the "outside" of $\mathbf{w}$ and $\mathbf{v}$. We will learn more about the outer product in Remark 3.20.

Remark 2.156. (Inner product conventions).

Some authors define inner products to satisfy the "positive-definitness" criterion rather than the nondegeneracy criterion. A function $V \times V \to K$ is *positive-definite* iff $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ for all $\mathbf{v} \in V$, with $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ occurring only when $\mathbf{v} = \mathbf{0}$. A *negative-definite* inner product is defined similarly and involves the condition $\langle \mathbf{v}, \mathbf{v} \rangle \leq 0$ for all $\mathbf{v} \in V$.

Definition 2.157. (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.*

Example 2.158.

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

The dot product in K^n, defined analogously to the dot product in $\mathbb{R}^n$, 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 2.159. (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 in $\mathbb{R}^n$ (see Theorem **??**), 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 2.160. (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 2.161. (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 in $\mathbb{R}^n$, which was discussed in Theorem **??**).

Theorem 2.162. (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:
$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle \leq ||\mathbf{v}_1|| \, ||\mathbf{v}_2||$

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 2.163. (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 2.164. (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 2.165. (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

$$\mathbf{f}_i := \mathbf{e}_i - \text{proj}(\mathbf{f}_i \rightarrow \text{span}(\mathbf{e}_1, ..., \cancel{\mathbf{e}_i}, ..., \mathbf{e}_n)) = \mathbf{e}_i - \sum_{j \neq i} \text{proj}(\mathbf{f}_i \rightarrow \mathbf{e}_j) = \mathbf{e}_i - \sum_{j \neq i} \frac{\langle \mathbf{f}_i, \mathbf{e}_j \rangle}{\langle \mathbf{e}_j, \mathbf{e}_j \rangle}, \quad i \geq 2.$$

(In the last equality in the line above, we've used an analogue of Theorem 2.153 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 $\hat{\mathbf{u}}_i := \frac{\mathbf{f}_i}{||\mathbf{f}_i||}$.

Remark 2.166. The above theorem reveals that the algebraic dot product (in $\mathbb{R}^2$) 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 $||\mathbf{v}_i||^2 = \sum_{j=1}^{2}(([\mathbf{v}_i]_{\hat{\mathbf{e}}})_j)^2$ to discover that we must have $([\mathbf{v}_1]_{\hat{\mathbf{e}}})_1([\mathbf{v}_2]_{\hat{\mathbf{e}}})_1 + ([\mathbf{v}_1]_{\hat{\mathbf{e}}})_2([\mathbf{v}_2]_{\hat{\mathbf{e}}})_2 = 0$.

2.5   Systems of linear equations

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 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[23], and that systems of linear equations should be viewed as an application of linear algebra.

Definition 2.167. (Linear equation).

Let V and W be vector spaces, let $\mathbf{f} : V \to W$ be a function, and let $\mathbf{w} \in W$. When $\mathbf{f}$ is linear, the equation

$$\mathbf{f}(\mathbf{v}) = \mathbf{w} \text{ for all } \mathbf{v} \in V$$

is called a *linear equation.*

The set of solutions to the equation, $\{\text{solns}\} := \{\mathbf{v} \in V \mid \mathbf{f}(\mathbf{v}) = \mathbf{w}\}$, is a vector space whenever it is nonempty.

We first present a recursive characterization of solutions to linear equations. This isn't an algorithm for finding solutions; it's only a characterization.

Theorem 2.168. (Solutions to linear equations).

Let V and W be vector spaces, let $\mathbf{f} : V \to W$ be a linear function, and let $\mathbf{w} \in W$. Consider the linear equation

$$\mathbf{f}(\mathbf{v}) = \mathbf{w} \text{ for all } \mathbf{v} \in V.$$

When the set of solutions $\{\text{solns}\}$ is nonempty, it can be described in terms of any particular solution:

$$\{\text{solns}\} = \{\mathbf{v} + \mathbf{v}_0 \mid \mathbf{v}_0 \in \ker(\mathbf{f})\} \text{ for any } \mathbf{v} \in \{\text{solns}\}.$$

This theorem is very useful when in the case when $\mathbf{f}$ is not invertible, as it says that we can obtain the entire solution set by finding a *single* solution to the equation, finding the kernel of $\mathbf{f}$, and then combining these two pieces of information in a simple way.

Proof.

Since the theorem assumes solutions to the linear equation exist, we know there is some $\mathbf{v} \in \{\text{solns}\}$.

(Case: $\mathbf{f}$ is invertible). In this case, $\ker(\mathbf{f}) = \{\mathbf{0}\}$ and $\mathbf{v} = \mathbf{f}^{-1}(\mathbf{w})$, so $\{\mathbf{v} + \mathbf{v}_0 \mid \mathbf{v}_0 \in \ker(\mathbf{f})\} = \{\mathbf{v}\} = \{\mathbf{f}^{-1}(\mathbf{w})\}$. Since $\mathbf{f}$ is invertible, the solution to the linear equation is unique, and $\{\text{solns}\} = \{\mathbf{f}^{-1}(\mathbf{w})\}$. Thus $\{\mathbf{v} + \mathbf{v}_0 \mid \mathbf{v}_0 \in \ker(\mathbf{f})\} = \{\mathbf{f}^{-1}(\mathbf{w})\} = \{\text{solns}\}$.

(Case: $\mathbf{f}$ is not invertible). We need to show $\{\mathbf{v} + \mathbf{v}_0 \mid \mathbf{v}_0 \in \ker(\mathbf{f})\} \subseteq \{\text{solns}\}$ and $\{\mathbf{v} + \mathbf{v}_0 \mid \mathbf{v}_0 \in \ker(\mathbf{f})\} \supseteq \{\text{solns}\}$.

($\subseteq$). Let $\mathbf{u}$ be in the first set, so $\mathbf{u} = \mathbf{v} + \mathbf{v}_0$ for some $\mathbf{v}_0 \in \ker(\mathbf{f})$. We have $\mathbf{f}(\mathbf{u}) = \mathbf{f}(\mathbf{v} + \mathbf{v}_0) = \mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{v}_0) = \mathbf{f}(\mathbf{v}) + \mathbf{0} = \mathbf{f}(\mathbf{v}) = \mathbf{w}$. Overall, we have $\mathbf{f}(\mathbf{u}) = \mathbf{w}$, so $\mathbf{u}$ is in the second set.

($\supseteq$). Let $\mathbf{u}$ be in the second set, so $\mathbf{f}(\mathbf{u}) = \mathbf{w}$. We have $\mathbf{f}(\mathbf{u} - \mathbf{v}) = \mathbf{f}(\mathbf{u}) - \mathbf{f}(\mathbf{v}) = \mathbf{w} - \mathbf{w} = \mathbf{0}$, so $\mathbf{u} - \mathbf{v} \in \ker(\mathbf{f})$. That is, $\mathbf{u} - \mathbf{v} = \mathbf{v}_0$, where $\mathbf{v}_0 \in \ker(\mathbf{f})$. We conclude that $\mathbf{u} = \mathbf{v} + \mathbf{v}_0$, where $\mathbf{v}_0 \in \ker(\mathbf{f})$, so $\mathbf{u}$ is in the first set. □

The above characterization implies the following theorem.

---

[23]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!

Theorem 2.169. (Number of solutions to a linear equation).

A linear equation has either zero, one, or infinitely many solutions.

Proof. Let V and W be finite-dimensional vector spaces, let $\mathbf{f} : V \to W$ be a linear function, and let $\mathbf{w} \in W$. Consider the linear equation $\mathbf{f}(\mathbf{v}) = \mathbf{w}$.

(Case: there are no solutions).

(Case: there are solutions). There are zero solutions in this case.

(Case: $\mathbf{f}$ is invertible). We saw in the previous proof that the solution set is $\{\mathbf{f}^{-1}(\mathbf{w})\}$ in this case. There is one solution in this case.

(Case: $\mathbf{f}$ is not invertible). In general, the solution set is $\{\mathbf{v} + \mathbf{v}_0 \mid \mathbf{v}_0 \in \ker(\mathbf{f})\}$. Since $\mathbf{f}$ is not invertible, then $\ker(\mathbf{f})$ has dimension at least 1, and so the solution set is a vector-space[24] of dimension at least 1. Vector spaces of dimension at least 1 contain infinitely many elements, so the solution set is infinite when $\mathbf{f}$ is not invertible. $\qquad\square$

The first theorem in this chapter recursively characterizes the problem of solving linear equations, telling us that in order to solve ($\mathbf{f}(\mathbf{v}) = \mathbf{w}$ for all $\mathbf{v} \in V$), we have to solve another linear equation, ($\mathbf{f}(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$). We haven't yet discussed a method for actually solving either of these linear equations, though!

We address this now. In order to solve a linear equation, we will transform it into an equivalent system of linear equations, and then solve that.

Derivation 2.170. (System of linear equations corresponding to a linear equation).

Let V and W be n- and m-dimensional vector spaces over a field K, let $\mathbf{f} : V \to W$ be a linear function, let $\mathbf{w} \in W$, and consider the linear equation

$$\mathbf{f}(\mathbf{v}) = \mathbf{w} \text{ for all } \mathbf{v} \in V.$$

Making use of the characterizing property of the matrix $[\mathbf{f}(E)]_F$ of $\mathbf{f}$ relative to E and F (recall the result at the end of Derivation 2.103), the above equation is equivalent to

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

If we let $[\mathbf{f}(E)]_F$ vary[25] over $K^{m \times n}$, $\mathbf{x} := [\mathbf{v}]_E$ vary over K^n, and $\mathbf{b} := [\mathbf{w}]_F$ vary over K^m, we see that the above equation is of the form

$$\mathbf{A}\mathbf{x} = \mathbf{b} \text{ for all } \mathbf{x} \in K^n.$$

Equations of the above form are called *(linear) matrix equations*[26].

Notice that if $\mathbf{x} \in K^n$ has ith component x_i and $\mathbf{b} \in K^m$ has ith component b_i, then the above linear matrix equation is the same as

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \text{ for all } x_1, ..., x_n \in K$$

$$\begin{pmatrix} x_1 a_{11} + ... + x_n a_{1n} \\ \vdots \\ x_1 a_{m1} + ... + x_n a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \text{ for all } x_1, ..., x_n \in K,$$

which is equivalent to the system of linear equations

$$\begin{cases} x_1 a_{11} + ... + x_n a_{1n} = b_1 \\ \vdots \\ x_1 a_{m1} + ... + x_n a_{mn} = b_m \end{cases} \text{ for all } x_1, ..., x_n \in K.$$

Each equation in this system is itself a linear equation because the function $(x_1, ..., x_n) \mapsto x_1 a_{i1} + ... + x_n a_{in}$ is linear for all i.

---

[24]Check for yourself that when $\ker(\mathbf{f})$ has dimension of at least one, the solution set is a vector space with dimension equal to that of $\ker(\mathbf{f})$.

[25]We can vary $\mathbf{A} := [\mathbf{f}(E)]_F$ over $K^{m \times n}$ because for any set of vectors, there exists a linear function sending basis vectors to those vectors (recall Theorem 2.55). We can let $[\mathbf{v}]_E$ vary over K^n and $[\mathbf{w}]_F$ vary over K^m because $[\cdot]_E$ and $[\cdot]_F$ are invertible.

[26]Unfortunately, in practice, *matrix equation* usually means *linear matrix equation*.

Finding bases for the kernel and image of a matrix

Theorem 2.171. ().
   row operations preserve the kernel
   column operations preserve the image

Theorem 2.172. (Basis for the kernel of a matrix).
   Let $\mathbf{A}$ be an $m \times n$ matrix with entries in a field K. Since row operations preserve the kernel, then $\ker(\mathrm{rref}(\mathbf{A})) = \ker(\mathbf{A})$.
   The following algorithm produces a basis for $\ker(\mathrm{rref}(\mathbf{A})) = \ker(\mathbf{A})$.

1. Since $\mathbf{x} \in \ker(\mathrm{rref}(\mathbf{A}))$ iff $\mathrm{rref}(\mathbf{A})\mathbf{x} = \mathbf{0}$, consider the system of equations corresponding to the linear matrix equation $\mathrm{rref}(\mathbf{A})\mathbf{x} = \mathbf{0}$. Let $x_{j_1}, ..., x_{j_p} \in \{x_1, ..., x_n\}$ be the pivot variables of $\mathrm{rref}(\mathbf{A})$ and let $x_{k_1}, ..., x_{k_q} \in \{x_1, ..., x_n\}$ be the free variables of $\mathrm{rref}(\mathbf{A})$.

2. Obtain an equivalent system of equations by solving each equation of the previous system for its pivot variable, so that the ith equation in the new system is $x_{j_i} = c_{i1}x_{k_1} + ... + c_{ip}x_{k_q}$ for some $c_{i1}, ..., c_{ip} \in K$.

3. Obtain another equivalent system by appending the q equations $x_{k_1} = x_{k_1}, ..., x_{k_q} = x_{k_q}$, which express the free variables in terms of themselves, to the previous system of equations.

4. Consider the vector equation that corresponds to the previous system of linear equations. The left side of this vector equation is $(x_1, ..., x_n)^\top$. The right side has ith entry $c_{i_1}x_{k_1} + ... + c_{i_p}x_{k_q}$ when i is the index of a pivot variable and ith entry x_i when i is the index of a free variable.

5. Rewrite the previous vector equation so that it is of the form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_{k_1}\mathbf{v}_1 + \ldots + x_{k_q}\mathbf{v}_q,$$

   for some $\mathbf{v}_1, ..., \mathbf{v}_q \in K^n$. In this equation, $\mathbf{x}$ is expressed as a weighted sum of vectors, where the weights are the free variables.

6. Since all of the previous systems of equations and vector equations are equivalent, we clearly have $\mathbf{x} \in \ker(\mathrm{rref}(\mathbf{A}))$ iff $\mathbf{x} \in \mathrm{span}(\{\mathbf{v}_1, ..., \mathbf{v}_q\})$. Thus $\ker(\mathbf{A}) = \ker(\mathrm{rref}(\mathbf{A})) = \mathrm{span}(\{\mathbf{v}_1, ..., \mathbf{v}_q\})$. In the below, we show that $\{\mathbf{v}_1, ..., \mathbf{v}_q\}$ is linearly independent and is thus a basis for $\ker(\mathrm{rref}(\mathbf{A})) = \ker(\mathbf{A})$.

Proof. We prove that the set $\{\mathbf{v}_1, ..., \mathbf{v}_q\}$ returned by the algorithm is linearly independent.
   The form of the system of equations in (4) implies that in (5), the vector $\mathbf{v}_j$ has an ith entry of δ_{ij} when i is restricted to be an index of a free variable. It follows from this constraint that $\mathbf{v}_1, ..., \mathbf{v}_q$ are linearly independent.
   To see this, consider the equation $c_1\mathbf{v}_1 + \ldots + c_q\mathbf{v}_q = 0$, where $c_1, ..., c_q \in K$. Since, when i is the index of a free variable, the ith entry of $c_j\mathbf{v}_j$ is $c_j\delta_{ij}$, then when i is the index of a free variable, the ith entry of the sum is $\sum_{j=1}^q c_j\delta_{ij} = c_i$. Thus the entries of the sum corresponding to free variables are $c_1, ..., c_q$. For the sum to be the zero vector, all entries of the sum must be zero, including all of $c_1, ..., c_q$. Thus $\mathbf{v}_1, ..., \mathbf{v}_q$ are linearly independent. $\square$

Example 2.173. (Basis for the kernel of a matrix).
   If $\mathbf{A} \in \mathbb{R}^{3 \times 5}$ is such that

$$\mathrm{rref}(\mathbf{A}) = \begin{pmatrix} 1 & 0 & 2 & 0 & 5 \\ 0 & 1 & -1 & 0 & -3 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix},$$

then $(x_1, x_2, x_3, x_4, x_5)^\top \in \ker(\mathbf{A})$ iff

$$\underbrace{\begin{pmatrix} 1 & 0 & 2 & 0 & 5 \\ 0 & 1 & -1 & 0 & -3 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}}_{\mathbf{A}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This linear matrix equation is equivalent to the following system of linear equations:

$$\begin{cases} x_1 + 2x_3 + 5x_5 = 0 \\ x_2 - x_3 - 3x_5 = 0 \qquad \text{for all } x_1, x_2, x_3, x_4, x_5 \in \mathbb{R}. \\ x_4 + 2x_5 = 0 \end{cases}$$

Solving for the pivot variables (x_1, x_2, x_4) in terms of the free variables (x_3, x_5), we obtain the equivalent system

$$\begin{cases} x_1 = -2x_3 - 5x_5 \\ x_2 = x_3 + 3x_5 \qquad \text{for all } x_1, x_2, x_3, x_4, x_5 \in \mathbb{R}. \\ x_4 = -2x_5 \end{cases}$$

Appending equations relating the free variables to themselves, we obtain the system

$$\begin{cases} x_1 = -2x_3 - 5x_5 \\ x_3 = x_3 \\ x_2 = x_3 + 3x_5 \qquad \text{for all } x_1, x_2, x_3, x_4, x_5 \in \mathbb{R}. \\ x_4 = -2x_5 \\ x_5 = x_5 \end{cases}$$

This system is equivalent to the vector equation

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -2x_3 - 5x_5 \\ x_3 + 3x_5 \\ x_3 \\ -2x_5 \\ x_5 \end{pmatrix} = x_3 \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -5 \\ 3 \\ 0 \\ -2 \\ 1 \end{pmatrix} \quad \text{for all } x_1, x_2, x_3, x_4, x_5 \in \mathbb{R}.$$

Thus,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \in \text{span} \left(\left\{ \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ 3 \\ 0 \\ -2 \\ 1 \end{pmatrix} \right\} \right).$$

The result of the previous theorem ensures us that

$$\left\{ \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ 3 \\ 0 \\ -2 \\ 1 \end{pmatrix} \right\}$$

is a basis for $\ker(\mathbf{A})$.

Theorem 2.174. (Basis for the image of a matrix).

Let $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{pmatrix}$ be a matrix, consider $\text{rref}(\mathbf{A})$, and let $x_{j_1}, ..., x_{j_p}$ be the pivot variables of $\text{rref}(\mathbf{A})$. The columns of $\mathbf{A}$ corresponding to the pivot columns of $\text{rref}(\mathbf{A})$ is a basis for $\text{im}(\mathbf{A})$. More precisely, $\{\mathbf{a}_{j_1}, ..., \mathbf{a}_{j_p}\}$ is a basis for $\text{im}(\mathbf{A})$.

Proof. First, we show that the pivot columns of $\text{rref}(\mathbf{A})$ are a basis for $\text{im}(\mathbf{A})$. Secondly, we show that the corresponding columns of $\mathbf{A}$ are a basis for $\text{im}(\mathbf{A})$.

1. (Pivot columns of $\text{rref}(\mathbf{A})$ are a basis for $\text{im}(\mathbf{A})$). Let $\mathbf{r}_j$ be the jth column of $\text{rref}(\mathbf{A})$, and let $\mathbf{r}_{j_1}, ..., \mathbf{r}_{j_p} \in \{\mathbf{r}_1, ..., \mathbf{r}_n\}$ be the pivot columns of $\text{rref}(\mathbf{A})$. Since $\text{rref}(\mathbf{A})$ is in RREF, then $\mathbf{r}_{j_\alpha}$ has in its ith row the entry δ_{ij_α}. Equivalently, $\mathbf{r}_{j_\alpha} = \hat{\mathbf{e}}_{j_\alpha}$. The pivot columns $\mathbf{r}_{j_1} = \hat{\mathbf{e}}_{j_1}, ..., \mathbf{r}_{j_p} = \hat{\mathbf{e}}_{j_p}$ are thus linearly independent.

We now show that the pivot columns span $\text{im}(\mathbf{A})$. Since $\text{rref}(\mathbf{A})$ is in RREF, then the columns of $\text{rref}(\mathbf{A})$ containing free variables, $\mathbf{r}_{k_1}, ..., \mathbf{r}_{k_q} \in \{\mathbf{r}_1, ..., \mathbf{r}_n\}$, are linear combinations of the pivot columns of $\text{rref}(\mathbf{A})$, so $R_\beta := \{\mathbf{r}_{j_1}, ..., \mathbf{r}_{j_p}, \mathbf{r}_{k_1}, ..., \mathbf{r}_{k_\beta}\}$ is linearly dependent for every $\beta \in [1, q]$. Since $\text{im}(\mathbf{A}) = \text{span}(\{\mathbf{r}_{j_1}, ..., \mathbf{r}_{j_p}, \mathbf{r}_{k_1}, ..., \mathbf{r}_{k_q}\}) = \text{span}(R_q)$, and since removing a vector from a linearly dependent set does not change the span of the set (recall Lemma 2.35), then $R_\beta = R_{\beta+1} - \{\mathbf{r}_{k_{\beta+1}}\}$ spans $\text{im}(\mathbf{A})$ for all $\beta \in [0, q]$. In particular, $R_0 = \{\mathbf{r}_{j_1}, ..., \mathbf{r}_{j_p}\}$, which is the set of pivot columns, spans $\text{im}(\mathbf{A})$.

Since $\mathbf{r}_{j_1}, ..., \mathbf{r}_{j_p}$ are linearly independent and span $\text{im}(\mathbf{A})$, they are a basis of $\text{im}(\mathbf{A})$.

2. (Columns of $\mathbf{A}$ corresponding to the pivot columns of $\text{rref}(\mathbf{A})$ are a basis of $\text{im}(\mathbf{A})$). Since row-reduction is an invertible linear operation, then there exists an invertible linear function $\mathbf{f}$ such that $\mathbf{f}(\mathbf{A}) = \text{rref}(\mathbf{A})$. We then have $\mathbf{A} = \mathbf{f}^{-1}(\text{rref}(\mathbf{A}))$, i.e., $\mathbf{a}_i = \mathbf{f}^{-1}(\mathbf{r}_i)$, and, in particular, $\{\mathbf{a}_{j_1}, ..., \mathbf{a}_{j_p}\} = \{\mathbf{f}^{-1}(\mathbf{r}_{j_1}), ..., \mathbf{f}^{-1}(\mathbf{r}_{j_p})\}$. Since invertible linear functions send bases to bases (recall Theorem 2.70), then $\{\mathbf{f}^{-1}(\mathbf{r}_{j_1}), ..., \mathbf{f}^{-1}(\mathbf{r}_{j_p})\} = \{\mathbf{a}_{j_1}, ..., \mathbf{a}_{j_p}\}$ is a basis of $\text{im}(\mathbf{A})$.

$\square$

Derivation 2.175. (Main dimension theorem with matrices).

Let $\mathbf{A}$ be a matrix with entries in a field K. The theorem about the basis for the kernel of a matrix showed that the number of vectors in a basis for $\text{ker}(\mathbf{A})$ is the number of free variables of $\text{rref}(\mathbf{A})$, and the theorem about the basis for the image of a matrix showed that the number of vectors in a basis for $\text{im}(\mathbf{A})$ is the number of pivot variables of $\text{rref}(\mathbf{A})$. Thus, $\dim(\text{ker}(\mathbf{A}))$ is the number of free variables of $\text{rref}(\mathbf{A})$ and $\dim(\text{im}(\mathbf{A}))$ is the number of pivot variables of $\text{rref}(\mathbf{A})$.

Recall the main dimension theorem (Theorem 2.65), which says that when V is a finite-dimensional vector space we have $\dim(\text{im}(\mathbf{A})) = \dim(V) - \dim(\text{ker}(\mathbf{A}))$. Applying this theorem, we see that if $\mathbf{A} \in K^{m \times n}$, then, since the function $\mathbf{v} \mapsto \mathbf{A}\mathbf{v}$ is a linear function $K^n \to K^m$, we have

$$(\text{number of pivot variables in } \text{rref}(\mathbf{A})) = n - (\text{number of free variables in } \text{rref}(\mathbf{A}))$$

.

Outline

(T) Theory

1. defn. linear equation

2. thm. $\{\text{solns}\} = \{\mathbf{v} + \mathbf{v}_0 \mid \mathbf{v}_0 \in \ker(\mathbf{f})\}$ for any $\mathbf{v} \in \{\text{solns}\}$

3. thm. number of solutions to a linear equation

(M) Motivating study of systems of linear equations

1. deriv. linear equations <-> linear matrix equations <-> systems of linear equations

2. Remark. (Using intuition about planes with systems of linear equations).

   Notice that a linear equation in n variables $v_1, ..., v_n$ is of the form $\mathbf{v} \cdot \mathbf{n} = 0$ for some $\mathbf{n} \in \mathbb{R}^n$ and $c \in K$.

   not all linear equations are of this form. some are of the form $\mathbf{v} \cdot \mathbf{n} = c$ for $c \neq 0$

   In general, the *orthogonal complement* to a vector subspace W of vector space V with inner product $\langle \cdot, \cdot \rangle$, is defined to be the vector subspace $W^\perp := \{\langle \mathbf{v}, \mathbf{w} \rangle = 0 \mid \mathbf{v} \in V \text{ and } \mathbf{w} \in W\}$ of vectors in V perpendicular to all vectors in W. An $(n-1)$-*dimensional plane in* $\mathbb{R}^n$ is then defined to be the the the orthogonal complement of $\text{span}(\mathbf{n})^\top = \{\mathbf{v} \cdot \mathbf{w} = 0 \mid \mathbf{v} \in \mathbb{R}^n \text{ and } \mathbf{w} \in \text{span}(\mathbf{n})\} = \{\mathbf{v} \cdot \mathbf{n} = 0 \mid \mathbf{v} \in V\}$ for some $\mathbf{n} \in \mathbb{R}^n$.

   Thus, a linear equation in n variables is the equation of an $(n-1)$-dimensional plane, and so a system of m linear equations in n variables can be interpreted to be an intersection of m many $(n-1)$-dimensional planes.

   Intuition about 2-dimensional planes in $\mathbb{R}^3$ can therefore be used to gain intuition about the number of solutions to a system of linear equations.

(S) Solving linear equations with augmented matrices

1. Deriv. Augmented matrix corresponding to a system of linear equations

2. Because adding equations, multiplying equations by a scalar, and swapping equations doesn't change a system of equations, performing the analogous row operations on the corresponding matrix doesn't change the system to which the augmented matrix correspond. The following algorithm takes advantage of this fact and uses it to put every matrix in a form from which it is easy to read the solution to the corresponding system of linear equations

3. Definition. (Gaussian elimination algorithm).

4. (Gaussian elimination produces matrices in RREF). After Gaussian elimination, an augmented matrix is in the following form:

   - All zero rows (rows equal to $\mathbf{0}^\top$) are at the bottom of the matrix.
   - Every nonzero row's leftmost nonzero entry is a 1. Such 1's are called "leading 1's", or "pivots".
   - Each column containing some row's leading 1 has zeros in all its other entries.

   This form is known as *reduced row echelon form*, or *RREF*.

   [Discuss how solutions to the corresponding system are easy to read off...]

   Variables whose corresponding columns have a pivot are called *pivot variables*. Notice that, after considering the system of equations corresponding to an augmented matrix in RREF, we can express each pivot variable as the evaluation of a (linear) function on the non-pivot variables:

   ...

   Thus, [we see the solution to the system is easy to read off.]

   Since the non-pivot variables are treated as independent variables, and since independent variables can be thought of as being allowed to freely vary, the non-pivot variables are called *free variables*.

5. Lemma about RREF.

   - Every row has at most 1 pivot and every column has at most 1 pivot. (Proof. (Rows). A row is either zero or nonzero. If it is zero, it has no pivots; if it is nonzero, it has one pivot. (Columns). A column is either zero or nonzero. If it is zero, it has no pivots. If it is nonzero, it either contains some row's leading 1 or doesn't. If it does, it contains one pivot. If it doesn't, it contains zero pivots.)

6. reading solutions to a linear equation from a RREF

7. RREF as linear function represented by product of elementary matrices

(F) Finding ker($\mathbf{A}$)

1. $\mathbf{A}$ is invertible (and thus has a trivial kernel) iff rref($\mathbf{A}$) $= \mathbf{I}$

   - ($\implies$). We prove the contrapositive, rref($\mathbf{A}$) $\neq \mathbf{I} \implies \mathbf{A}$ is not invertible. Suppose rref($\mathbf{A}$) $\neq \mathbf{I}$, and consider the augmented matrix (rref($\mathbf{A}$) | $\mathbf{0}$). Since rref($\mathbf{A}$) $\neq \mathbf{I}$, then for some i the set $\{j \mid j$ is a free column in the ith row of rref(($\mathbf{A}$ | $\mathbf{0}$))$\}$ is nonempty. The ith row of (rref($\mathbf{A}$) | $\mathbf{0}$) then corresponds to the equation $v_i + \sum_j c_j w_j = 0$. Since the w_j are linearly independent [???], then $v_i = -\sum_j w_j \neq 0$. This means that the solution $\mathbf{v} \in K^n$ to $\mathbf{A}\mathbf{v} = \mathbf{0}$ has a nonzero ith component, and is thus itself nonzero; $\mathbf{v} \neq \mathbf{0}$. The kernel of $\mathbf{A}$ is therefore nontrivial, so $\mathbf{A}$ is not invertible.

     ($\impliedby$). If rref($\mathbf{A}$) $= \mathbf{I}$ then $\mathbf{E}_k...\mathbf{E}_1\mathbf{A} = \mathbf{I}$ and thus $\mathbf{A}^{-1} = \mathbf{E}_1^{-1}...\mathbf{E}_k^{-1}$.

2. lemma. rref($\mathbf{A}$) has no pivots iff rref($\mathbf{A}$) $= \mathbf{0}$.

   - ($\implies$). It suffices to show that if a row has no pivots then it is $\mathbf{0}^\top$. Assume for contradiction that a row $\mathbf{r} \in K^{1 \times n}$ has no pivots and is not equal to $\mathbf{0}^\top$. Then $\mathbf{r}$ has some nonzero entries $r_1, ..., r_k$. Consider the leftmost such entry, r_j. This r_i is the leading entry of $\mathbf{r}$. Either $r_i = 1$ or $r_i \neq 1$. If $r_i = 1$, then $\mathbf{r}$ has a pivot and thus rref($\mathbf{A}$) has a pivot; contradiction. If $r_i \neq 1$, then $\mathbf{r}$ has a leading entry that is not equal to 1, so rref($\mathbf{A}$) is not in RREF; contradiction. ($\impliedby$). If rref($\mathbf{A}$) $= \mathbf{0}$ then there are no 1's in rref($\mathbf{A}$). Thus there are no leading 1's, i.e., no pivots.

3. (lemma for basis for kernel). row swaps preserve kernel and column swaps preserve image

4. basis for kernel

5. basis for image

6. $\dim(V)$ is number of columns, dimension of image is number of pivots, dimension of kernel is number of free variables, so n is number of pivots plus number of free variables

misc.

- test for linear independence: $\mathbf{v}_1, ..., \mathbf{v}_k$ linearly independent iff the kernel of $\begin{pmatrix} \mathbf{v}_1 & ... & \mathbf{v}_k \end{pmatrix}$ is trivial. i.e. if its RREF is $\mathbf{I}$

- row operations don't change the kernel and column operations don't change the image

2.6 Eigenvectors and eigenvalues

[This section not required for rest of book]

Definition 2.176. (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 $\lambda \in K$ such that $\mathbf{f}(\mathbf{v}) = \lambda \mathbf{v}$. Iff $\mathbf{v}$ is indeed an eigenvector of $\mathbf{f}$, then the $\lambda \in K$ for which $\mathbf{f}(\mathbf{v}) = \lambda \mathbf{v}$ 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 2.177. ("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 in mind, as we will see that, in some cases, a linear function is completely specified if its "characteristic vectors" and corresponding "characteristic values" are known.

Theorem 2.178. Let V be a vector space. A linear function $\mathbf{f} : V \to V$ is invertible iff $\mathbf{0}$ is the only eigenvector of $\mathbf{f}$ whose eigenvalue is 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. $\square$

Remark 2.179. (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 in their definition of "eigenvector" so that the condition "$\mathbf{0}$ is the only eigenvector of $\mathbf{f}$ whose eigenvalue is 0" is equivalent to the condition "0 is not an eigenvalue of $\mathbf{f}$". The later is easier to say than the former. This convention is a bit too contrived, though, so we will not use it.

Theorem 2.180. (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 2.181. (Eigenvectors).

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 an eigenvector of $\mathbf{f}$ with eigenvalue λ iff

$$\mathbf{f}(\mathbf{v}) = \lambda \mathbf{v}$$
$$\Longleftrightarrow \mathbf{f}(\mathbf{v}) - \lambda \mathbf{v} = \mathbf{0}$$
$$\Longleftrightarrow \mathbf{f}(\mathbf{v}) - \lambda \mathbf{I}(\mathbf{v}) = \mathbf{0}$$
$$\Longleftrightarrow (\mathbf{f} - \lambda \mathbf{I})(\mathbf{v}) = \mathbf{0}$$
$$\Longleftrightarrow \mathbf{v} \in \ker(\mathbf{f} - \lambda \mathbf{I}).$$

Thus,

$$\boxed{\mathbf{v} \text{ is an eigenvector of } \mathbf{f} \text{ with eigenvalue } \lambda \text{ iff } \mathbf{v} \in \ker(\mathbf{f} - \lambda \mathbf{I})}$$

Derivation 2.182. (Eigenvalues).

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

A scalar $\lambda \in K$ is an eigenvalue of $\mathbf{f}$ if and only if $\ker(\mathbf{f} - \lambda \mathbf{I})$ is nonempty, which is the case if and only if $\mathbf{f} - \lambda \mathbf{I}$ is not invertible. This is equivalent to the condition $\det(\mathbf{f} - \lambda \mathbf{I}) = 0$. Thus,

$$\boxed{\lambda \text{ is an eigenvalue of } \mathbf{f} \text{ iff } \det(\mathbf{f} - \lambda \mathbf{I}) = 0}$$

Part I

Multilinear algebra and tensors

3

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.

Notice, the case $k = 1$ is consistent with our definition in Theorem 2.58 of $\mathcal{L}(V_1 \to W)$ as being the set of linear functions $V_1 \to W$.

Proof. The proof that $\mathcal{L}(V_1 \times ... \times V_k \to W)$ is indeed a vector space is left as an exercise. $\qquad\square$

Elements of a vector space can be considered to be "linear elements" because their decompositions relative to a basis are respected by linear functions (see Definition 2.50). 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. [Lee, p. 307] (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:

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

and

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

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 3.39). 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 2.66 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 2.71).

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. $\square$

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. $\square$

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.

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.

---

[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 $\sum_{i,j} T_{i,j} \mathbf{e}_i \otimes \mathbf{f}_j = \mathbf{0} = \mathbf{v} \otimes \mathbf{0}$ for any $\mathbf{v} \in V$.

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. Other consequences of this are explored in Derivation 4.27). 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. $\qquad\square$

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 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 V. We can see that the two key ideas (the first being thinking of linear functions as vectors and the second being "multilinear elements") are represented in this theorem with formal notation: the theorem includes 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.

We first need a lemma.

Lemma 3.12. (Standard basis matrices).

Let K be a field, and consider the standard bases $\hat{\mathfrak{F}} = \{\hat{\mathfrak{f}}_1, ..., \hat{\mathfrak{f}}_m\}$ and $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ for K^m and K^n. The $m \times n$ matrix whose only nonzero entry is a 1 in the *ith* row and *j*th column is equal to $\hat{\mathfrak{f}}_i \hat{\mathfrak{e}}_j^\top$. For example, if $m = 4$ and $n = 3$, we have

$$\hat{\mathfrak{f}}_3 \hat{\mathfrak{e}}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Now we begin our investigation.

Derivation 3.13. (Decomposition of a linear function).

Let V and W be n- and m-dimensional vector spaces over a field K, and consider the matrix $[\mathbf{f}(E)]_F$ of $\mathbf{f}$ relative to E and F.

Let a_{ij} be the ij entry of $[\mathbf{f}(E)]_F$. Making use of the lemma, we have

$$[\mathbf{f}(E)]_F = \sum_{\substack{i \in \{1, ..., n\} \\ j \in \{1, ..., m\}}} a_{ij} \hat{\mathfrak{f}}_i \hat{\mathfrak{e}}_j^\top.$$

Applying the linear isomorphism $\mathbf{F}^{-1}$ between $m \times n$ matrices and linear functions $V \to W$ from Theorem 2.113 to this equation causes

- $[\mathbf{f}(E)]_F$ to be replaced with $\mathbf{f}$

- matrix multiplication to be replaced by function composition $\circ$

- $\hat{\mathfrak{e}}_j^\top$ to be replaced the element $\phi^{\hat{\mathfrak{e}}_j}$ of $\mathcal{L}(V \to K)$ whose matrix relative to E is $\hat{\mathfrak{e}}_j^\top$ (so $\phi^{\mathbf{e}_j}(\mathbf{v}) = \hat{\mathfrak{e}}_j^\top [\mathbf{v}]_E$)

- $\hat{\mathfrak{f}}_i$ to be replaced by the element $\widetilde{\hat{\mathfrak{f}}_i}$ of $\mathcal{L}(K \to W)$ whose standard matrix is $\hat{\mathfrak{f}}_i$ (so $\widetilde{\hat{\mathfrak{f}}_i}(c) = c\hat{\mathfrak{f}}_i$)

So, we have

$$\mathbf{f} = \sum_{\substack{i \in \{1, ..., n\} \\ j \in \{1, ..., m\}}} a_{ij} \widetilde{\hat{\mathfrak{f}}_i} \circ \phi^{\mathbf{e}_i},$$

where $\phi^{\mathbf{e}_i} \in \mathcal{L}(V \to K)$ and $\widetilde{\hat{\mathfrak{f}}_i} \in \mathcal{L}(K \to W)$.

Since we have seen $\mathcal{L}(V \to K)$ and $\mathcal{L}(K \to W)$ have fundamental roles in the above decomposition, we make the following definitions.

Definition 3.14. (Dual vector space).
Let V be a vector space over a field K. The *dual vector space* V^* to V is the vector space $V^* := \mathcal{L}(V \to K)$.

Definition 3.15. (Natural identification of W with $\mathcal{L}(K \to W)$).
Let W be a vector space over a field K. We define a natural linear isomorphism $\sim: W \to \mathcal{L}(K \to W)$ by $\sim(\mathbf{w}) = (c \mapsto c\mathbf{w})$, and use the notation $\widetilde{\mathbf{w}} :=\sim(\mathbf{w})$. That is, we have $\widetilde{\mathbf{w}}(c) = c\mathbf{w}$.

Using these definitions together with the result from the end of the above derivation yields the following theorem.

Theorem 3.16. (Decomposition of a linear function).
Let V and W be n- and m-dimensional vector spaces. Any linear function $\mathbf{f} : V \to W$ can be decomposed as

$$\mathbf{f} = \sum_{\substack{i \in \{1,\ldots,n\} \\ j \in \{1,\ldots,m\}}} a_{ij} \widetilde{\mathbf{w}}_i \circ \phi_j \text{ for some } a_{ij} \in K,$$

where $\phi_j \in V^*$, $\mathbf{w}_i \in W$, and where we recall from the previous definition that $\widetilde{\mathbf{w}}_i(c) = c\mathbf{w}_i$.

Proof. See the previous derivation. $\qquad\square$

One final "big leap" will complete our discovery. Recall, our original goal was to show $\mathcal{L}(V \to W) \cong W \otimes V^*$. We've seen that dual spaces (i.e. V^*) have become involved. Tensor product spaces (i.e. $\otimes$) have not, yet, but will now.

Derivation 3.17. ($\mathcal{L}(V \to W) \cong W \otimes V^*$ naturally).
We know from the previous theorem that $\mathbf{f} \in \mathcal{L}(V \to W)$ decomposes as a sum of linear functions of the form $\widetilde{\mathbf{w}} \circ \phi$, where $\phi \in V^*$ and $\mathbf{w} \in W$. The idea is to take advantage of this decomposition and define a linear isomorphism $\mathbf{F} : \mathcal{L}(V \to W) \to W \otimes V^*$ that sends $(\widetilde{\mathbf{w}} \circ \phi) \in \mathcal{L}(V \to W)$ to the elementary tensor $\mathbf{w} \otimes \phi \in W \otimes V^*$:

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

We need to show that $\mathbf{F}$ is a linear bijection. Ultimately, this will be the case because $\circ$ is a bilinear map, because $\mathbf{w} \mapsto \widetilde{\mathbf{w}}$ is a linear isomorphism, and because $\otimes$ appears to be bilinear.

First, we show $\mathbf{F}$ is linear on *elementary compositions*, which we define to be linear functions of the form $(\mathbf{w} \circ \phi) \in \mathcal{L}(V \to W)$, where $\phi \in V^*$ and $\mathbf{w} \in W$. These elementary compositions are similar to elementary tensors in the sense that they are not fundamentally a linear combination of two or more other elements of the form $\mathbf{w} \circ \phi$.

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

$$\widetilde{\mathbf{w}}_1 \circ \phi + \widetilde{\mathbf{w}}_2 \circ \phi \overset{\mathbf{F}}{\longmapsto} \mathbf{w}_1 \otimes \phi + \mathbf{w}_2 \otimes \phi$$
$$\widetilde{\mathbf{w}} \circ \phi_1 + \widetilde{\mathbf{w}} \circ \phi_2 \overset{\mathbf{F}}{\longmapsto} \mathbf{w} \otimes \phi_1 + \mathbf{w} \otimes \phi_2$$
$$c(\widetilde{\mathbf{w}} \circ \phi) \overset{\mathbf{F}}{\longmapsto} c(\mathbf{w} \otimes \phi),$$

for all $\mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in W$, $\phi, \phi_1, \phi_2 \in V^*$, and $c \in K$.

The above is achieved because $\circ$ is bilinear[2], because $\mathbf{w} \mapsto \widetilde{\mathbf{w}}$ is a linear isomorphism, and because $\otimes$ is bilinear:

$$\widetilde{\mathbf{w}}_1 \circ \phi + \widetilde{\mathbf{w}}_2 \circ \phi = (\widetilde{\mathbf{w}}_1 + \widetilde{\mathbf{w}}_2) \circ \phi = (\widetilde{\mathbf{w}_1 + \mathbf{w}_2}) \circ \phi \overset{\mathbf{F}}{\longmapsto} (\mathbf{w}_1 + \mathbf{w}_2) \otimes \phi = \mathbf{w}_1 \otimes \phi + \mathbf{w}_2 \otimes \phi$$

$$\widetilde{\mathbf{w}} \circ \phi_1 + \widetilde{\mathbf{w}} \circ \phi_2 = \widetilde{\mathbf{w}} \circ (\phi_1 + \phi_2) \overset{\mathbf{F}}{\longmapsto} \mathbf{w} \otimes (\phi_1 + \phi_2) = \mathbf{w} \otimes \phi_1 + \mathbf{w} \otimes \phi_2$$

$$c(\widetilde{\mathbf{w}} \circ \phi) = (c\widetilde{\mathbf{w}}) \circ \phi \overset{\mathbf{F}}{\longmapsto} (c\mathbf{w}) \otimes \phi = c(\mathbf{w} \otimes \phi).$$

So, $\mathbf{F}$ is linear on elementary compositions. We now *impose* that it be 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 "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 onto because any nonelementary tensor coorresponds to a "nonelementary composition", i.e., a linear combination of elementary compositions. $\mathbf{F}$ is one-to-one because it is one-to-one 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 3.18. ($\mathcal{L}(V \to W) \cong W \otimes V^*$ 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 2.71).

Proof. See the previous derivation. $\qquad \square$

Remark 3.19. (The two key ideas).
Now that we have gone through the derivation, we can specifically see how the two key ideas- 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 $\otimes$ to think of these elementary compositions as "multilinear elements" (speficially, we thought of them as "bilinear elements").

When we defined inner products (Definition 2.155), we gave motivation for why inner products are called "inner products", and said that we would later explain the relevance of "outer products". Here is that explanation.

Definition 3.20. (Outer products).
Let K be a field. We define the *outer product of* $\mathbf{w} \in K^m$ *and* $\mathbf{v} \in K^n$ to be the matrix $\mathbf{w}\mathbf{v}^\top \in K^{m \times n}$.

Now let V and W be finite-dimensional vector spaces. Throughout this chapter, we've seen that matrices of the form $\mathbf{w}\mathbf{v}^\top$ are identified with linear functions of the form $\mathbf{w} \circ \phi \in \mathcal{L}(V \to W)$, where $\phi \in V^*$, and that those linear functions are identified with tensors of the form $\mathbf{w} \otimes \phi \in W \otimes V^*$. Thus the significance of outer products is that they are fundamentally at play whenever tensors are involved.

Since when $\mathbf{v} \in K^n$ and $\mathbf{w} \in K^m$, the matrix $\mathbf{w}\mathbf{v}^\top$ can be identified with the tensor $\mathbf{w} \otimes \mathbf{v} \in W \otimes V$, then the tensor product $\otimes$ is unfortunately sometimes referred to as "the outer product". Some authors even take advantage of the fact that for $\mathbf{v} \in K^n$ and $\mathbf{w} \in K^n$ the function $(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{w}\mathbf{v}^\top$ is bilinear, and define the tensor product on K^n so that it actually is the outer product: "$\mathbf{w} \otimes \mathbf{v} := \mathbf{w}\mathbf{v}^\top$".

Note: there is no sort of fundamental duality between inner products (Definition 2.155) and outer products. The two topics are relatively unrelated.

There is one last definition whose name sounds like it is related to inner products and outer products, and that is the *exterior product of vectors.* We discuss this in Remark **??**.

---

[2]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, $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C})$

3.3   Dual vector spaces

Recall that dual vector spaces are crucial to the concept of a (p, q) tensor because they allow us to think of linear functions as vectors. We now restate the definition of a dual space and make some additional remarks.

Definition 3.21. (Dual vector space).

Let V be a vector space over a field K. The *dual vector space* V^* to V is the vector space $V^* := \mathcal{L}(V \to K)$.

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, which we have not yet defined).

Definition 3.22. (Covariance and contravariance).

Let V be a vector space. For reasons that will be explained later, in Remark 4.23, 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, co-varaint 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 4.23.

Bases for dual spaces

Derivation 3.23. (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 decomposing the matrix corresponding to an element of V^*.

Specifically, any element $\phi \in V^* = \mathcal{L}(V \to K)$ is represented by its matrix $\phi(E)$ relative to E (recall Derivation 2.103),

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

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

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

Applying the linear isomorphism $\mathbf{F}^{-1}$ between $1 \times n$ matrices and linear functions $V \to K$ from Theorem 2.113 to this equation causes

- $\phi(E)$ to be replaced with ϕ

- matrix multiplication to be replaced by function composition $\circ$

- $\hat{\mathbf{\epsilon}}_i^\top$ to be replaced the element $\phi^{\hat{\mathbf{\epsilon}}_i}$ of $\mathcal{L}(V \to K)$ whose matrix relative to E is $\hat{\mathbf{\epsilon}}_i^\top$; that is, $\phi^{\mathbf{e}_i}(\mathbf{v}) = \hat{\mathbf{\epsilon}}_i^\top[\mathbf{v}]_E$

Therefore

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

where $\phi^{\mathbf{e}_i}(\mathbf{v}) = \hat{\mathbf{\epsilon}}_i^\top[\mathbf{v}]_E$.

We see that any $\phi \in V^*$ is a linear combination of $\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}$, so $\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}$ span V^*. Since $\hat{\mathbf{\epsilon}}_1^\top, ..., \hat{\mathbf{\epsilon}}_n^\top$ are linearly independent, and as $\mathbf{F}^{-1}$ is a linear isomorphism, then $\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}$ are also linearly independent (recall Theorem 2.64). 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*.

Theorem 3.24. (Dimension of dual space to a finite-dimensional vector space).

If V is a finite-dimensional vector space, then $\dim(V^*) = \dim(V)$.

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

Theorem 3.25. (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:

1. $\phi^{\mathbf{e}_i}$ is the element of V^* whose matrix $\phi^{\mathbf{e}_i}(E)$ relative to E is $\phi^{\mathbf{e}_i}(E) = \hat{\mathfrak{e}}_i^\top$. 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 $\mathbf{v} = \mathbf{e}_j$. For the reverse direction, we have $\phi^{\mathbf{e}_i}(\mathbf{v}) = \sum_{j=1}^n ([\mathbf{v}]_E)^j \phi(\mathbf{e}_j) = \sum_{j=1}^n ([\mathbf{v}]_E)^j \delta_{ij} = ([\mathbf{v}]_E)^i$. $\qquad\square$

Remark 3.26. (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 2.71) because it depends on the basis E for V. Additionally, the map $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ is onto but not one-to-one when V is finite-dimensional, because when V is infinite-dimensional the cardinality of V^* is strictly greater than the cardinality of V.

Remark 3.27. (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 3.23 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^*.

The double dual

The following theorem explains why $V^* = \mathcal{L}(V \to K)$ is called the "dual vector space" to V.

Theorem 3.28. ($V \cong V^{**}$ naturally).

Let V be a finite-dimensional vector space. The function $\mathbf{v} \mapsto \Phi_{\mathbf{v}}$, where $\Phi_{\mathbf{v}} : V^* \to K$ is the element of $(V^*)^* = V^{**}$ defined by $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$, is a natural linear isomorphism $V \to V^{**}$. Thus $V \cong V^{**}$ naturally.

This theorem tells us that not only is V^* the dual space to V, but that, since $(V^*)^* \cong V$ naturally, it is also true in some sense that V is the dual space to 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 3.24 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 (recall Theorem 2.68).

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. $\qquad\square$

For more on this theory of duality, see Derivation 11.1 in the appendix.

Corresponding elements of dual spaces induced by bases

Definition 3.29. (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 3.30. ($\Phi_{\phi^{\mathbf{v}}} = \Phi_{\mathbf{v}}$).
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 3.28).

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 3.25, we have $([\mathbf{v}]_E)^i = \phi^{\mathbf{e}_i}(\mathbf{v})$ and $\Phi_{\phi^{\mathbf{e}_i}}(\phi) = ([\phi]_{E^*})^i$, so $\sum_{i=1}^n \Big(([\mathbf{v}]_E)^i \Phi_{\phi^{\mathbf{e}_i}}(\phi) \Big) = \sum_{i=1}^n \Big(\phi^{\mathbf{e}_i}(\mathbf{v})([\phi]_{E^*})^i \Big) = \sum_{i=1}^n \Big(([\phi]_{E^*})^i \phi^{\mathbf{e}_i}(\mathbf{v}) \Big) = \Big(\sum_{i=1}^n ([\phi]_{E^*})^i \phi^{\mathbf{e}_i} \Big)(\mathbf{v}) = \phi(\mathbf{v}) = \Phi_{\mathbf{v}}(\phi).$ $\square$

Theorem 3.31. ($[\mathbf{v}]_E = [\phi^{\mathbf{v}}]_{E^*}$).
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 because $\mathbf{F}(\mathbf{e}_i) = \phi^{\mathbf{e}_i}$. A more explicit check is left to the reader. $\square$

Remark 3.32. (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 dual transformation

Definition 3.33. (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 3.34. (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 from Derivation 2.103 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 $[\mathbf{f}^*(F^*)]_{E^*} = [\mathbf{f}(E)]_F^\top$.

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 Theorem 4.16, we have $([\mathbf{f}^*(\delta^j)]_{E^*})_i = \mathbf{f}^*(\delta^j)(\mathbf{e}_i)$. We have $\mathbf{f}^*(\delta^j)(\mathbf{e}_i) = (\delta^j \circ \mathbf{f})(\mathbf{e}_i) = \delta^j(\mathbf{f}(\mathbf{e}_i)) = ([\mathbf{f}(\mathbf{e}_i)]_F)_j$, which is the ji entry of $[\mathbf{f}(E)]_F$, i.e., the ij entry of $[\mathbf{f}(E)]_F^\top$. $\square$

Remark 3.35. (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 3.33, 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 3.36. (Dual of a composition).
Let V, W, and Y be finite-dimensional vector spaces, and let $\mathbf{f} : V \to W$ and $\mathbf{g} : W \to Y$ be linear functions. Then $(\mathbf{g} \circ \mathbf{f})^* = \mathbf{f}^* \circ \mathbf{g}^*$.
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. For all $\psi \in Y^*$ we have $(\mathbf{g} \circ \mathbf{f})^*(\psi) = \psi \circ (\mathbf{g} \circ \mathbf{f}) = (\psi \circ \mathbf{g}) \circ \mathbf{f} = \mathbf{g}^*(\psi) \circ \mathbf{f}$, where $\mathbf{g}^*(\psi) \in V^*$. Then for all $\psi \in Y^*$ we have $\mathbf{g}^*(\psi) \circ \mathbf{f} = \mathbf{f}^*(\mathbf{g}^*(\psi)) = (\mathbf{f}^* \circ \mathbf{g}^*)(\psi)$. $\square$

3.4 (p, q) tensors

Before our detour into the world of dual spaces, we derived Theorem 3.18, and learned that when V and W are finite-dimensional vector spaces, there is a natural isomorphism $\mathcal{L}(V \to W) \cong W \otimes V^*$. 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 3.37. (Tensor space on a vector space).

Let V be a vector space. We define a *tensor space on V*, or more colloquially 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[3, 4]:

- The type of $V \otimes V^* \otimes V$ as a tensor space on V is ${}^1{}_1{}^1$.

- The type of $V^* \otimes (V)^{\otimes 3}$ as a tensor space on V is ${}_1{}^3$.

- The type of the tensor space $(V)^{\otimes 2} \otimes V^* \otimes V$ as a tensor space on V is ${}^2{}_1{}^1$.

(Superscripts coorrespond to V and subscripts coorrespond to V^*).

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 on V of which the tensor is a member.

Definition 3.38. (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 in K 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 $S_{i_1}{}^{i_2}{}_{i_3} \in K$ 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 3.39. ((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 $T_{p,q}(V) := V^{\otimes p} \otimes (V^*)^{\otimes q}$.

Recall, the below theorem is what motivated our definition of tensors in the first place.

Theorem 3.40. ($(1, 1)$ tensors are naturally identified with linear functions).

Since $T_{1,1}(V) = V \otimes V^*$ and $V \otimes V^* \cong \mathcal{L}(V \to V)$ naturally, then $T_{1,1}(V) \cong \mathcal{L}(V \to V)$ naturally.

---

[3]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.

[4]This convention goes against Definition 3.22, 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.

Theorem 3.41. (Coordinates of (p, q) tensors).

Let V be a vector space over a field K, let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be a basis for V, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be a basis for V^*. Since (p, q) tensors are of type ^p_q, then the coordinates of $\mathbf{T} \in T_{p,q}(V)$ relative to E and E^* are of the form $T^{i_1...i_p}{}_{j_1...j_q} \in K$, with all of the superscripts occurring before all of the subscripts. Using the coordinates to express $\mathbf{T}$, we have

$$\mathbf{T} = \sum_{\substack{i_1,...,i_p \in \{1,...,n\} \\ j_1,...,j_q \in \{1,...,m\}}} T^{i_1...i_p}{}_{j_1...j_q} \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_k} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q}.$$

Example 3.42. (Linear function corresponding to a $(1, 1)$ tensor).

Let V be a vector space, let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be a basis for V, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be a basis for V^*. Any $(1, 1)$ tensor on V, $\mathbf{T} \in T_{1,1}(V)$, can be expressed as

$$\mathbf{T} = \sum_{i,j \in \{1,...,n\}} T^i{}_j \mathbf{e}_i \otimes \epsilon^j.$$

Recall from Derivation 3.17 that $(1, 1)$ tensors are naturally identified with elements of $\mathcal{L}(V \to V)$ via the linear isomorphism $\mathbf{G} : T_{1,1}(V) = V \otimes V^* \to \mathcal{L}(V \to V)$ defined on elementary tensors by $\mathbf{G}(\mathbf{e}_i \otimes \epsilon^j) := \widetilde{\mathbf{e}}_i \circ \epsilon^j$, where $\widetilde{\mathbf{e}}_i \in \mathcal{L}(K \to V)$ is defined by $\widetilde{\mathbf{e}}_i(c) = c\mathbf{e}_i$. Thus, the corresponding element of $\mathcal{L}(V \to V)$ is

$$\mathbf{G}(\mathbf{T}) = \sum_{i,j \in \{1,...,n\}} T^i{}_j \widetilde{\mathbf{e}}_i \circ \epsilon^j.$$

Remark 3.43. (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 to denote coordinates of a $(2, 0)$ tensor, $\delta^i{}_j$ is used to denote coordinates of a $(1, 1)$ tensor, and δ_{ij} is used to denote 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.

Definition 3.44. (Valence and order of a (p, q) tensor).

The *valence* of a (p, q) tensor is the tuple (p, q). The *order* of a (p, q) tensor is $p + q$.

Theorem 3.45. (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

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

Importantly, application of the first and third line, together with 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

$$T_{p,q}(V) = V^{\otimes p} \otimes (V^*)^{\otimes q} \cong \mathcal{L}((V^*)^{\times p} \times V^{\times q} \to K)$$

Proof. The first line in the first box is Theorem 3.10, and the second line in the first box is Theorem 3.18. 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 3.46. (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:

$$
\begin{array}{ccc}
\text{multilinear element} & & \\
\text{(element of a tensor product space)} & \longleftrightarrow & \text{multilinear function} \\
\updownarrow & & \updownarrow \\
\text{linear element} & & \\
\text{(element of a vector space)} & \longleftrightarrow & \text{linear function}
\end{array}
$$

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.

Remark 3.47. (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 4.25). 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 3.45), 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 "order-n tensor" on V to be[5] a linear map that sends an order-$(n-1)$ tensor to a vector in V, where an order-1 tensor is defined to be a vector in V. Thus, an order-2 tensor is an element of $\mathcal{L}(V \to V) \cong V \otimes V^*$, and an order-3 tensor is an element of $\mathcal{L}(\mathcal{L}(V \to V) \to V) \cong \mathcal{L}(V \otimes V^* \to V) \cong V \otimes V^* \otimes V^*$. We can use induction to prove that an n-order tensor of this definition is identified with an element of $V \otimes (V^*)^{n-1} = T_{1,n-1}(V)$.

Theorem 3.48. (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

$$
V \otimes K \cong V
$$
$$
V \otimes W \cong W \otimes V.
$$

The proof of this theorem is left as an exercise.

---

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

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 4.16, 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.

4.1   Bilinear forms

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 3.45, we have $\mathcal{L}(V^{\times k} \to K) \cong \mathcal{L}(V^{\otimes k} \to K) = (V^{\otimes k})^* \cong (V^*)^{\otimes k} = T_k^0(V)$. Therefore a linear k-form is naturally identified with a $(0, k)$ tensor.

Definition 4.3. (The musical functions).

Let V and W be finite-dimensional vector spaces. If B is a bilinear form on V and W, then there are linear functions $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ defined[2] by $\flat_1(\mathbf{v}) := B(\mathbf{v}, \cdot)$ and $\flat_2(\mathbf{w}) := B(\cdot, \mathbf{w})$. These are called the *musical functions (induced by B)*. We denote $\mathbf{v}^{\flat_1} := \flat_1(\mathbf{v})$ and $\mathbf{w}^{\flat_2} := \flat_2(\mathbf{w})$.

Theorem 4.4. (The musical isomorphisms).

Let V and W be finite-dimensional vector spaces and let B be a bilinear form on V and W. The musical functions $\flat_1$ and $\flat_2$ are linear isomorphisms iff the following equivalent conditions hold:

1. $\flat_1$ and $\flat_2$ are one-to-one

2. $B(\mathbf{v}, \mathbf{w}) = \mathbf{0} \iff \mathbf{v} = \mathbf{0}$ or $\mathbf{w} = \mathbf{0}$ for all $\mathbf{v} \in V, \mathbf{w} \in W$

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 $\sharp_2 = \flat_2^{-1} : V^* \to W$.

Note that the musical isomorphisms are natural isomorphisms because they do not depend on a choice of basis (see Definition 2.71).

---

[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 $\mathbf{v} \mapsto B(\mathbf{v}, \mathbf{w})$.

Proof.

1. ($\flat_1$ and $\flat_2$ $\implies$ $\flat_1$ and $\flat_2$ are linear isomorphisms). Assume $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ are one-to-one. Then $\dim(V) \leq \dim(W^*) = \dim(W)$ and $\dim(W) \leq \dim(V^*) = \dim(V)$, so $\dim(V) = \dim(W)$. Since $\flat_1$ and $\flat_2$ one-to-one linear functions between finite-dimensional vectors paces of the same dimension, then $\flat_1$ and $\flat_2$ are linear isomorphisms (recall Theorem 2.68).

   ($\flat_1$ and $\flat_2$ are linear isomorphisms $\implies$ $\flat_1$ and $\flat_2$ are one-to-one). This is trivial.

2. We have: ($\flat_1$ is one-to-one) $\iff$ ($\mathbf{v}^{\flat_1} = \mathbf{0} \implies \mathbf{v} = \mathbf{0}$) $\iff$ (($\mathbf{v}^{\flat_1}(\mathbf{w}) = \mathbf{0}$ for all $\mathbf{w}$) $\implies$ ($\mathbf{v} = \mathbf{0}$)) $\iff$ $B(\mathbf{v}, \mathbf{w}) = \mathbf{0} \implies \mathbf{v} = \mathbf{0}$. Similarly, ($\flat_2$ is one-to-one) $\iff$ ($B(\mathbf{v}, \mathbf{w}) = \mathbf{0} \implies \mathbf{w} = \mathbf{0}$). The fact (2) is equivalent to these two results.

$\square$

The second condition of the previous theorem describes bilinear forms that generate musical isomorphisms and therefore motivates a new definition.

Definition 4.5. (Nondegenerate bilinear form).
Let V and W be finite-dimensional vector spaces. A bilinear form B on V and W is called *nondegenerate* iff ($B(\mathbf{v}, \mathbf{w}) = \mathbf{0} \iff \mathbf{v} = \mathbf{0}$ or $\mathbf{w} = \mathbf{0}$) for all $\mathbf{v} \in V, \mathbf{w} \in W$.
We saw above that a bilinear form generates musical isomorphisms iff it is nondegenerate.

Theorem 4.6. (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 $\widetilde{\flat}_1 : V^* \to W^{**}$ and $\widetilde{\flat}_2 : W^* \to V^{**}$ defined by $\phi^{\widetilde{\flat}_1}(\psi) = \widetilde{B}(\psi, \phi)$ and $\psi^{\widetilde{\flat}_2}(\phi) = \widetilde{B}(\psi, \phi)$.

Proof. The proof that $\widetilde{B}$ is actually a nondegenerate bilinear form is left as an exercise. $\square$

Lemma 4.7. Let V and W be finite-dimensional vector spaces with bases $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_m\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_m}\}$ be the bases for V^* and W^* induced by E and F. Let B be a nondegenerate bilinear form on V and W, and let $\widetilde{B}$ be the induced bilinear form on W^* and V^*. Then

$$B(\mathbf{v}, \mathbf{w}) = [\mathbf{v}]_E^\top \mathbf{B}[\mathbf{w}]_F = [\mathbf{w}]_F^\top \mathbf{B}^\top [\mathbf{v}]_E, \text{ where } \mathbf{B} := \left(B(\mathbf{e}_i, \mathbf{e}_j) \right)$$

$$\widetilde{B}(\psi, \phi) = [\psi]_{F^*}^\top \widetilde{\mathbf{B}}[\phi]_{E^*} = [\phi]_{E^*}^\top \widetilde{\mathbf{B}}^\top [\psi]_{F^*}, \text{ where } \widetilde{\mathbf{B}} := \left(\widetilde{B}(\psi^{\mathbf{f}_i}, \phi^{\mathbf{e}_j}) \right).$$

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.

We first consider the special case in which the vector spaces are $V = K^n$ and $W = K^m$ and where the bases are the standard bases for K^n and K^n, $E = \hat{\mathfrak{E}}$ and $F = \hat{\mathfrak{F}}$. In this special case, we have $B(\mathbf{v}, \mathbf{w}) = \sum_{ij}([\mathbf{v}]_{\hat{\mathfrak{E}}})^i([\mathbf{w}]_{\hat{\mathfrak{E}}})^j B_{ij} = \sum_i ([\mathbf{v}]_{\hat{\mathfrak{E}}})^i \sum_j B_{ij}([\mathbf{w}]_{\hat{\mathfrak{E}}})^j = \sum_i ([\mathbf{v}]_{\hat{\mathfrak{E}}})^i([\mathbf{Bw}]_{\hat{\mathfrak{F}}})^i = [\mathbf{v}]_{\hat{\mathfrak{E}}} \cdot [\mathbf{Bw}]_{\hat{\mathfrak{F}}} = [\mathbf{v}]_{\hat{\mathfrak{E}}}^\top \mathbf{B}[\mathbf{w}]_{\hat{\mathfrak{F}}}$, 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 $\mathbf{C} = (C(\mathbf{e}_i, \mathbf{f}_j)) = (B([\mathbf{e}_i]_E, [\mathbf{f}_i]_F)) = (B(\hat{\mathbf{e}}_i, \hat{\mathbf{f}}_i)) = \mathbf{B}$, 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. $\square$

Theorem 4.8. (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

$$[\flat_1(E)]_{F^*} = \mathbf{B}^\top \text{ and } [\flat_2(F)]_{E^*} = \mathbf{B}, \text{ where } \mathbf{B} := \left(B(\mathbf{e}_i, \mathbf{e}_j) \right)$$

$$[\widetilde{\flat}_1(F^*)]_{E^{**}} = (\widetilde{\mathbf{B}})^\top \text{ and } [\widetilde{\flat}_2(E^*)]_{F^{**}} = \widetilde{\mathbf{B}}, \text{ where } \widetilde{\mathbf{B}} := \left(\widetilde{B}(\mathbf{e}_i, \mathbf{e}_j) \right).$$

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 4.9. (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 3.38, 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 4.10. ($\flat_1$ and $\widetilde{\flat}_1$ are "kind of" inverses).

Let V and W be finite-dimensional vector spaces with bases E and F, let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_m}\}$ 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 $\widetilde{\mathbf{B}} = \left(\widetilde{B}(\psi^{\mathbf{f}_i}, \phi^{\mathbf{e}_j})\right)$.

Proof.

1. We have: $\widetilde{\flat}_1 \circ \flat_1 = (\mathbf{v} \mapsto \Phi_{\mathbf{v}}) \iff ((\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1} = \Phi_{\mathbf{v}}$ for all $\mathbf{v} \in V) \iff ((\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1}(\phi) = \Phi_{\mathbf{v}}(\phi)$ for all $\mathbf{v} \in V, \phi \in V^*)$. Since $(\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1}(\phi) = \widetilde{B}(\mathbf{v}^{\flat_1}, \phi) = B(\mathbf{v}, \phi^{\sharp_2}) = \mathbf{v}^{\flat_1}(\phi^{\sharp_2})$ and $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$, an equivalent condition is $(\mathbf{v}^{\flat_1}(\phi^{\sharp_2}) = \phi(\mathbf{v})$ for all $\mathbf{v}, \mathbf{w} \in V)$. Then since $\sharp_2$ is an isomorphism we can substitute $\mathbf{w} \in W$ for $\phi^{\sharp_2}$ to obtain another equivalent condition, $(\mathbf{v}^{\flat_1}(\mathbf{w}) = \mathbf{w}^{\flat_2}(\mathbf{v})$ for all $\mathbf{v} \in V, \phi \in V^*)$, which is equivalent to the obviously true condition $(B(\mathbf{v}, \mathbf{w}) = B(\mathbf{v}, \mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V)$.

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 4.8 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.

□

Metric tensors

Definition 4.11. (Metric tensor).

Let V be a finite-dimensional vector space. A nondegenerate bilinear form g on V that is also *symmetric*, in the sense that $g(\mathbf{v}, \mathbf{w}) = g(\mathbf{w}, \mathbf{v})$ for all $\mathbf{v}, \mathbf{w} \in V$, is called a *metric tensor on V*.

Remark 4.12. (Inner products are metric tensors).

Let V be a vector space over a field K. Recall from Definition 2.155 that an *inner product* on V is a nondegenerate symmetric bilinear form on V and V.

Definition 4.13. (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 $\sharp := \sharp_1 = \flat_1^{-1} = \sharp_2 = \flat_2^{-1}$.

Self-duality

Theorem 4.14. (Self-dual $\iff B(\mathbf{e}_i, \mathbf{e}_j) = \delta^i{}_j$).

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, $B = \langle \cdot, \cdot \rangle$, this means that $\flat$ sends basis vectors from E to induced dual basis vectors iff E is orthonormal with respect to the inner product.

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 $\mathbf{v} \in V \iff \mathbf{e}_i^\flat(\mathbf{e}_j) = \phi^{\mathbf{e}_i}(\mathbf{e}_j) \iff B(\mathbf{e}_i, \mathbf{e}_j) = \phi^{\mathbf{e}_i}(\mathbf{e}_j) \iff B(\mathbf{e}_i, \mathbf{e}_j) = \delta^i{}_j$. $\qquad\square$

Remark 4.15. (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 3.26 actually becomes natural exactly whenever we have a bilinear form B on V and $B(\mathbf{e}_i, \mathbf{e}_j) = \delta^i_j$.

4.2 Coordinates of (p, q) tensors

Theorem 4.16. (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 3.28 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:

$$([\mathbf{v}]_E)^i = C(\mathbf{v}, \phi^{\mathbf{e}_i})$$
$$([\phi]_{E^*})_i = C(\mathbf{e}_i, \phi).$$

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. To prove the first equation in the first line, $([\mathbf{v}]_E)^i = \phi^{\mathbf{e}_i}(\mathbf{v})$, we decompose $\mathbf{v}$ in the basis E:

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

To prove the second line, we decompose ϕ in the basis E^*:

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

Now we prove the second equality of the first line. Recall from Theorem 3.28 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 $\Phi_{\mathbf{v}}(\phi^{\mathbf{e}_i}) = ([\Phi_{\mathbf{v}}]_{E^{**}})^i$. $\qquad\square$

Theorem 4.17. (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

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

.

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 $\phi(E) = [\phi]_{E^*}^\top$. $\qquad\square$

Coordinates with nondegenerate bilinear form

Theorem 4.18. (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

$$[\mathbf{v}]_E = \mathbf{B}^{-\top}[\mathbf{v}^{\flat_1}]_{F^*} \text{ and } [\mathbf{w}]_F = \mathbf{B}^{-1}[\mathbf{w}^{\flat_2}]_{E^*}$$
$$[\phi]_{E^*} = \mathbf{B}[\phi^{\sharp_2}]_F \text{ and } [\psi]_{F^*} = \mathbf{B}^\top[\psi^{\sharp_1}]_E,$$

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 $\widetilde{\mathbf{B}} = (\widetilde{B}(\psi^{\mathbf{f}_i}, \phi^{\mathbf{e}_j}))$.

In the special case when $V = W$ and $E = F$, we have

$$[\mathbf{v}]_E = \mathbf{B}^{-\top}[\mathbf{v}^{\flat_1}]_{E^*} = \mathbf{B}^{-1}[\mathbf{v}^{\flat_2}]_{E^*}$$
$$[\phi]_{E^*} = \mathbf{B}^\top[\phi^{\sharp_1}]_E = \mathbf{B}[\phi^{\sharp_2}]_E,$$

for all $\mathbf{v} \in V, \phi \in V^*$, where $\mathbf{B} = (B(\mathbf{e}_i, \mathbf{e}_j))$ and $\widetilde{\mathbf{B}} = (\widetilde{B}(\phi^{\mathbf{e}_i}, \phi^{\mathbf{e}_j}))$.

Proof. We will prove (1) that $[\psi]_{F^*} = \mathbf{B}^\top[\psi^{\sharp_1}]_E$ for all $\psi \in W^*$ and (2) that $[\mathbf{v}]_E = \mathbf{B}^{-\top}[\mathbf{v}^{\flat_1}]_{F^*}$ for all $\mathbf{v} \in V$. The top right equation 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 bottom left equation 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 (1). 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 4.8 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 (2). 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 $[\Phi]_{E^{**}} = \widetilde{\mathbf{B}}^\top[\Phi^{\widetilde{\sharp}_1}]_{F^*}$ 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 4.16 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 4.10. 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 4.10 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. $\square$

Remark 4.19. (Metric tensors in physics).

In physics, the typical situation is to have a vector space with metric tensor g. (Recall that a metric tensor on a vector space V is a nondegenerate symmetric bilinear form on V and V). In this situation, the symmetry of g further simplifies the special case of the above theorem:

$$[\mathbf{v}]_E = \mathbf{g}^{-1}[\mathbf{v}^\flat]_{E^*} \iff [\mathbf{v}^\flat]_{E^*} = \mathbf{g}[\mathbf{v}]_E$$
$$[\phi]_{E^*} = \mathbf{g}[\phi^\sharp]_E \iff [\phi^\sharp]_E = \mathbf{g}^{-1}[\phi]_{E^*}.$$

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

$$v^i = \sum_j g^{ij} v_j \iff v_i = \sum_j g_{ij} v^j$$
$$\phi_i = \sum_j g_{ij} \phi^j \iff \phi^i = \sum_j g^{ij} \phi_j.$$

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.

Derivation 4.20. (Using a metric tensor to convert between vectors and dual vectors in a (p, q) tensor).

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 consider a (p, q) tensor $\mathbf{T} \in T_{p,q}(V)$ with coordinates $T^{i_1 \cdots i_p}{}_{j_1 \ldots j_q}$ relative to E and E^*,

$$\mathbf{T} = \sum_{\substack{i_1, \ldots, i_p \in \{1, \ldots, n\} \\ j_1, \ldots, j_q \in \{1, \ldots, m\}}} T^{i_1 \cdots i_p}{}_{j_1 \ldots j_q} \mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_k} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes \ldots \otimes \phi^{\mathbf{e}_{j_q}}.$$

If we have a metric tensor g on V, then we can send a basis (p, q) tensor $\mathbf{S}^{i_1 \cdots i_p}{}_{j_1 \ldots j_q}$ to a $(p-1, q+1)$ tensor by applying the map $\flat : V \to V^*$ to one of the p vectors (as opposed to one of the q dual vectors):

$$\mathbf{S}^{i_1 \cdots i_p}{}_{j_1 \ldots j_q} := \mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_k} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes \ldots \otimes \phi^{\mathbf{e}_{j_q}}$$

$$\longmapsto$$

$$\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_k}^{\flat} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes \ldots \otimes \phi^{\mathbf{e}_{j_q}}.$$

Using the equation $[\phi]_{E^*} = \mathbf{g}[\phi^{\sharp}]_E \iff ([\phi]_{E^*})^r = \sum_{j=1}^{n} g_{rj} ([\phi^{\sharp}]_E)^j$ from the previous remark, we compute $\mathbf{e}_{i_k}^{\flat}$ to be

$$\mathbf{e}_{i_k}^{\flat} = \sum_{r=1}^{n} ([\mathbf{e}_{i_k}^{\flat}]_{E^*})^r \phi^{\mathbf{e}_r} = \sum_{r=1}^{n} \left(\sum_{j=1}^{n} g_{rj} ([\mathbf{e}_{i_k}]_E)^j \right) \phi^{\mathbf{e}_r} = \sum_{r=1}^{n} \left(\sum_{j=1}^{n} g_{rj} \delta^j{}_{i_k} \right) \phi^{\mathbf{e}_r} = \sum_{r=1}^{n} g_{r i_k} \phi^{\mathbf{e}_r},$$

so $\mathbf{S}^{i_1 \cdots i_p}{}_{j_1 \ldots j_q}$ is sent to

$$\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_{k-1}} \otimes \sum_{r=1}^{n} \left(g_{r i_k} \phi^{\mathbf{e}_r} \right) \otimes \mathbf{e}_{i_{k+1}} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes \ldots \otimes \phi^{\mathbf{e}_{j_q}}$$

$$= \sum_{r=1}^{n} g_{r i_k} \left(\mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_{k-1}} \otimes \phi^{\mathbf{e}_r} \otimes \mathbf{e}_{i_{k+1}} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \phi^{\mathbf{e}_{j_1}} \otimes \ldots \otimes \phi^{\mathbf{e}_{j_q}} \right).$$

Thus, if the coordinates of $\mathbf{T}$ relative to E and E^* were originally $T^{i_1 \cdots i_p}{}_{j_1 \ldots j_q}$, then they get sent to

$$\sum_{r=1}^{n} g_{r i_k} T^{i_1 \ldots i_{k-1}}{}_r{}^{i_{k+1} \ldots i_p}{}_{j_1 \ldots j_q}.$$

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 $T^{i_1 \cdots i_p}{}_{j_1 \ldots j_q}$ to

$$\sum_{r=1}^{n} g^{r j_k} T^{i_1 \cdots i_p}{}_{j_1 \ldots j_{k-1}}{}^r{}_{j_{k+1} \ldots j_q}.$$

So, using the notation at the end of the previous remark, we have the following "index lowering" and "index raising" mappings:

$$\boxed{\begin{aligned} T^{i_1 \cdots i_p}{}_{j_1 \ldots j_q} &\mapsto \sum_{r=1}^{n} g_{r i_k} T^{i_1 \ldots i_{k-1}}{}_r{}^{i_{k+1} \ldots i_p}{}_{j_1 \ldots j_q} \quad \text{(index lowering)} \\ T^{i_1 \cdots i_p}{}_{j_1 \ldots j_q} &\mapsto \sum_{r=1}^{n} g^{r j_k} T^{i_1 \cdots i_p}{}_{j_1 \ldots j_{k-1}}{}^r{}_{j_{k+1} \ldots j_q} \quad \text{(index raising)} \end{aligned}}$$

Remark 4.21. (Index raising and lowering in greater generality).

It is possible to prove a version of the above result that applies when we have a tensor from $V_1 \otimes ... \otimes V_k$, where $V_i \in \{V, V^*, W, W^*\}$, and a not-necessarily-symmetric nondegenerate bilinear form on finite-dimensional vector spaces V and W.

In this situation, we wouldn't necessarily have $\flat_1 = \flat_2$, $\sharp_1 = \sharp_2$, and $\mathbf{B} = \mathbf{B}^\top$. The full generality of the previous theorem would imply that there would be two different index-lowering operations (one involving multiplication by $B_{i_k r}$ and one involving multiplication by $B_{r i_k}$) and two different index-raising operations (one involving multiplication by $B_{j_k r}$ and one involving multiplication by $B^{r j_k}$).

If the form of $V_1 \otimes ... \otimes V_k$ were set in stone for our purposes, e.g., $V_1 \otimes ... \otimes V_k = V \otimes W^* \otimes V^* \otimes W^{\otimes 2}$, then we would have a natural way to distinguish between indices that are lowered or raised as a result of applying $\flat_1$ or $\sharp_1$ and indices that are lowered or raised as a result of applying $\flat_2$ or $\sharp_2$: if an index corresponds to V or W^*, then $\flat_1$ and $\sharp_1$ are applicable, and if an index corresponds to W or V^*, then $\flat_2$ and $\sharp_2$ are applicable.

If the form of $V_1 \otimes ... \otimes V_k$ weren't set in stone, however- perhaps because we wanted to investigate with as much generality as possible- then things would be much uglier. We would have to distinguish between index operations from resulting from $\flat_1$ and $\sharp_1$ and index operations resulting from $\flat_2$ and $\sharp_2$. This could be achieved by putting primes $'$ or tildes $\sim$ on indices that are raised or lowered with $\flat_2$ or $\sharp_2$, and leaving indices that are raised or lowered with $\flat_1$ or $\sharp_1$ alone. But as you can see, that is quite cumbersome.

Change of basis for (p, q) tensors

Theorem 4.22. (Change of basis for vectors and dual vectors).

Let V be a finite-dimensional vector space with bases E and F, and let E^* and F^* be the corresponding induced dual bases for V^*. Then

$$
\begin{aligned}
[\mathbf{v}]_F &= [\mathbf{E}]_F [\mathbf{v}]_E = [\mathbf{F}]_E^{-1} [\mathbf{v}]_E \\
[\phi]_{F^*} &= [\mathbf{E}]_F^{-\top} [\phi]_{E^*} = [\mathbf{F}]_E^{\top} [\phi]_{E^*} \\
[\mathbf{E}]_F &= [\mathbf{E}^*]_{F^*} = [\mathbf{F}]_E^{-1}
\end{aligned}
$$

where $\mathbf{v} \in V$ and $\phi \in V^*$.

Proof. The first line of the boxed equation is Theorem 2.105, and the equation $[\mathbf{E}]_F = [\mathbf{F}]_E^{-1}$ from the third line is Theorem 2.108.

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 3.31).

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. $\qquad\square$

Remark 4.23. (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 4.24. (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

$$
\begin{aligned}
\mathbf{f}_i &= \sum_{j=1}^{n} ([\mathbf{f}_i]_E)_j \mathbf{e}_j = \sum_{j=1}^{n} ([\mathbf{F}]_E)_i^j \mathbf{e}_j \\
\psi^{\mathbf{f}_i} &= \sum_{j=1}^{n} ([\psi^{\mathbf{f}_i}]_{E^*})_j \phi^{\mathbf{e}_j} = \sum_{j=1}^{n} ([\mathbf{F}]_E^{-\top})_i^j \phi^{\mathbf{e}_j}
\end{aligned}
$$

Proof. The first line in the boxed equation follows directly from the definition of $[\cdot]_F$. (The first line is also Theorem 2.109)). 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 $\psi^{\mathbf{f}_i} = \sum_{j=1}^{n}([\mathbf{F}^*]_{E^*})_i^j \phi^{\mathbf{e}_j}$, where we have $[\mathbf{F}^*]_{E^*} = [\mathbf{F}]_E^{-\top}$ due to the previous theorem. $\qquad\square$

Theorem 4.25. (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 $\mathbf{T}$.

$$\mathbf{f}_{i_1} \otimes ... \otimes \mathbf{f}_{i_p} \otimes \psi^{\mathbf{f}_{j_1}} \otimes ... \otimes \psi^{\mathbf{f}_{j_q}}$$

$$= \Big(\sum_{j_1=1}^{n} ([\mathbf{F}]_E)_{i_1}^{j_1} \mathbf{e}_{j_1} \Big) \otimes ... \otimes \Big(\sum_{j_p=1}^{n} ([\mathbf{F}]_E)_{i_p}^{j_p} \mathbf{e}_{j_p} \Big) \otimes \Big(\sum_{i_1=1}^{n} ([\mathbf{F}_E]^{-1})_{i_1}^{j_1} \phi^{\mathbf{e}_1} \Big) \otimes ... \otimes \Big(\sum_{i_q=1}^{n} ([\mathbf{F}_E]^{-1})_{i_q}^{j_q} \phi^{\mathbf{e}_{i_q}} \Big)$$

$$= \sum_{j_1=1}^{n} ... \sum_{j_p}^{n} \sum_{i_1=1}^{n} ... \sum_{i_q}^{n} \Big(([\mathbf{F}]_E)_{i_1}^{j_1} ... ([\mathbf{F}]_E)_{i_p}^{j_p} ([\mathbf{F}_E]^{-1})_{i_1}^{j_1} ... ([\mathbf{F}_E]^{-1})_{i_q}^{j_q} \mathbf{e}_{j_1} \otimes ... \otimes \mathbf{e}_{j_p} \otimes \phi^{\mathbf{e}_{i_1}} \otimes ... \otimes \phi^{\mathbf{e}_{i_q}} \Big).$$

After substituting this expression back into the basis sum for $\mathbf{T}$, we see that an arbitrary (p,q) tensor with an $\binom{i_1...i_p}{j_1...j_q}$ component of $T^{i_1...i_p}{}_{j_1...j_q}$ relative to F and F^* has a $\binom{i_1...i_p}{j_1...j_q}$ component relative to E and E^* of

$$\sum_{k_1=1}^{n} ... \sum_{k_p}^{n} \sum_{\ell_1=1}^{n} ... \sum_{\ell_q=1}^{n} \Big(([\mathbf{F}]_E)_{\ell_1}^{k_1} ... ([\mathbf{F}]_E)_{\ell_p}^{k_p} ([\mathbf{F}_E]^{-1})_{\ell_1}^{k_1} ... ([\mathbf{F}_E]^{-1})_{\ell_q}^{k_q} T^{k_1...k_p}{}_{\ell_1...\ell_q} \Big).$$

(It is possible to "simplify" this expression by using the fact that $([\mathbf{F}]_E)_j^i ([\mathbf{F}]_E)^{-1})_j^i = \delta^i{}_j$. 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 $\binom{i_1...i_p}{j_1...j_q}$ component of $\mathbf{T}$ relative to E and E^* as

$$([\mathbf{F}]_E)_{\ell_1}^{k_1} ... ([\mathbf{F}]_E)_{\ell_p}^{k_p} ([\mathbf{F}_E]^{-1})_{\ell_1}^{k_1} ... ([\mathbf{F}_E]^{-1})_{\ell_q}^{k_q} T^{k_1...k_p}{}_{\ell_1...\ell_q} \quad \text{(Einstein notation)}.$$

Remark 4.26. (Tensors as "multidimensional matrices" that "transform like tensors").

As was mentioned in Remark 3.47, physicists often define tensors to be "multidimensional matrices" that follow the change of basis formula of the previous theorem.

Tensor contraction

Derivation 4.27. (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 $\mathcal{L}(V \to W) \times \mathcal{L}(W \to Z) \xrightarrow{\circ} \mathcal{L}(V, Z)$. (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 4.28. (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 4.29. (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 3.22 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 4.30. (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 3.45, 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 4.27, 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 4.31. (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 $\left(^{i_1...i_p}_{j_1...j_q}\right)$ coordinate of $\mathbf{T}$ relative to E and E^* is $T^{i_1...i_p}{}_{j_1...j_q}$, then the $\left(^{i_1...i_{p-1}}_{j_1...j_{q-1}}\right)$ component of the (k, ℓ) contraction of $\mathbf{T}$ relative to E and E^* is **come back here** $\sum_{r=1}^n T^{i_1...i_{k-1}\ r\ i_k...i_{p-1}}{}_{j_1...j_{\ell-1}\ r\ j_\ell...j_{q-1}}$.

Proof. Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$. Assume $\mathbf{T}$ has a $\left(^{i_1...i_p}_{j_1...j_q}\right)$ component of $T^{i_1...i_p}{}_{j_1...j_q}$ relative to E and E^*, so

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

Using that $\phi^{\mathbf{e}_{j_\ell}}(\mathbf{e}_{i_k}) = \mathbf{e}_{j_\ell}^\flat(\mathbf{e}_{i_k}) = g(\mathbf{e}_{j_\ell}, \mathbf{e}_{i_k}) = g_{j_\ell i_k} = g_{i_k j_\ell}$, we see that the (k, ℓ) contraction of $\mathbf{T}$ is

$$\sum_{\substack{i_1\ldots,i_p\in\{1,\ldots,n\}\\ j_1,\ldots,j_q\in\{1,\ldots,n\}}} \phi^{\mathbf{e}_{j_\ell}}(\mathbf{e}_{i_k})\, T^{i_1\ldots i_p}{}_{j_1\ldots j_q} \mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_k} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \ldots \otimes \epsilon^{j_\ell} \otimes \ldots \otimes \epsilon^{j_q}$$

$$= \sum_{\substack{i_1\ldots,i_p\in\{1,\ldots,n\}\\ j_1,\ldots,j_q\in\{1,\ldots,n\}}} g_{i_k j_\ell}\, T^{i_1\ldots i_p}{}_{j_1\ldots j_q} \mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_k} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \ldots \otimes \epsilon^{j_\ell} \otimes \ldots \otimes \epsilon^{j_q}$$

$$= \sum_{i_k, j_\ell \in \{1,\ldots,n\}} g_{i_k j_\ell} \sum_{\substack{i_1,\ldots,\not{i_k},\ldots,i_p\in\{1,\ldots,n\}\\ j_1,\ldots,\not{j_\ell},\ldots,j_q\in\{1,\ldots,n\}}} T^{i_1\ldots i_p}{}_{j_1\ldots j_q} \mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_k} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \ldots \otimes \epsilon^{j_\ell} \otimes \ldots \otimes \epsilon^{j_q}$$

$$= \sum_{i_k, j_\ell \in \{1,\ldots,n\}} g_{i_k j_\ell} \sum_{\substack{i_1,\ldots,\not{i_k},\ldots,i_p\in\{1,\ldots,n\}\\ j_1,\ldots,\not{j_\ell},\ldots,j_q\in\{1,\ldots,n\}\\ r\in\{1,\ldots,n\}}} T^{i_1\ldots i_{k-1}\, r\, i_{k+1}\ldots i_p}{}_{j_1\ldots j_{\ell-1}\, r\, j_{\ell+1}\ldots j_q} \mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_k} \otimes \ldots \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes \ldots \otimes \epsilon^{j_\ell} \otimes \ldots \otimes \epsilon^{j_q}$$

So, we can see that

The $\left({}^{i_1\ldots\not{i_k}\ldots i_p}_{\,j_1\ldots\not{j_\ell}\ldots j_p}\right)$ component of the (k, ℓ) contraction of $\mathbf{T}$ 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 $i_{k+1}, \ldots, i_p$ down by one (this is valid because the indices i_k and j_ℓ are no longer "occupied"), we see

The $\left({}^{i_1\ldots i_{p-1}}_{\,j_1\ldots j_{q-1}}\right)$ component of the (k, ℓ) contraction of $\mathbf{T}$ 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

The $\left({}^{i_1\ldots i_{p-1}}_{\,j_1\ldots j_{q-1}}\right)$ component of the (k, ℓ) contraction of $\mathbf{T}$ is $\sum_r T^{i_1\ldots i_{k-1}\, r\, i_k\ldots i_{p-1}}{}_{j_1\ldots j_{\ell-1}\, r\, j_\ell\ldots j_{q-1}}$

when the basis for V is orthonormal.

In Einstein notation, this is stated as

The $\left({}^{i_1\ldots i_{p-1}}_{\,j_1\ldots j_{q-1}}\right)$ component of the (k, ℓ) contraction of $\mathbf{T}$ is $T^{i_1\ldots i_{k-1}\, r\, i_k\ldots i_{p-1}}{}_{j_1\ldots j_{\ell-1}\, r\, j_\ell\ldots j_{q-1}}$

when the basis for V is orthonormal.

$\square$

Theorem 4.32. 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. $\square$

Theorem 4.33. (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: $\operatorname{tr}(a^i{}_j) := \sum_{i=1}^n a^i_i$. We have that $\operatorname{tr}(a^i{}_j)$ 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 $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, \ldots, \hat{\mathfrak{e}}_n\}$ is the standard basis for K^n. Recall from Theorem 3.18 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 $E^* = \{\phi^{\hat{\mathfrak{e}}_1}, \ldots, \phi^{\hat{\mathfrak{e}}_n}\}$ is the basis for $(K^n)^*$ induced by the standard basis $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, \ldots, \hat{\mathfrak{e}}_n\}$ for K^n. The $(1, 1)$ contraction of this $(1, 1)$ tensor is $\sum_{ij} a^i{}_j \phi^{\hat{\mathfrak{e}}_j}(\hat{\mathfrak{e}}^i) = \sum_{ij} a^i{}_j \delta^i{}_j = \sum_i a^i_i = \operatorname{tr}(a^i{}_j)$. $\square$

5

The determinant and orientation

Permutations

Definition 5.1. (Permutation).
A *permutation (on $\{1, ..., n\}$)* is a bijection $\{1, ..., n\} \to \{1, ...n\}$. The set of permutations on $\{1, ..., n\}$ is denoted S_n.

Definition 5.2. (Cycle).
A permutation in S_n is a *cycle* iff it is defined by

$$\begin{cases} i_j \mapsto i_{j+1} & j \in \{1, ..., k-1\} \\ i_j \mapsto i_1 & j = k \end{cases}$$

for some $i_1, ..., i_k \in \{1, ..., n\}$. If such $i_1, ..., i_k$ exist, then the cycle is represented as the following tuple without commas: $\begin{pmatrix} i_1 & ... & i_k \end{pmatrix}$.

Definition 5.3. (Transposition).
A permutation in S_n is a *transposition* iff it swaps two elements and leaves all of the others fixed. Notice, every transposition can be written as the cycle $\begin{pmatrix} i & j \end{pmatrix}$ for some $i, j \in \{1, ..., n\}$.

Now, we prove an intuitive theorem.

Theorem 5.4. (Every permutation is a composition of transpositions).
Every permutation $\sigma \in S_n$ can be decomposed as $\sigma = \sigma_k \circ ... \circ \sigma_1$, where $\sigma_1, ..., \sigma_k \in S_n$ are transpositions. Note, the set of transpositions in the decomposition does not uniquely correspond to σ.

Proof. Notice that

- The function $\text{sort}_\sigma \in S_n$ that sorts $(\sigma(1), ..., \sigma(n))$ from least to greatest satisfies $\text{sort}_\sigma((\sigma(1), ..., \sigma(n))) = (1, ..., n)$.

- We also have $\sigma^{-1}((\sigma(1), ..., \sigma(n))) = (\sigma^{-1}(\sigma(1)), ..., \sigma^{-1}(\sigma(n)) = (1, ..., n)$.

Thus, $\text{sort}_\sigma = \sigma^{-1}$. Since there are many ways to describe sort_σ as a composition of transpositions[1], we have $\text{sort}_\sigma = \sigma^{-1} = \sigma_k \circ ... \circ \sigma_1$, where $\sigma_1 \circ ... \circ \sigma_k$ are transpositions. The inverse of a transposition is a transposition, so we have $\sigma = \sigma_1^{-1} \circ ... \circ \sigma_k^{-1}$, where $\sigma_1^{-1}, ..., \sigma_k^{-1}$ are transpositions. So, we have shown that σ decomposes into a composition of transpositions. This decomposition is not unique, as it depends on which sorting algorithm is used to compute $\text{sort}_\sigma = \sigma^{-1}$.

$\square$

The following definition and theorem are the main reason for our excursion into permutations.

Definition 5.5. (Parity of a permutation).
A permutation in S_n is said to have *even parity*, or to be *even*, iff it is equal to the composition of an even number of transpositions. A permutation is said to have *odd parity*, or to be *odd*, iff it is equal to the composition of an odd number of transpositions.

---

[1]See Definition 11.6 in the appendix for a description of one such description, the *bubble sort algorithm*.

Theorem 5.6. (The parity of a permutation is unique).

A permutation in S_n has a unique parity. That is, no permutation is both even and odd.

Proof. Let $\sigma \in S_n$, and suppose $\sigma = \sigma_k \circ ... \circ \sigma_1$ and $\sigma = \pi_\ell \circ ... \circ \pi_1$, where $\sigma_1, ..., \sigma_k$ and $\pi_1, .., \pi_\ell$ are transpositions. We need to prove that k and ℓ have the same parity.

Suppose for contradiction that k and ℓ have different parities. We have $\sigma_k \circ ... \circ \sigma_1 = \pi_k \circ ... \circ \pi_1$, and so $\mathrm{id} = \pi_1^{-1} \circ ... \circ \pi_\ell^{-1} \circ \sigma_k \circ ... \circ \sigma_1 = \pi_1 \circ ... \circ \pi_\ell \circ \sigma_k \circ ... \circ \sigma_1$. Since k and ℓ have different parities, then $k + \ell$ is odd, and the identity permutation is a composition of an odd number of transpositions.

Since the identity is a composition of an odd number of transpositions, it cannot be a composition of zero transpositions. It also cannot be a composition of one of one transposition, as no single transposition is the identity! Therefore, the identity is a composition of more than one transposition. Let r be the smallest integer with $r > 1$ such that $\mathrm{id} = \tau_r \circ ... \circ \tau_1$, where $\tau_1, ..., \tau_r$ are transpositions.

Consider the following cases.

(Case: $\tau_r = \tau_{r-1}$). Then $\tau_r \circ \tau_{r-1} = \mathrm{id}$, and $\tau_r \circ ... \circ \tau_1 = \tau_{r-2} \circ ... \circ \tau_1$ is a composition of $r-2$ transpositions. We see r is not minimal; this is a contradiction, so this case is impossible.

(Case: $\tau_r \neq \tau_{r-1}$). Suppose $\tau_r = \begin{pmatrix} i & j \end{pmatrix}$.

(Case: τ_r and τ_{r-1} are disjoint). If τ_r and τ_{r-1} are disjoint, then $\tau_r \circ \tau_{r-1} = \tau_{r-1} \circ \tau_r = \tau_{r-1} \circ \begin{pmatrix} i & j \end{pmatrix}$.

(Case: τ_r and τ_{r-1} share one element).

(Case: $\tau_r \circ \tau_{r-1} = \begin{pmatrix} i & j \end{pmatrix} \circ \begin{pmatrix} i & r \end{pmatrix}$ for some $r \neq j$). We have $\tau_r \circ \tau_{r-1} = \begin{pmatrix} i & j \end{pmatrix} \circ \begin{pmatrix} i & r \end{pmatrix} = \begin{pmatrix} i & r & j \end{pmatrix} = \begin{pmatrix} j & i & r \end{pmatrix} = \begin{pmatrix} j & r \end{pmatrix} \circ \begin{pmatrix} i & j \end{pmatrix}$.

(Case: $\tau_r \circ \tau_{r-1} = \begin{pmatrix} i & j \end{pmatrix} \circ \begin{pmatrix} j & s \end{pmatrix}$ for some $s \neq i$). We have $\tau_r \circ \tau_{r-1} = \begin{pmatrix} i & j \end{pmatrix} \circ \begin{pmatrix} j & s \end{pmatrix} = \begin{pmatrix} i & j & s \end{pmatrix} = \begin{pmatrix} s & j \end{pmatrix} \circ \begin{pmatrix} i & s \end{pmatrix}$.

To summarize the above, in all of the possible cases, we have $\tau_r \circ \tau_{r-1} = \begin{pmatrix} i & j \end{pmatrix} \circ \tau_{r-1} = \widetilde{\tau}_r \circ \begin{pmatrix} i & a_1 \end{pmatrix}$ for some $a_1 \in \{1, ..., n\} - \{i\}$ and some transposition $\widetilde{\tau}_r$ that fixes a_1. Thus, $\mathrm{id} = \tau_r \circ ... \circ \tau_1 = \widetilde{\tau}_r \circ \begin{pmatrix} i & a_1 \end{pmatrix} \circ \tau_{r-2} \circ ... \circ \tau_1$, where $\widetilde{\tau}_r$ fixes i. Repeating this process, we see $\mathrm{id} = \tau_r \circ \tau_{r-1} \circ \begin{pmatrix} i & a_2 \end{pmatrix} \circ ... \circ \tau_1$. It follows from induction that there exist $a \in \{1, ..., n\} - \{i\}$ and transpositions $\widetilde{\tau}_2, ..., \widetilde{\tau}_r$ such that $\mathrm{id} = \tau_r \circ ... \circ \tau_1 = \widetilde{\tau}_r \circ ... \circ \widetilde{\tau}_2 \circ \begin{pmatrix} i & a \end{pmatrix}$, where all of $\widetilde{\tau}_2, ..., \widetilde{\tau}_r$ fix i. Thus $\mathrm{id}(i) = a \neq i$. This is a contradiction, since the identity fixes every element of $\{1, ..., n\}$. $\qquad \square$

Definition 5.7. The *sign* $\mathrm{sgn}(\sigma)$ of a permutation $\sigma \in S_n$ is defined to be

$$\mathrm{sgn}(\sigma) := \begin{cases} 1 & \sigma \text{ is even} \\ -1 & \sigma \text{ is odd} \end{cases}.$$

Notice, sgn is a well-defined function because the parity of a permutation is unique.

It's useful to note that $\mathrm{sgn}(\sigma) = (-1)^k$, where k is the number of transpositions in one of the decompositions for σ.

Theorem 5.8. (Sign of a composition of permutations).

We have $\mathrm{sgn}(\pi \circ \sigma) = \mathrm{sgn}(\pi)\mathrm{sgn}(\sigma)$ for all permutations $\sigma, \pi \in S_n$.

Proof. σ is a composition of k transpositions for some k, and π is a composition of ℓ transpositions for some ℓ. Notice that the theorem is equivalent to the statement "the parity of $k + \ell$ is even iff k and ℓ have the same parity and odd iff k and ℓ have different parity". This statement is an elementary fact about integers, so the theorem is true. $\qquad \square$

The determinant

Definition 5.9. (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. $\det(\mathbf{c}_1, ..., \mathbf{c}, ..., \mathbf{c}, ..., \mathbf{c}_n) = 0$ for all $\mathbf{c} \in K^n$. This should hold because when two sides of a parallelapiped coincide, its *n-dimensional* volume is zero.

Derivation 5.10. (Alternatingness).

Let $\mathbf{c}, \mathbf{d} \in K^n$. By the third axiom of the determinant, we have $\det(\mathbf{c}_1, ..., \mathbf{c}, ..., \mathbf{d}, ..., \mathbf{c}_n) + \det(\mathbf{c}_1, ..., \mathbf{d}, ..., \mathbf{c}, ..., \mathbf{c}_n) = \det(\mathbf{c}_1, ..., \mathbf{c} + \mathbf{d}, ..., \mathbf{d} + \mathbf{c}, ..., \mathbf{c}_n) = 0$. Thus,

$$\det(\mathbf{c}_1, ..., \mathbf{c}, ..., \mathbf{d}, ..., \mathbf{c}_n) = -\det(\mathbf{c}_1, ..., \mathbf{d}, ..., \mathbf{c}, ..., \mathbf{c}_n) \text{ for all } \mathbf{c}, \mathbf{d} \in K^n.$$

This condition is known as the *alternatingness* property of the determinant. It is quick to show that alternatingness property not only follows from the third axiom of the determinant, but is equivalent to it.

Notice that the alternatingness property of the determinant implies that there exist $\mathbf{c}_1, ..., \mathbf{c}_n \in K^n$ with $\det(\mathbf{c}_1, ..., \mathbf{c}_n) < 0$. So, we see our intuitive assumptions about volume require that volume be *signed*. In Theorem [...] we will see how signed volume corresponds to the concept of orientation.

We quickly define alternating functions to be functions that have this alternatingness property, as this definition allows for a succinct characterization of the determinant.

Definition 5.11. (Alternating function).

If V and W are vector spaces, then a function $\mathbf{f} : V^{\times k} \to W$ is *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 $\mathbf{v}_1, ..., \mathbf{v}_k \in V$. Equivalently, $\mathbf{f} : V^{\times k} \to W$ is alternating iff $\mathbf{f}(\mathbf{v}_{\sigma(1)}, ..., \mathbf{v}_{\sigma(k)}) = \text{sgn}(\sigma)\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_k)$ for all $\mathbf{v}_1, ..., \mathbf{v}_k \in V$ and $\sigma \in S_k$.

Theorem 5.12. (Succinct characterization of the determinant).

The determinant is the unique[2] alternating linear k-form sending $\hat{\mathfrak{E}} \mapsto 1$. (Recall Definition 4.1).

We now investigate more properties of the determinant that follow from this succint characterization.

Theorem 5.13. (Consequent properties of the determinant).

4. 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 $i \in \{1, ..., n\}$.

5. $\det(\mathbf{c}_1, ..., \mathbf{c}_n) = 0$ iff $\{\mathbf{c}_1, ..., \mathbf{c}_n\}$ is a linearly dependent set.

Proof.

5. Using the axiom "$\det(\mathbf{c}_1, ..., \mathbf{c}, ..., \mathbf{c}, ..., \mathbf{c}_n) = 0$ for all $\mathbf{c} \in K^n$" together with the multilinearity of the determinant, we have

$$\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_n) = \det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n) + \sum_{j=1, j\neq i}^n d_j \det(\mathbf{c}_1, ..., \mathbf{c}_{i-1}, \mathbf{c}_j, \mathbf{c}_{i+1}, ..., \mathbf{c}_j, ..., \mathbf{c}_n)$$

$$= \det\left(\mathbf{c}_1, ..., \mathbf{c}_i + \sum_{j=1, j\neq i}^n d_j\mathbf{c}_j, ..., \mathbf{c}_n\right).$$

---

[2]We have not yet explained why a function satisfying the determinant axioms is unique. We do so in Remark [...].

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($\hat{\mathbf{\epsilon}}_1, ..., \hat{\mathbf{\epsilon}}_n$) = 1.

$\square$

Using the properties we have amassed, we can derive a formula for the determinant.

Definition 5.14. (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.

Derivation 5.15. (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 5.9 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[3] set $(a^i_j) := ([\mathbf{c}_i]_{\hat{\mathbf{\epsilon}}})^j$. Then...

$$\det((a^i_j)) = \det(\mathbf{c}_1, ..., \mathbf{c}_n) = \det\left(\sum_{i_1=1}^{n} a^1_{i_1}\hat{\mathbf{\epsilon}}_{i_1}, ..., \sum_{i_n=1}^{n} a^n_{i_n}\hat{\mathbf{\epsilon}}_{i_n}\right)$$

$$= \sum_{i_1=1}^{n} \det\left(a^1_{i_1}\hat{\mathbf{\epsilon}}_{i_1}, ..., \sum_{i_n=1}^{n} a^n_{i_n}\hat{\mathbf{\epsilon}}_{i_n}\right)$$

$$\vdots$$

$$= \sum_{i_1=1}^{n} ... \sum_{i_n=1}^{n} \det(a^1_{i_1}\hat{\mathbf{\epsilon}}_{i_1}, ..., a^n_{i_n}\hat{\mathbf{\epsilon}}_{i_n})$$

$$= \sum_{i_1=1}^{n} ... \sum_{i_n=1}^{n} \det(a^1_{i_1}\hat{\mathbf{\epsilon}}_{i_1}, ..., a^n_{i_n}\hat{\mathbf{\epsilon}}_{i_n}), \text{ where } i_1, ..., i_n \text{ are distinct from each other}$$

$$= \sum_{\sigma \in S_n} \det(a^1_{\sigma(1)}\hat{\mathbf{\epsilon}}_{\sigma(1)}, ..., a^n_{\sigma(n)}\hat{\mathbf{\epsilon}}_{\sigma(n)})$$

$$= \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)} \det(\hat{\mathbf{\epsilon}}_{\sigma(1)}, ..., \hat{\mathbf{\epsilon}}_{\sigma(n)})$$

$$= \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)} \text{sgn}(\sigma) \det(\hat{\mathbf{\epsilon}}_1, ..., \hat{\mathbf{\epsilon}}_n)$$

$$= \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)} \text{sgn}(\sigma)$$

Therefore, we have

$$\boxed{\det(\mathbf{c}_1, ..., \mathbf{c}_n) = \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)} \text{sgn}(\sigma)}$$

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 $\mathbf{c}_i = \mathbf{c}_j$.

---

[3]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.

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 $\text{sgn}(\sigma)$.

Remark 5.16. (Determinant is well-defined).

Since the boxed formula was derived by assuming the axioms of the determinant, and since the right side of the boxed formula is a well-defined function of every variable it involves, then the determinant is the unique function satisfying the axioms. Thus, the determinant is the unique alternating linear k-form sending $\hat{\mathfrak{E}} \mapsto 1$.

Derivation 5.17. (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 each been shuffled by a different permutation σ:

$$\det(\mathbf{A}) = \sum_{\sigma \in S_n} \det(a^1_{\sigma(1)} \hat{\mathfrak{e}}_{\sigma(1)}, ..., a^n_{\sigma(n)} \hat{\mathfrak{e}}_{\sigma(n)}).$$

Now, notice that a shuffled diagonal matrix that has been shuffled in accordance to a permutation σ can be converted to a diagonal matrix via application of σ^{-1}, transposed (the transpose of a diagonal matrix $\mathbf{D}$ is $\mathbf{D}$), and then re-shuffled via application of σ. Doing this does not change the determinant, since the $\text{sgn}(\sigma^{-1})$ introduced by the shuffle is canceled by the $\text{sgn}(\sigma)$ factor introduced by the re-shuffle.

When this is done to every matrix in the sum, then the sum is what we would obtain if we were to compute $\det(\mathbf{A}^\top)$ by repeating the derivation of the permutation formula for the determinant. Therefore

$$\det(\mathbf{A}) = \det(\mathbf{A}^\top).$$

Theorem 5.18. (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 $\mathbf{B}_j^i$, then $\mathrm{sgn}(\sigma) = \mathrm{sgn}(\pi)(-1)^{i+j-2(k_1+1)} = \mathrm{sgn}(\pi)(-1)^{i+j}$. Thus $\mathrm{sgn}(\sigma) = (-1)^{i+j}\mathrm{sgn}(\pi) \iff \mathrm{sgn}(\pi) = (-1)^{i+j}\mathrm{sgn}(\sigma)$.

So,

$$
\begin{aligned}
a_j^i \det(\mathbf{B}_j^i) &= a_j^i \sum_{\pi \in S_n} a_1^{\pi(1)} ... a_j^{\pi(j)} ..., a_n^{\pi(n)} \mathrm{sgn}(\pi) \\
&= a_j^i \sum_{\sigma \in S_n} a_1^{\pi(1)} ... a_j^{\pi(j)} ..., a_n^{\pi(n)} (-1)^{i+j} \mathrm{sgn}(\sigma) \\
&= (-1)^{i+j} a_j^i \det(\mathbf{B})
\end{aligned}
$$

Thus $a_j^i \det(\mathbf{B}_j^i) = (-1)^{i+j} a_j^i \det(\mathbf{B}) \iff \det(\mathbf{B}) = (-1)^{i+j} a_j^i \det(\mathbf{B}_j^i)$. Now sum all of the $\mathbf{B}$'s (the diagonal shuffled matrices) to get $\det(\mathbf{A}) = \sum_{j=1}^n a_j^i \det(\mathbf{A}_j^i)$. $\square$

Definition 5.19. (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 5.20. 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 5.21. (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 11.2). Then $\det(\mathbf{f}^*) = \det(\mathbf{f})$.

Proof. Recall from[4] condition (3) of Definition 11.4 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 5.17,) we have $\det(\mathbf{f}) = \det(\mathbf{A}) = \det(\mathbf{A}^\top) = \det(\mathbf{f}^*)$. $\square$

Lemma 5.22. (Determinant as scale factor).
Let V and W be vector spaces over the same field. Let $\mathbf{h}$ be an alternating linear k-form. If $\mathbf{f} : V \to W$ is a linear function, then $h(\mathbf{v}_1, ..., \mathbf{v}_k) = \det(\mathbf{f}) h(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k))$.

Proof. Because h is multilinear and alternating, it 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 $(a_j^i) = [\mathbf{f}(E)]_F$, and then use essentially the same argument as was made to derive the permutation formula on the left side of the above. We obtain

$$
\begin{aligned}
h(\mathbf{f}(\mathbf{e}_1), ..., \mathbf{f}(\mathbf{e}_k)) &= \sum_{\sigma \in S_k} a_{\sigma(1)}^1 ... a_{\sigma(n)}^n \mathrm{sgn}(\sigma) h(\mathbf{e}_1, ..., \mathbf{e}_k) = \Big(\sum_{\sigma \in S_k} a_{\sigma(1)}^1 ... a_{\sigma(n)}^n \mathrm{sgn}(\sigma) \Big) h(\mathbf{e}_1, ..., \mathbf{e}_k) \\
&= \det([\mathbf{f}(E)]_F) h(\mathbf{e}_1, ..., \mathbf{e}_k) = \det(\mathbf{f}) h(\mathbf{e}_1, ..., \mathbf{e}_k).
\end{aligned}
$$

So, we have the statement on the basis E

$$
h(\mathbf{f}(\mathbf{e}_1), ..., \mathbf{f}(\mathbf{e}_k)) = \det(\mathbf{f}) h(\mathbf{e}_1, ..., \mathbf{e}_k).
$$

Using the seeming-multilinearity of $\wedge$, we can extend this fact to apply to any list of vectors in V and obtain the lemma. $\square$

Theorem 5.23. (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 $\det(\mathbf{BA}) = \det(\mathbf{B}) \det(\mathbf{A})$.

---

[4]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.

Proof. Set $n := \dim(V) = \dim(W) = \dim(Z)$, and let h be an alternating linear k-form. By the lemma, $\det(\mathbf{g} \circ \mathbf{f})$ satisfies

$$h((\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1), ..., (\mathbf{g} \circ \mathbf{f})(\mathbf{v}_n)) = \det(\mathbf{g} \circ \mathbf{f})h(\mathbf{v}_1, ..., \mathbf{v}_n) \text{ for all } \mathbf{v}_1, ..., \mathbf{v}_n \in V.$$

Notice that the left side is

$$h\Big(\mathbf{g}(\mathbf{f}(\mathbf{v}_1)), ..., \mathbf{g}(\mathbf{f}(\mathbf{v}_n))\Big) = \det(\mathbf{g})h(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_n)) = \det(\mathbf{g})\det(\mathbf{f})h(\mathbf{v}_1, ..., \mathbf{v}_n).$$

Thus

$$\det(\mathbf{g} \circ \mathbf{f})h(\mathbf{v}_1, ..., \mathbf{v}_n) = \det(\mathbf{g})\det(\mathbf{f})h(\mathbf{v}_1, ..., \mathbf{v}_n).$$

This is a statement on the tuple $(\mathbf{v}_1, ..., \mathbf{v}_n) \in V^{\times n}$. Extending this statement to a statement on any tuple $\mathbf{T} \in V^{\times n}$, we have $\det(\mathbf{g} \circ \mathbf{f})h(\mathbf{T}) = \det(\mathbf{g})\det(\mathbf{f})h(\mathbf{T})$. Thus $\det(\mathbf{g} \circ \mathbf{f}) - \det(\mathbf{g})\det(\mathbf{f}))h(\mathbf{T}) = \mathbf{0}$ for all $\mathbf{T} \in V^{\times n}$. Since we can choose h such that $h(\mathbf{T}) \neq 0$ for some $\mathbf{T} \in V^{\times n}$, this forces $\det(\mathbf{g} \circ \mathbf{f}) - \det(\mathbf{g})\det(\mathbf{f}) = 0$, giving us $\det(\mathbf{g} \circ \mathbf{f}) = \det(\mathbf{g})\det(\mathbf{f})$, as desired. $\square$

Theorem 5.24. (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 $\det(\mathbf{f})\det(\mathbf{f}^{-1}) = 1$. $\square$

5.1 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 a vector space* to be a choice of an orthonormal ordered 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.

[TO-DO: fix the terminology of the above]

First notions of orientation

First notions of orientation in two dimensions

Definition 5.25. (Ordered basis).

An *ordered basis* for a vector space is a simply a tuple containing some ordering of the basis vectors for that space.

For example, if V is a 2-dimensional vector space and has basis $E = \{\mathbf{e}_1, \mathbf{e}_2\}$, then $(\mathbf{e}_1, \mathbf{e}_2)$ and $(\mathbf{e}_2, \mathbf{e}_1)$ are both ordered bases of V. We have $(\mathbf{e}_1, \mathbf{e}_2) \neq (\mathbf{e}_2, \mathbf{e}_1)$.

We now discover a consequence of imposing that the bases under consideration be ordered.

Derivation 5.26. (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 orthonormal ordered bases*.

We would now like to work towards a more precise statement of the antisymmetry of orthonormal 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 5.27. (2-dimensional rotation).

Let V be a 2-dimensional inner product space, and let $\widehat{U}$ be an orthonormal ordered basis for V. A 2-*dimensional rotation on* V is a linear function $\mathbf{R}_\theta : V \to V$ whose matrix relative to $\widehat{U}$ and $\widehat{U}$ is

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \text{ where } \theta \in [0, 2\pi).$$

Definition 5.28. (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 $\widehat{W} = (\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2)$ of a 2-dimensional inner product space[5] to be *equivalent under rotation*, and write $\widehat{U} \sim \widehat{W}$, iff there exists a 2-dimensional rotation $\mathbf{R}_\theta$ for which $\widehat{W} = \mathbf{R}_\theta(\widehat{U})$. (Recall Definition 2.79 for the meaning of $\mathbf{R}_\theta(\widehat{U})$).

Note, this notion of rotational equivalence $\sim$ is an equivalence relation.

---

[5]Note, this is the first time we have required that the vector space under consideration be an inner product space. We need this constraint so that we can speak of orthonormal bases.

Theorem 5.29. (Antisymmetry of orthonormal 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 orthonormal 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. When $\theta \in \{\frac{\pi}{2}, \frac{3\pi}{2}\}$, the matrix of $\mathbf{R}_\theta$ relative to $\widehat{U}$ and $\widehat{U}$ is

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

Computing $\mathbf{R}_\theta(\widehat{U}) = \mathbf{R}_\theta((\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2)) = (\mathbf{R}_\theta(\hat{\mathbf{u}}_1), \mathbf{R}_\theta(\hat{\mathbf{u}}_2))$ for $\theta \in \{\frac{\pi}{2}, \frac{3\pi}{2}\}$, we see that the following ordered bases are rotationally equivalent:

$$(\hat{\mathbf{u}}_2, \hat{\mathbf{u}}_1) \sim (-\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2) \sim (\hat{\mathbf{u}}_1, -\hat{\mathbf{u}}_2)$$
$$(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2) \sim (-\hat{\mathbf{u}}_2, \hat{\mathbf{u}}_1) \sim (\hat{\mathbf{u}}_2, -\hat{\mathbf{u}}_1)$$

This is what we noticed in the informal discussion of Derivation 5.26.

First notions of orientation in n dimensions

To study orientation in n dimensions, we generalize the key notion of rotational equivalence by "extending" 2-dimensional rotations so that they can be applied to elements of n-dimensional inner product spaces.

Definition 5.30. (Extension of a function).

Let X and Y be sets, let $X_1 \subset X$ be subset of X, and consider a function $f : X_1 \to Y$. The *extension of f to X* is the function f_{ext} defined by

$$f_{\text{ext}}(x) = \begin{cases} f(x) & x \in X_1 \\ x & x \notin X_1 \end{cases}.$$

We will speak frequently of extensions of 2-dimensional rotations. In fact, we will engage in a bit of abuse of notation: when V is an inner product space, we will say "the 2-dimensional rotation $\mathbf{R}_\theta$ on V" to mean "the extension of the 2-dimensional rotation $\mathbf{R}_\theta$ to V", and use $\mathbf{R}_\theta$ to denote $(\mathbf{R}_\theta)_{\text{ext}}$.

It's helpful to spell out a further detail. If $\mathbf{R}_\theta$ is a 2-dimensional rotation on a finite-dimensional inner product space and $\widehat{U}$ is an orthonormal basis of the space, then the matrix of $\mathbf{R}_\theta$ relative to $\widehat{U}$ and $\widehat{U}$ is

$$[\mathbf{R}_\theta(\widehat{U})]_{\widehat{U}} = \begin{pmatrix} 1 & \ldots & \cos(\theta) & 0 & -\sin(\theta) & 0 \\ 0 & & 0 & \vdots & 0 & \vdots \\ 0 & & \vdots & 1 & \vdots & \vdots \\ \vdots & & 0 & \vdots & 0 & \vdots \\ 0 & \ldots & \sin(\theta) & 0 & \cos(\theta) & 1 \end{pmatrix}.$$

with the ith column and jth column labeled above.

(The columns other than the ith and jth columns are the columns of the $n \times n$ identity matrix).

Now that we are equipped with 2-rotations that can be applied on finite-dimensional inner product spaces, we can generalize our definition of rotational equivalence to n dimensions.

Definition 5.31. (Equivalence under rotation).

We define orthonormal ordered bases $\widehat{U}$ and $\widehat{W}$ of a finite-dimensional inner product space to be *equivalent under rotation*, and write $\widehat{U} \sim \widehat{W}$, iff there exists a 2-dimensional rotation $\mathbf{R}_\theta$ for which $\widehat{W} = \mathbf{R}_\theta(\widehat{U})$. (Recall Definition 2.79 for the meaning of $\mathbf{R}_\theta(\widehat{U})$).

Note, this notion of rotational equivalence $\sim$ is an equivalence relation.

Definition 5.32. (Permutation acting on a tuple).

Let $X = (x_1, ..., x_n)$ be an n-tuple. Given a permutation $\sigma \in S_n$, we define $X^\sigma := (x_{\sigma(1)}, ..., x_{\sigma(n)})$. For example, if $E = (\mathbf{e}_1, ..., \mathbf{e}_n)$ is a basis of a finite-dimensional vector space, then $E^\sigma = (\mathbf{e}_{\sigma(1)}, ..., \mathbf{e}_{\sigma(n)})$.

Theorem 5.33. (Permutations preserve rotational equivalence).

If $\widehat{U}$ and $\widehat{W}$ are orthonormal ordered bases of an n-dimensional inner product space, then $\widehat{U} \sim \widehat{W} \iff \widehat{U}^\sigma \sim \widehat{W}^\sigma$ for all permutations $\sigma \in S_n$.

Proof. Let $\widehat{U} = (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n)$ and $\widehat{W} = (\hat{\mathbf{w}}_1, ..., \hat{\mathbf{w}}_n)$. We have that $\widehat{U} \sim \widehat{W}$ iff there exists $\theta \in [0, 2\pi)$ such that $\widehat{W} = \mathbf{R}_\theta(\widehat{U})$, which is true iff $\hat{\mathbf{w}}_i = \mathbf{R}_\theta(\hat{\mathbf{u}}_i)$ for all i. This is true iff for all $\sigma \in S_n$ we have $\hat{\mathbf{w}}_{\sigma(i)} = \mathbf{R}_\theta(\hat{\mathbf{u}}_{\sigma(i)})$, which is equivalent to $\widehat{U}^\sigma \sim \widehat{W}^\sigma$. $\qquad \square$

Theorem 5.34. (Antisymmetry of orthonormal ordered bases for an n-dimensional inner product space).

Let $\widehat{U}$ be an orthonormal ordered basis of an n-dimensional inner product space. Similarly to what was done in Theorem 5.29, we use $\theta \in \{\frac{\pi}{2}, \frac{3\pi}{2}\}$ in the matrix relative to $\widehat{U}$ and $\widehat{U}$ of a 2-dimensional rotation to obtain the formal characterization of the antisymmetry of orthonormal ordered bases.

For any orthonormal ordered basis $\widehat{U} = (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n)$ of an n-dimensional inner product space, we have

$$(\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_i, ..., \hat{\mathbf{u}}_j, ..., \hat{\mathbf{u}}_n) \sim (\hat{\mathbf{u}}_1, ..., -\hat{\mathbf{u}}_j, ..., \hat{\mathbf{u}}_i, ..., \hat{\mathbf{u}}_n) \sim (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_j, ..., -\hat{\mathbf{u}}_i, ..., \hat{\mathbf{u}}_n) \text{ for all } i, j \in \{1, ..., n\}.$$

This fundamental result leads to the following [...].

Theorem 5.35. (Orthonormal ordered bases of signed vectors).

Let $\widehat{U} = (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n)$ be an orthonormal ordered basis of an n-dimensional inner product space. If $s_1, ..., s_n \in \{-1, 1\}$, then

$$(s_1\hat{\mathbf{u}}_1, ..., s_n\hat{\mathbf{u}}_n) \sim \begin{cases} (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n), & \text{the number of } s_i \text{ equal to } -1 \text{ is } -1 \text{ is even} \\ (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_{j-1}, -\hat{\mathbf{u}}_j, \hat{\mathbf{u}}_{j+1}, ..., \hat{\mathbf{u}}_n) \text{ for any } j \in \{1, ..., n\}, & \text{the number of } s_i \text{ equal to } -1 \text{ is } -1 \text{ is odd} \end{cases}.$$

Proof.

(Case: the number of s_i equal to -1 is even). Since the number of s_i equal to -1 is even, we can pair every $s_i\hat{\mathbf{u}}_i$ with $s_i = -1$ to some $s_j\hat{\mathbf{u}}_j$ with $s_j = -1$, so that all $s_i\hat{\mathbf{u}}_i$ with $s_i = -1$ appear as a member of only one pair. If we swap the vectors in each pair twice, then, as illustrated by the example $(-\hat{\mathbf{u}}_i, -\hat{\mathbf{u}}_j) \sim (\hat{\mathbf{u}}_j, -\hat{\mathbf{u}}_i) \sim (\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$, the coefficients on these vectors in the rotationally equivalent orthonormal ordered basis will be 1, with the order of the vectors being the same as in the original pair. Thus $(s_1\hat{\mathbf{u}}_1, ..., s_n\hat{\mathbf{u}}_n)$ is rotationally equivalent to $(\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n)$.

(Case: the number of s_i equal to -1 is odd). Since the number of s_i equal to -1 is odd, we can pair every $s_i\hat{\mathbf{u}}_i$ with $s_i = -1$ to some $s_j\hat{\mathbf{u}}_j$ with $s_j = -1$, so that all but one $s_i\hat{\mathbf{u}}_i$ with $s_i = -1$ appear as a member of only one pair. Let $s_k\hat{\mathbf{u}}_k$ be the vector with $s_k = -1$ that is not a member of a pair. Applying the same reasoning as in the previous case, we see that $(s_1\hat{\mathbf{u}}_1, ..., s_n\hat{\mathbf{u}}_n)$ is rotationally equivalent to $(\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_{k-1}, -\hat{\mathbf{u}}_k, \hat{\mathbf{u}}_{k+1}, ...\hat{\mathbf{u}}_n)$. Making use of the example $(-\hat{\mathbf{u}}_k, \hat{\mathbf{u}}_\ell) \sim (\hat{\mathbf{u}}_\ell, \hat{\mathbf{u}}_k) \sim (\hat{\mathbf{u}}_k, -\hat{\mathbf{u}}_\ell)$, we see that this orthonormal ordered basis is rotationally equivalent to $(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_{\ell-1}, -\hat{\mathbf{u}}_\ell, \hat{\mathbf{u}}_{\ell+1}, ..., \hat{\mathbf{u}}_n)$ for any $\ell \in \{1, ..., n\}$. Thus, when the number of s_i equal to -1 is odd, the orthonormal ordered basis $(s_1\hat{\mathbf{u}}_1, ..., s_n\hat{\mathbf{u}}_n)$ is rotationally equivalent to $(\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_{\ell-1}, -\hat{\mathbf{u}}_\ell, \hat{\mathbf{u}}_{\ell+1}, ..., \hat{\mathbf{u}}_n)$ for any $\ell \in \{1, ..., n\}$. Replace ℓ with j to obtain the result. $\qquad \square$

Lemma 5.36. Let $\widehat{U}$ be an orthonormal ordered basis of an n-dimensional inner product space. We have $\text{sgn}(\sigma) = 1 \implies \widehat{U}^\sigma \sim \widehat{U}$ for all $\sigma \in S_n$.

Proof. Let $\widehat{U} = (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n)$. The sign of a permutation is 1 iff the permutation consists of an even number of swaps. So, an equivalent statement to the theorem is: for all n, if σ consists of $2n$ transpositions, then $\widehat{U}^\sigma \sim \widehat{U}$. We prove this statement by induction on n.

(Base case). Consider $\widehat{U}^\sigma$, where $\sigma = \tau \circ \pi$, and where π, τ are transpositions.

The permutation π swaps some i, j. Using the example $(\hat{\mathbf{u}}_j, \hat{\mathbf{u}}_i) \sim (-\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$, we see $\widehat{U}^\pi \sim (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_{i-1}, -\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}, ..., \hat{\mathbf{u}}_n)$. Since this orthonormal ordered basis has only one basis vector with a coefficient of -1, it follows from the previous theorem that it is rotationally equivalent to every other such orthonormal ordered basis; in particular, it is rotationally equivalent to $\widehat{W} := (-\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n)$.

So far, we've shown $\widehat{U}^\pi \sim \widehat{W}$. Since permutations preserve rotational equivalence (see Theorem 5.33), we may apply the permutation τ to each side of the rotational equivalence $\widehat{U}^\pi \sim \widehat{W}$ to obtain $(\widehat{U}^\pi)^\tau = \widehat{U}^{\tau \circ \pi} = \widehat{U}^\sigma \sim \widehat{W}^\tau$.

We now compute the right side of this rotational equivalence, $\widehat{W}^\tau$.

(Case: τ swaps 1 with some $\ell > 1$). By studying the example $(\hat{\mathbf{u}}_\ell, -\hat{\mathbf{u}}_1) \sim (\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_\ell)$, we see $\widehat{W}^\tau = (-\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n)^\tau \sim (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n) = \widehat{U}$.

(Case: τ swaps $k \neq 1$ with some $\ell > k$). By studying the example $(\hat{\mathbf{u}}_\ell, \hat{\mathbf{u}}_k) \sim (-\hat{\mathbf{u}}_k, \hat{\mathbf{u}}_\ell)$, we see $\widehat{W}^\tau = (-\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n)^\tau \sim (-\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_{k-1}, -\hat{\mathbf{u}}_k, \hat{\mathbf{u}}_{k+1}, ..., \hat{\mathbf{u}}_n)$. In this last orthonormal ordered basis, the number of basis vectors with a coefficient of -1 is even. It follows from the previous theorem that this last orthonormal ordered basis is rotationally equivalent to $\widehat{U}$.

We've shown that $\widehat{W}^\tau = \widehat{U}$ for all $\tau \in S_n$. Therefore, we have $\widehat{U}^\sigma \sim \widehat{W}^\tau \sim \widehat{U}$, as desired.

(Inductive case). Assume as the inductive hypothesis if σ consists of $2n$ transpositions, then $\widehat{U}^\sigma \sim \widehat{U}$. We need to prove that if σ consists of $2(n+1)$ transpositions, then $\widehat{U}^\sigma \sim \widehat{U}$.

So, assume σ consists of $2(n+1) = 2n+2$ transpositions. Then $\sigma = \tau \circ \pi$, where π consists of $2n$ transpositions and τ consists of 2 transpositions. By the inductive hypothesis, $\widehat{U}^\pi \sim \widehat{U}$. Since permutations preserve rotational equivalence (see Theorem 5.33), then $(\widehat{U}^\pi)^\tau = \widehat{U}^{\tau \circ \pi} = \widehat{U}^\sigma \sim \widehat{U}^\tau$. Since τ consists of 2 transpositions, then from the base case it follows that $\widehat{U}^\tau \sim \widehat{U}$. Thus $\widehat{U}^\sigma \sim \widehat{U}^\tau \sim \widehat{U}$, as desired. $\qquad\square$

Theorem 5.37.

$$\widehat{U}^\sigma \sim \widehat{U}^\pi \iff \operatorname{sgn}(\sigma) = \operatorname{sgn}(\pi) \text{ for all } \sigma, \pi \in S_n.$$

Proof.

($\Longleftarrow$).

The reverse implication is true when the following two statements are:

1. If $\operatorname{sgn}(\pi) = 1$, then $\operatorname{sgn}(\sigma) = 1 \implies \widehat{U}^\sigma \sim \widehat{U}^\pi$ for all $\sigma \in S_n$.

2. If $\operatorname{sgn}(\pi) = -1$, then $\operatorname{sgn}(\sigma) = -1 \implies \widehat{U}^\sigma \sim \widehat{U}^\pi$ for all $\sigma \in S_n$.

We restate the previous lemma for convenience:

- $\operatorname{sgn}(\sigma) = 1 \implies \widehat{U}^\sigma \sim \widehat{U}$ for all $\sigma \in S_n$.

Now we prove (1) and (2).

1. If $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\pi) = 1$, then $\widehat{U}^\sigma \sim \widehat{U} \sim \widehat{U}^\pi$ by the above bullet.

2. Let π be any odd permutation. If σ is an even permutation, then applying the above bullet yields $(\widehat{U}^\pi)^\sigma \sim \widehat{U}^\pi$, i.e. $\widehat{U}^{\sigma \circ \pi} \sim \widehat{U}^\pi$. Thus $\operatorname{sgn}(\sigma) = 1 \implies \widehat{U}^{\sigma \circ \pi} = \widehat{U}^\pi$. In general, every odd permutation is of the form $\sigma \circ \pi$, where σ is some even permutation. Thus, as σ varies over the even permutations, the permutation $\tau := \sigma \circ \pi$ varies over the odd permutations. This makes the statement $\operatorname{sgn}(\sigma) = 1 \implies \widehat{U}^{\sigma \circ \pi} \sim \widehat{U}^\pi$ equivalent to the statement $\operatorname{sgn}(\tau) = -1 \implies \widehat{U}^\tau \sim \widehat{U}^\pi$. Since the former statement is true, so is the later, as desired.

($\Longrightarrow$).

The forward implication is true when the following two statements are:

3. If $\operatorname{sgn}(\pi) = 1$, then $\widehat{U}^\sigma \sim \widehat{U}^\pi \implies \operatorname{sgn}(\sigma) = 1$ for all $\sigma \in S_n$.

4. If $\operatorname{sgn}(\pi) = -1$, then $\widehat{U}^\sigma \sim \widehat{U}^\pi \implies \operatorname{sgn}(\sigma) = -1$ for all $\sigma \in S_n$.

Before we can prove (3) and (4), we need to prove the following bullet point:

- $\widehat{U} \sim \widehat{U}^\sigma \implies \operatorname{sgn}(\sigma) = 1$ for all $\sigma \in S_n$.

  - (Proof). We show the contrapositive, $\operatorname{sgn}(\sigma) = -1 \implies \widehat{U} \not\sim \widehat{U}^\sigma$ for all $\sigma \in S_n$. Because of (2), it suffices to show that $\widehat{U} \not\sim \widehat{U}^\sigma$ for *any* σ with $\operatorname{sgn}(\sigma) = -1$.
  Show $(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2) \not\sim (-\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2)$, and then that $(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_n) \not\sim (-\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_n)$ by induction.

Now we prove (3) and (4).

3. Let π be any even permutation. If σ is an even permutation, then applying the above bullet yields $(\widehat{U}^\pi)^\sigma \sim \widehat{U}^\pi \implies \operatorname{sgn}(\sigma) = 1$, i.e., $\widehat{U}^{\sigma \circ \pi} \sim \widehat{U}^\pi \implies \operatorname{sgn}(\sigma) = 1$. If we allow π to vary over the even permutations, then since σ varies over the even permutations, $\tau := \sigma \circ \pi$ varies over the even permutations, and we obtain the statement $\widehat{U}^\tau \sim \widehat{U}^\pi \implies \operatorname{sgn}(\tau) = 1$ for all $\tau \in S_n$, as desired.

4. The proof is analogous to that of (3). Begin the proof with "Let π be any odd permutation", and use the fact that since σ varies over the even permutations, $\tau := \sigma \circ \pi$ varies over the odd permutations.

$\qquad\square$

Definition 5.38. (Orientation of permuted ordered bases).

LOOK OVER THIS. MAYBE EDIT IN LIGHT OF NEW PREVIOUS THEOREM

This means that there are only two equivalence classes[6] 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 5.39. (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.

Orientation in n dimensions

Definition 5.40. (*n*-dimensional rotations by Euler angles).

Let V be an n-dimensional inner product space and let $\widehat{U} = (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n)$ be an orthonormal ordered basis for V. An *(n-dimensional) Euler rotation* is a function $\mathbf{R}_{\theta_1,...,\theta_k} : V \to V$ of the form $\mathbf{R}_{\theta_1,...,\theta_k} = \mathbf{R}_{k,\theta_k} \circ ... \circ \mathbf{R}_{1,\theta_1}$, where, for each i, $\mathbf{R}_{i,\theta_i}$ is the 2-dimensional rotation by $\theta_i \in [0, 2\pi)$ on the subspace $\text{span}(\widehat{U} - \{\hat{\mathbf{u}}_i\})$. The angles $\theta_1, ..., \theta_k$ are called *Euler angles*.

Theorem 5.41. All Euler rotations are orthogonal linear functions.

Proof. This follows because 2-dimensional rotations are orthogonal linear functions, and since a composition of orthogonal linear functions is another orthogonal linear function. $\square$

Lemma 5.42. (In three dimensions, there is an Euler rotation taking any nonzero vector to any other vector).

Let V be a 3-dimensional inner product space. If $\mathbf{v} \neq \mathbf{0}$, then for all $\mathbf{w} \in V$ with $||\mathbf{w}|| = ||\mathbf{v}||$ there exists an Euler rotation $\mathbf{R}$ such that $\mathbf{R}(\mathbf{v}) = \mathbf{w}$.

Proof. Let $\widehat{U} = (\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \hat{\mathbf{u}}_3)$ be an orthornormal ordered basis of V. If we can find angles α_1, β_1 that rotate $\mathbf{v}$ to align with $\hat{\mathbf{u}}_3$,

$$(\mathbf{R}_{2,\beta_1} \circ \mathbf{R}_{1,\alpha_1})(\mathbf{v}) = ||\mathbf{v}||\hat{\mathbf{u}}_3,$$

and angles α_2, β_2 that rotate $\hat{\mathbf{u}}_3$ to align with $\mathbf{w}$,

$$(\mathbf{R}_{2,\beta_2} \circ \mathbf{R}_{1,\alpha_2})(\hat{\mathbf{u}}_3) = \hat{\mathbf{w}},$$

then the Euler rotation $\mathbf{R} := \mathbf{R}_{2,\beta_2} \circ \mathbf{R}_{1,\alpha_2} \circ \mathbf{R}_{2,\beta_1} \circ \mathbf{R}_{1,\alpha_1}$ satisfies $\mathbf{R}(\mathbf{v}) = \mathbf{w}$, the lemma is proven.

First, we find α_1 and β_1 such that $(\mathbf{R}_{2,\beta_1} \circ \mathbf{R}_{1,\alpha_1})(\mathbf{v}) = ||\mathbf{v}||\hat{\mathbf{u}}_3$.

---

[6]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.

Letting $(v_1, v_2, v_3)^\top := [\mathbf{v}]_{\widehat{U}}$, we have

$$[\mathbf{R}_{1,\alpha_1}(\mathbf{v})]_{\widehat{U}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha_1) & -\sin(\alpha_1) \\ 0 & \sin(\alpha_1) & \cos(\alpha_1) \end{pmatrix}}_{[\mathbf{R}_{1,\alpha_1}(\widehat{U})]_{\widehat{U}}} \underbrace{\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}}_{[\mathbf{v}]_{\widehat{U}}} = \begin{pmatrix} v_1 \\ v_2 \cos(\alpha_1) - v_3 \sin(\alpha_1) \\ v_2 \sin(\alpha_1) + v_3 \cos(\alpha_1) \end{pmatrix}.$$

α_1 is such that the second component of the above is zero,

$$v_2 \cos(\alpha_1) - v_3 \sin(\alpha_1) = 0.$$

One value of α_1 that solves this equation is $\alpha_1 = \arctan(v_2/v_3)$. With this choice of α_1, we have $\cos(\alpha_1) = v_3/\sqrt{v_2^2 + v_3^2}$ and $\sin(\alpha_1) = v_2/\sqrt{v_2^2 + v_3^2}$. Thus the above is

$$[\mathbf{R}_{1,\alpha_1}(\mathbf{v})]_{\widehat{U}} = \begin{pmatrix} v_1 \\ (v_2 v_3)/\sqrt{v_2^2 + v_3^2} - (v_3 v_2)/\sqrt{v_2^2 + v_3^2} \\ (v_2^2 + v_3^2)/\sqrt{v_2^2 + v_3^2} \end{pmatrix} = \begin{pmatrix} v_1 \\ 0 \\ \sqrt{v_2^2 + v_3^2} \end{pmatrix}.$$

Now we have

$$[(\mathbf{R}_{2,\beta_1} \circ \mathbf{R}_{1,\alpha_1})(\mathbf{v})]_{\widehat{U}} = [\mathbf{R}_{2,\beta_1}(\mathbf{R}_{1,\alpha_1}(\mathbf{v}))]_{\widehat{U}} = [\mathbf{R}_{2,\beta_1}(\widehat{U})]_{\widehat{U}} [\mathbf{R}_{1,\alpha_1}(\mathbf{v})]_{\widehat{U}}$$

$$= \underbrace{\begin{pmatrix} \cos(\beta_1) & 0 & -\sin(\beta_1) \\ 0 & 1 & 0 \\ \sin(\beta_1) & 0 & \cos(\beta_1) \end{pmatrix}}_{[\mathbf{R}_{2,\beta_1}(\widehat{U})]_{\widehat{U}}} \underbrace{\begin{pmatrix} v_1 \\ 0 \\ \sqrt{v_2^2 + v_3^2} \end{pmatrix}}_{[\mathbf{R}_{1,\alpha_1}(\mathbf{v})]_{\widehat{U}}} = \begin{pmatrix} v_1 \cos(\beta_1) - \sqrt{v_2^2 + v_3^2} \sin(\beta_1) \\ 0 \\ v_1 \sin(\beta_1) + \sqrt{v_2^2 + v_3^2} \cos(\beta_1) \end{pmatrix}.$$

β_1 is such that $(\mathbf{R}_{2,\beta_1} \circ \mathbf{R}_{1,\alpha_1})(\mathbf{v}) = ||\mathbf{v}|| \hat{\mathbf{u}}_3 \iff [(\mathbf{R}_{2,\beta_1} \circ \mathbf{R}_{1,\alpha_1})(\mathbf{v})]_{\widehat{U}} = ||\mathbf{v}|| \hat{\mathbf{\varepsilon}}_3$. So β_1 satisfies the following system of equations:

$$\begin{cases} v_1 \cos(\beta_1) - \sqrt{v_2^2 + v_3^2} \sin(\beta_1) = 0 \\ 0 = 0 \\ v_1 \sin(\beta_1) + \sqrt{v_2^2 + v_3^2} \cos(\beta_1) = ||\mathbf{v}|| \end{cases}.$$

One value of β_1 that solves the first equation is $\beta_1 = \arctan(v_1/\sqrt{v_2^2 + v_3^2})$.
Since $\cos(\beta_1) = \sqrt{v_2^2 + v_3^2}/||\mathbf{v}||$ and $\sin(\beta_1) = v_1/||\mathbf{v}||$, the third equation is also satisfied:

$$\frac{v_1^2}{||\mathbf{v}||} + \frac{\sqrt{v_2^2 + v_3^2}^2}{||\mathbf{v}||} = \frac{v_1^2 + v_2^2 + v_3^2}{||\mathbf{v}||} = \frac{||\mathbf{v}||^2}{||\mathbf{v}||} = ||\mathbf{v}||.$$

Thus we have found α_1 and β_1 such that $(\mathbf{R}_{2,\beta_1} \circ \mathbf{R}_{1,\alpha_1})(\mathbf{v}) = ||\mathbf{v}|| \hat{\mathbf{u}}_3$.

Now, we find α_2 and β_2 such that $(\mathbf{R}_{2,\beta_2} \circ \mathbf{R}_{1,\alpha_2})(\hat{\mathbf{u}}_3) = \hat{\mathbf{w}}$.
We have

$$[(\mathbf{R}_{2,\beta_2} \circ \mathbf{R}_{1,\alpha_2})(\hat{\mathbf{u}}_3)]_{\widehat{U}} = [\mathbf{R}_{2,\beta_2}(\widehat{U})]_{\widehat{U}} [\mathbf{R}_{1,\alpha_2}(\widehat{U})]_{\widehat{U}} \hat{\mathbf{u}}_3$$

$$= \underbrace{\begin{pmatrix} \cos(\beta_2) & 0 & -\sin(\beta_2) \\ 0 & 1 & 0 \\ \sin(\beta_2) & 0 & \cos(\beta_2) \end{pmatrix}}_{[\mathbf{R}_{2,\beta_2}(\widehat{U})]_{\widehat{U}}} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha_2) & -\sin(\alpha_2) \\ 0 & \sin(\alpha_2) & \cos(\alpha_2) \end{pmatrix}}_{[\mathbf{R}_{1,\alpha_2}(\widehat{U})]_{\widehat{U}}} \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}_{\hat{\mathbf{u}}_3}$$

$$= \begin{pmatrix} \cos(\beta_2) & 0 & -\sin(\beta_2) \\ 0 & 1 & 0 \\ \sin(\beta_2) & 0 & \cos(\beta_2) \end{pmatrix} \begin{pmatrix} 0 \\ -\sin(\alpha_2) \\ \cos(\alpha_2) \end{pmatrix} = \begin{pmatrix} -\cos(\alpha_2)\sin(\beta_2) \\ -\sin(\alpha_2) \\ \cos(\alpha_2)\cos(\beta_2) \end{pmatrix}.$$

α_2 and β_2 are such that $\hat{\mathbf{w}} = (\mathbf{R}_{2,\beta_2} \circ \mathbf{R}_{1,\alpha_2})(\hat{\mathbf{u}}_3) \iff [\hat{\mathbf{w}}]_{\widehat{U}} = [(\mathbf{R}_{2,\beta_2} \circ \mathbf{R}_{1,\alpha_2})(\hat{\mathbf{u}}_3)]_{\widehat{U}}$. Letting $\hat{w}_i$ denote the ith component of $[\hat{\mathbf{w}}]_{\widehat{U}}$, we see α_2 and β_2 must satisfy the following system of equations:

$$\begin{cases} \hat{w}_1 = -\cos(\alpha_2)\sin(\beta_2) \\ \hat{w}_2 = -\sin(\alpha_2) \\ \hat{w}_3 = \cos(\alpha_2)\cos(\beta_2) \end{cases} .$$

One value of α_2 that solves the second equation is $\alpha_2 = -\arcsin(\hat{w}_2)$. Dividing the first equation by the third, we have $\hat{w}_1/\hat{w}_3 = -\tan(\beta_2)$ and thus one value of β_2 that solves the first equation is $\beta_2 = \arctan(-\hat{w}_1/\hat{w}_3) = -\arctan(\hat{w}_1/\hat{w}_3)$.

Thus we have found α_2 and β_2 such that $(\mathbf{R}_{2,\beta_2} \circ \mathbf{R}_{1,\alpha_2})(\hat{\mathbf{u}}_3) = \hat{\mathbf{w}}$. $\qquad\square$

The previous lemma can easily be generalized so that it holds for all dimensions greater than or equal to 2.

Lemma 5.43. (There is an Euler rotation taking any nonzero vector to any other vector).

Let V be an n-dimensional inner product space. If $\mathbf{v} \neq \mathbf{0}$, then for all $\mathbf{w} \in V$ with $||\mathbf{w}|| = ||\mathbf{v}||$ there exists an Euler rotation $\mathbf{R}$ such that $\mathbf{R}(\mathbf{v}) = \mathbf{w}$.

Proof.

(Case: $n = 2$). The rotation $\mathbf{R}_{\theta(\mathbf{v},\mathbf{w})}$ by $\theta(\mathbf{v},\mathbf{w})$ sends $\mathbf{v}$ to $\mathbf{w}$.
(Case: $n = 3$). See the previous lemma.
(Case: $n > 3$).
(Case: n is even). [TO-DO]
(Case: n is odd). [TO-DO] $\qquad\square$

Theorem 5.44. (Any ordered basis is "rotationally close" to a permuted orthonormal basis).

Let V be an finite-dimensional inner product space. If $\widehat{W} = (\hat{\mathbf{w}}_1, ..., \hat{\mathbf{w}}_k)$ and $\widehat{U} = (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_k)$ are orthonormal ordered bases of subspaces of V, then there is an Euler rotation taking $\widehat{W}$ to an ordered basis that is rotationally equivalent to some permutation $\widehat{U}^\sigma$ of $\widehat{U}$.

Proof. By the previous lemma, we know there is an Euler rotation $\mathbf{R}$ that satisfies $\mathbf{R}(\hat{\mathbf{w}}_1) = \hat{\mathbf{u}}_1$.

We first prove that for all i we have $\mathbf{R}(\hat{\mathbf{w}}_i) = s_i\hat{\mathbf{u}}_i$, where $s_i \in \{-1, 1\}$. And we show this by using induction to show that for all i, when $j \leq i$, we have $\mathbf{R}(\hat{\mathbf{w}}_j) = s_j\hat{\mathbf{u}}_j$, where $s_j \in \{-1, 1\}$.

(Base case). We have $R(\hat{\mathbf{w}}_1) = \hat{\mathbf{u}}_1 = s_1 u_1$, where $s_1 = 1$.

(Inductive case). Assume that when $j \leq i$, we have $\mathbf{R}(\hat{\mathbf{w}}_j) = s_j\hat{\mathbf{u}}_j$, where $s_j \in \{-1, 1\}$.

Since $\mathbf{R}$ is an orthogonal linear function, it preserves orthonormality (**see Theorem [...]**), and so $(\mathbf{R}(\hat{\mathbf{w}}_1), ..., \mathbf{R}(\hat{\mathbf{w}}_k))$ is orthonormal. Thus, for each j, $\mathbf{R}(\hat{\mathbf{w}}_j)$ is perpendicular to all of $\mathbf{R}(\hat{\mathbf{w}}_1), ..., \mathbf{R}(\hat{\mathbf{w}}_j), ..., \mathbf{R}(\hat{\mathbf{w}}_k)$. In particular, when $j \geq i + 1$, each $\mathbf{R}(\hat{\mathbf{w}}_j)$ is perpendicular to all of $\mathbf{R}(\hat{\mathbf{w}}_1), ..., \mathbf{R}(\hat{\mathbf{w}}_i)$. By the inductive hypothesis, $\mathbf{R}(\hat{\mathbf{w}}_1) = s_1\hat{\mathbf{u}}_1, ..., \mathbf{R}(\hat{\mathbf{w}}_i) = s_i\hat{\mathbf{u}}_i$, where $s_1, ..., s_i \in \{-1, 1\}$. So when $j \geq i + 1$, each $\mathbf{R}(\hat{\mathbf{w}}_j)$ is perpendicular to all of $s_1\hat{\mathbf{u}}_1, ..., s_i\hat{\mathbf{u}}_i$; it follows that each $\mathbf{R}(\hat{\mathbf{w}}_j)$ with $j \geq i + 1$ is perpendicular to all of $\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_i$. This means that each $\mathbf{R}(\hat{\mathbf{w}}_j)$ with $j \geq i + 1$ is in the orthogonal complement of $\text{span}((\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_i))$.

This orthogonal complement is equal to $\text{span}((\hat{\mathbf{u}}_{i+1}, \ldots, \hat{\mathbf{u}}_k))$. Therefore $\mathbf{R}(\hat{\mathbf{w}}_{i+1}), ..., \mathbf{R}(\hat{\mathbf{w}}_k) \in \text{span}(\hat{\mathbf{u}}_{i+1}, ..., \hat{\mathbf{u}}_k)$. Since $\mathbf{R}$ preserves orthonormality (**again, see Theorem [...]**), then $\mathbf{R}(\hat{\mathbf{w}}_{i+1}), ..., \mathbf{R}(\hat{\mathbf{w}}_k)$ are also orthonormal.

Since $\mathbf{R}(\hat{\mathbf{w}}_{i+1}) \in \text{span}(\hat{\mathbf{u}}_{i+1}, ..., \hat{\mathbf{u}}_k)$, then $\mathbf{R}(\hat{\mathbf{w}}_{i+1}) = c_{i+1}\hat{\mathbf{u}}_{i+1} + c_{i+2}\hat{\mathbf{u}}_{i+2} + ... + c_k\hat{\mathbf{u}}_k$ for some scalars $c_{i+1}, ..., c_k$. We claim that we must have $c_{i+2} = ... = c_k = 0$; assume for contradiction that $c_{\ell_1}, ..., c_{\ell_p}$ are nonzero scalars with $\ell_1, ..., \ell_p > i + 1$. A quick inner product computation shows that $\mathbf{R}(\hat{\mathbf{w}}_{i+1})$ is not perpendicular to any of $\hat{\mathbf{u}}_{\ell_1}, ..., \hat{\mathbf{u}}_{\ell_p}$. Since $\mathbf{R}(\hat{\mathbf{w}}_{i+2}), ..., \mathbf{R}(\hat{\mathbf{w}}_k)$ must all be perpendicular to $\mathbf{R}(\hat{\mathbf{w}}_{i+1})$, it follows that $\mathbf{R}(\hat{\mathbf{w}}_{i+2}), ..., \mathbf{R}(\hat{\mathbf{w}}_k) \notin \text{span}((\hat{\mathbf{u}}_{\ell_1}, ..., \hat{\mathbf{u}}_{\ell_p}))$. Thus $\mathbf{R}(\hat{\mathbf{w}}_{i+2}), ..., \mathbf{R}(\hat{\mathbf{w}}_k) \in \text{span}((\hat{\mathbf{u}}_{i+2}, ..., \hat{\mathbf{u}}_k) - (\hat{\mathbf{u}}_{\ell_1}, ..., \hat{\mathbf{u}}_{\ell_p}))$, which is a subspace of dimension $|[i + 2, k] \cap \mathbb{Z}| - p$. On the other hand, since $(\mathbf{R}(\hat{\mathbf{w}}_{i+2}), ..., \mathbf{R}(\hat{\mathbf{w}}_k))$ is an orthonormal basis, then $(\mathbf{R}(\hat{\mathbf{w}}_{i+2}), ..., \mathbf{R}(\hat{\mathbf{w}}_k))$ span a $|[i + 2, k]|$-dimensional space. This is a contradiction; vectors cannot be in a space that has a lesser dimension than their span. Therefore, $c_{i+2} = ... = c_k = 0$, and $\mathbf{R}(\hat{\mathbf{w}}_{i+1}) = c_{i+1}\hat{\mathbf{u}}_{i+1}$, i.e., $\mathbf{R}(\hat{\mathbf{w}}_{i+1}) \in \text{span}(\hat{\mathbf{u}}_{i+1})$. Since $\mathbf{R}(\hat{\mathbf{w}}_{i+1})$ is a unit vector, then $||\mathbf{R}(\hat{\mathbf{w}}_{i+1})|| = ||s_{i+1}\hat{\mathbf{u}}_{i+1}|| = |s_{i+1}| = 1$. Thus $s_{i+1} \in \{-1, 1\}$, and $\mathbf{R}(\hat{\mathbf{w}}_{i+1}) = s_{i+1}\hat{\mathbf{u}}_{i+1}$, as desired.

Since $\mathbf{R}(\hat{\mathbf{w}}_i) = s_i\hat{\mathbf{u}}_i$ for all i, where $s_i \in \{-1, 1\}$, then $\mathbf{R}(\widehat{W}) = (s_1\hat{\mathbf{u}}_1, ..., s_k\hat{\mathbf{u}}_k)$, where $s_1, ..., s_k \in \{-1, 1\}$.

If the number of s_i equal to -1 is even, then $(s_1\hat{\mathbf{u}}_1, ..., s_k\hat{\mathbf{u}}_k) \sim \widehat{U}^\sigma$ for all σ with $\text{sgn}(\sigma) = 1$, and if the number of s_i equal to -1 is odd, then $(s_1\hat{\mathbf{u}}_1, ..., s_k\hat{\mathbf{u}}_k) \sim \widehat{U}^\sigma$ for all σ with $\text{sgn}(\sigma) = -1$. (Recall Derivation 5.35). We see that in either case, $\mathbf{R}(\widehat{W}) = (s_1\hat{\mathbf{u}}_1, ..., s_k\hat{\mathbf{u}}_k) \sim \widehat{U}^\sigma$ for some $\sigma \in S_k$. This proves the theorem. $\quad\square$

Completing the definition of orientation for finite-dimensional inner product spaces

Definition 5.45. (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 orthonormal ordered 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 2.165).

Theorem 5.46. (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 $[\mathbf{e}_1]_{\widehat{U}}, ..., [\mathbf{e}_i]_{\widehat{U}}, ..., [\mathbf{e}_n]_{\widehat{U}}$ is added to $[\mathbf{e}_i]_{\widehat{U}}$. Recall from Theorem 5.13 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 $\widehat{U}_E = (\hat{\mathbf{w}}_1, ..., \hat{\mathbf{w}}_n)$ is the orthonormal basis obtained by performing Gram-Schmidt on E, then

$$\det([\mathbf{E}]_{\widehat{U}}) = \det([\mathbf{e}_1]_{\widehat{U}}, ..., [\mathbf{e}_n]_{\widehat{U}}) = \det([\hat{\mathbf{w}}_1]_{\widehat{U}}, ..., [\hat{\mathbf{w}}_n]_{\widehat{U}}) = \det([\widehat{U}_E]_{\widehat{U}}).$$

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 5.44 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 **??** guarantees that $\det(\mathbf{R}) = 1$. Thus, since $\widehat{U}^\sigma = \mathbf{R}(\widehat{U}_E)$, we have

$$\det([\widehat{U}_E]_{\widehat{U}}) = \det([\mathbf{R}(\widehat{U}_E)]_{\widehat{U}}) \det([\widehat{U}_E]_{\widehat{U}}) = \det([(\mathbf{R} \circ \mathbf{I})(\widehat{U}_E)]_{\widehat{U}}) = \det([\mathbf{R}(\widehat{U}_E)]_{\widehat{U}}) = \det([\widehat{U}^\sigma]_{\widehat{U}}).$$

To conclude the proof, we will show that $\det([\widehat{U}^\sigma]_{\widehat{U}}) = \text{sgn}(\sigma) \det([\widehat{U}]_{\widehat{U}})$; once we have shown this, we are done, since $\text{sgn}(\sigma) \det([\widehat{U}]_{\widehat{U}}) = \text{sgn}(\sigma) \det(\mathbf{I}) = \text{sgn}(\sigma)$. Since any permutation is a composition of transpositions, 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 $\text{sgn}(\sigma)$ factor in the equation $\det([\widehat{U}^\sigma]_{\widehat{U}}) = \text{sgn}(\sigma) \det([\widehat{U}]_{\widehat{U}})$. $\qquad\square$

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 5.47. (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 5.48. (Antisymmetry of orthonormal ordered bases).

Notice that we still have the previous antisymmetry of orthonormal ordered bases due to the antisymmetry of the determinant.

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 5.49. 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 **??**.

($\supseteq$). 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-dimensional 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 $\text{span}(\hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_n)$ to $\text{span}(\hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_n)$.

We now use the above idea finitely many times. Define $\mathbf{R}_{i+1}$ to be the 2-dimensional 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-dimensional 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-dimensional rotations). By induction, $\mathbf{R}_n \circ \mathbf{f}$ is a composition of 2-dimensional 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-dimensional rotations $\mathbf{R}_n^{-1}$ and $\mathbf{R}_n^{-1} \circ \mathbf{f}$. $\square$

6

The wedge product and exterior powers

The wedge product

We restate an earlier definition for convenience.

Definition 6.1. (Alternating function).
If V and W are vector spaces, then a function $\mathbf{f} : V^{\times k} \to W$ is *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 $\mathbf{v}_1, ..., \mathbf{v}_k \in V$. Equivalently, $\mathbf{f} : V^{\times k} \to W$ is alternating iff $\mathbf{f}(\mathbf{v}_{\sigma(1)}, ..., \mathbf{v}_{\sigma(k)}) = \mathrm{sgn}(\sigma)\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_k)$ for all $\mathbf{v}_1, ..., \mathbf{v}_k \in V$ and $\sigma \in S_k$.

Definition 6.2. (Vector space of alternating functions).
If $V_1, ..., V_k, W$ are vector spaces over a field K, then we use $(\mathrm{alt}\mathcal{L})(V_1 \times ... \times V_k \to W)$ 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.

Derivation 6.3. (The wedge product).
We have a way to represent multilinear functions as vectors. (The way: tensors). Given that we've seen multilinear alternating functions are also highly relevant [determinant], we now seek a way to represent such functions as vectors.

Just as multilinear functions correspond to symbols of the form $\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k$, where $\otimes$ is multilinear, we will have multilinear alternating functions correspond to symbols of the form $\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k$ such that

1. $\wedge$ looks as if it is bilinear, just as was the case with $\otimes$. That is, [...]

2. $\wedge$ is associative, just as was the case with $\otimes$: [...]

3. $\wedge$ looks as it it is an antisymmetric map: [...]

4. $\wedge$ is *skew-commutative*: [...]

5. $\mathbf{v}_1 \wedge ... \mathbf{v} \wedge ... \wedge \mathbf{v} \wedge ... \wedge \mathbf{v}_k = \mathbf{0}$ for all $\mathbf{v} \in V_i$ and all i

We define $V_1 \wedge ... \wedge V_k$ to be the vector space spanned by elements of the form $\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k$, where $\mathbf{v}_1 \in V_1, ..., \mathbf{v}_k \in V_k$.

Theorem 6.4. (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 6.9, 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. $\qquad\square$

Theorem 6.5. (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 6.6. (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 proposed basis, which would make our proposed basis a linearly dependent set. Since there is no $\mathbf{0}$ in the proposed basis, we can follow the proof of Theorem 3.8. $\qquad\square$

Construction of the wedge product

Definition 6.7. (Permutation of a tensor).

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 6.8. (Antisymmetric tensor).

Let $V_1, ..., V_k$ be vector spaces over the same field. A tensor $\mathbf{T} \in V_1 \otimes ... \otimes V_k$ is *antisymmetric* iff $\mathbf{T}^\sigma = -\mathbf{T}$ for all transpositions σ.

Remark 6.9. ("Antisymmetric" vs. "alternating").

Many authors refer to the "antisymmetric tensors" of Definition 6.10 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 define algebras or 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.

Theorem 6.10. (Antisymmetric tensor).

Let $V_1, ..., V_k$ be vector spaces over the same field. A tensor $\mathbf{T} \in V_1 \otimes ... \otimes V_k$ is *antisymmetric* iff $\mathbf{T}^\sigma = \operatorname{sgn}(\sigma)\mathbf{T}$ for all permutations $\sigma \in S_k$.

Proof. We must prove that $\mathbf{T}^\pi = -\mathbf{T}$ for all transpositions $\pi \in S_k \iff \mathbf{T}^\sigma = \operatorname{sgn}(\sigma)\mathbf{T}$ for all permutations $\sigma \in S_k$.

($\implies$). Decompose σ into a product of transpositions, $\sigma = \sigma_k \circ ... \circ \sigma_1$. We have $\mathbf{T}^\sigma = ((\mathbf{T}^{\sigma_1})^{\sigma_2}...)^{\sigma_k} = (-1)^k \mathbf{T} = \operatorname{sgn}(\sigma)\mathbf{T}$.

($\impliedby$). If π is a transposition, then $\operatorname{sgn}(\pi) = -1$ and we have $\mathbf{T}^\pi = \operatorname{sgn}(\pi)\mathbf{T} = -\mathbf{T}$. $\qquad\square$

Now, we provide a way to construct antisymmetric tensors (from arbitrary tensors). This is important, as we haven't shown that antisymmetric tensors actually exist.

Definition 6.11. (Antisymmetrization of elements of tensor product spaces).

Let $V_1, ..., V_k$ be vector spaces over the same field. We define alt : $V_1 \otimes ... \otimes V_k \to V_1 \otimes ... \otimes V_k$ 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

$$\text{alt}(\mathbf{T})^\pi = \left(\sum_{\sigma \in S_k} \text{sgn}(\sigma)(\mathbf{T}^\sigma) \right)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma)(\mathbf{T}^\sigma)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma)\mathbf{T}^{\pi \circ \sigma}.$$

Since S_k is closed under taking the inverse of a permutation (recall, a permutation in S_k is a bijection on $\{1, ..., k\}$), then for every $\sigma \in S_k$ there is a $\tau \in S_k$ such that $\tau = \pi \circ \sigma \iff \sigma = \pi^{-1} \circ \tau$. So the sum becomes

$$\sum_{\tau \in S_k} \text{sgn}(\pi^{-1} \circ \tau)\mathbf{T}^\tau = \text{sgn}(\pi^{-1}) \sum_{\tau \in S_k} \text{sgn}(\tau)\mathbf{T}^\tau = \text{sgn}(\pi)\text{alt}(\mathbf{T}),$$

as desired. $\square$

Definition 6.12. (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 $\mathbf{T} \wedge \mathbf{S} := \text{alt}(\mathbf{T} \otimes \mathbf{S})$.

Remark 6.13. ("Exterior product").

The wedge product is sometimes called the *exterior product*. One may notice that this sounds thematically similar to "outer product".

Recall that in Remark 3.20, we explained that the relationship between the suggestive terms "inner product" and "outer product" is more coincidental than fundamental. The situation is much the same between "outer product" and "exterior product", as the only real connection between "exterior product" and "outer product" is that since the exterior product involves a tensor product of vectors, which can be identified with an outer product of vectors.

In general, the terms "inner product", "outer product", and "exterior product" are more connected by coincidence than by a fundamental relationship.

Lemma 6.14. (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 $\mathbf{T} \wedge \mathbf{S} = \mathbf{0} = \mathbf{S} \wedge \mathbf{T}$.

Interestingly, it turns out that [...]

Theorem 6.15. (Wedge product spaces are spaces of antisymmetric tensors).

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 $\mathbf{T} = \mathbf{v}_{i_1} \otimes ... \otimes \mathbf{v}_{i_k}$ 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 cosets $\{H\sigma \mid \sigma \in S_{2k}\}$ of H in S_{2k}. Since these cosets partition S_{2k}, then

$$\text{alt}(\mathbf{T} \otimes \mathbf{S}) = \sum_{[\pi]_\tau \sigma \in \{\text{cosets}\}} \text{sgn}([\pi]_\tau \sigma)(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau \sigma} = \sum_{\sigma \in S_{2k}} \sum_{[\pi]_\tau \in H} \text{sgn}([\pi]_\tau \sigma)(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau \sigma}$$

$$= \sum_{\sigma \in S_{2k}} \left(\sum_{[\pi]_\tau \in H} \text{sgn}([\pi]_\tau)(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} \right)^\sigma.$$

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

$$\sum_{[\pi]_\tau \in H} \mathrm{sgn}([\pi]_\tau)(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} = \sum_{[\pi]_\tau \in H} \mathrm{sgn}([\pi]_\tau)\mathbf{T}^{[\pi]_\tau} \otimes \mathbf{S} = \Big(\sum_{[\pi]_\tau \in H} \mathrm{sgn}([\pi]_\tau)\mathbf{T}^{[\pi]_\tau} \Big) \otimes \mathbf{S}.$$

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

$$\Big(\sum_{\pi S_k} \mathrm{sgn}(\pi)\mathbf{T}^\pi \Big) \otimes \mathbf{S} = \mathrm{alt}(\mathbf{T}) \otimes \mathbf{S} = \mathbf{0} \otimes \mathbf{S} = \mathbf{0}.$$

The last equality follows by the seeming-multilinearity of $\otimes$. $\qquad\square$

Theorem 6.16. (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 $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \mathrm{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$; a similar argument shows $\mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R}) = \mathrm{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$. First, we have by definition of $\wedge$ that

$$(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \mathrm{alt}((\mathbf{T} \wedge \mathbf{S}) \otimes \mathbf{R}).$$

Subtracting $\mathrm{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$ from both sides and using linearity of alt, we get that

$$(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} - \mathrm{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R}) = \mathrm{alt}((\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \otimes \mathbf{R}) = (\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R}.$$

If we show $(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R} = \mathbf{0}$, then our claim is true. By the previous lemma, it suffices to show that $\mathrm{alt}(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) = \mathbf{0}$, since if this is true then $(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R} = \mathbf{0}$. And this the case: $\mathrm{alt}(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) = \mathrm{alt}(\mathbf{T} \wedge \mathbf{S}) - \mathrm{alt}(\mathbf{T} \otimes \mathbf{S}) = \mathbf{T} \wedge \mathbf{S} - \mathbf{T} \wedge \mathbf{S} = \mathbf{0}$, by linearity of alt and with use of the fact that $\mathbf{T} \wedge \mathbf{S}$ is antisymmetric. $\qquad\square$

Theorem 6.17. $\mathbf{T} \wedge \mathbf{S} := \mathrm{alt}(\mathbf{T} \otimes \mathbf{S})$ satisfies the desired properties of the wedge product

as before, $V_1 \wedge ... \wedge V_k$ is defined to be the vector space of elements spanned by elements of the form $\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k$, where $\mathbf{v}_1 \in V_1, ..., \mathbf{v}_k \in V_k$.

Proof. Property (1) follows by checking the definition $\mathbf{T} \wedge \mathbf{S} := \mathrm{alt}(\mathbf{T} \otimes \mathbf{S})$. Property (2) was proved in Theorem 6.16. Property (3) follows from the definition of alt. Conditions (3) and (4) are logically equivalent, and conditions (3) and (5) are logically equivalent. $\qquad\square$

In Remark **??**, 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.

Theorem 6.18. (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

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

Proof. See the proof of Theorem 3.8. $\qquad\square$

Theorem 6.19. (Fundamental natural isomorphisms for exterior powers).

Theorem 3.45 stated that there are natural isomorphisms

$$\mathcal{L}(V_1 \times ... \times V_k \to W) \cong \mathcal{L}(V_1 \otimes ... \otimes V_k \to W)$$
$$\mathcal{L}(V \to W) \cong W \otimes V^*$$
$$(V \otimes W)^* \cong V^* \otimes W^*$$
$$T_{p,q}(V) \cong T_{q,p}(V^*)$$

Analogously, there are natural isomorphisms

$$
\boxed{
\begin{aligned}
(\mathrm{alt}\mathcal{L})(V_1 \times ... \times V_k \to W) &\cong (\mathrm{alt}\mathcal{L})(V_1 \wedge ... \wedge V_k \to W) \\
(\mathrm{alt}\mathcal{L})(V \to W) &\cong W \wedge V^* \\
(V \wedge W)^* &\cong V^* \wedge W^* \\
\Lambda^k(V)^* &\cong \Lambda^k(V^*)
\end{aligned}
}
$$

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 $\mathbf{f} \in (\mathrm{alt}\mathcal{L})(V \to W)$, 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 3.45; 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.

$\square$

6.1 Wedge products and the determinant

Theorem 6.20. (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,

$$
\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n).
$$

had to introduce this fact in a different form earlier to prove det product rule. so, maybe use this lemma to prove this theorem: Lemma 5.22.

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 $(a_j^i) = [\mathbf{f}(E)]_F$, and then use essentially the same argument as was made to derive the permutation formula on the left side of the above. We obtain

$$
\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n) = \sum_{\sigma \in S_n} \left(a_{\sigma(1)}^1 ... a_{\sigma(n)}^n \mathrm{sgn}(\sigma) \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n \right) = \left(\sum_{\sigma \in S_n} a_{\sigma(1)}^1 ... a_{\sigma(n)}^n \mathrm{sgn}(\sigma) \right) \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n
$$
$$
= \det([\mathbf{f}(E)]_F)(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n) = \det(\mathbf{f})(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n).
$$

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 $\det(\mathbf{f})$.

Theorem 6.21. (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 11.2). 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,

$$\boxed{\mathbf{f}^*(\mathbf{w}_1) \wedge ... \wedge \mathbf{f}^*(\mathbf{w}_n) = \det(\mathbf{f})(\mathbf{w}_1 \wedge ... \wedge \mathbf{w}_n)}$$

6.2 Wedge products and orientation

We now show how the top exterior power of a vector space can be used to describe orientation.

Theorem 6.22. (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 $\text{sgn}(\sigma) < 0$.

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 orthonormal 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 6.20) 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 $\mathbf{e}_i \mapsto \mathbf{f}_i$.

6.3 Pushforward and pullback

[segway]

Definition 6.23. (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 $T_{k,0}(V) = V^{\otimes k}$ and $T_{0,k}(W) = (W^*)^{\otimes k}$, respectively, extending each with the seeming-multilinearity of $\otimes$:

$$\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k \in T_{k,0}(V) \overset{\otimes_{k,0}\mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \otimes ... \otimes \mathbf{f}(\mathbf{v}_k) \in T_{k,0}(W) \quad \text{(``pushforward on } T_{k,0}(V)\text{'')}$$

$$\psi_1 \otimes ... \otimes \psi_k \in T_{0,k}(W) \overset{\otimes_{0,k}\mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi_1) \otimes ... \otimes \mathbf{f}^*(\psi_k) \in T_{0,k}(V) \quad \text{(``pullback on } T_{0,k}(W)\text{'')}.$$

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 6.24. (Star notation for pushforward and pullback).

The above notation of using $\otimes_{0,k}\mathbf{f} : T_{k,0}(V) \to T_{k,0}(W)$ and $\Lambda^k\mathbf{f} : \Lambda^k(V) \to \Lambda^k(W)$ for the pushforwards and using $\otimes_{k,0}\mathbf{f}^*$ and $\Lambda^k\mathbf{f}^*$ for the pullbacks is nonstandard. It is more common to denote the pushforwards by $\mathbf{f}_* : T_{k,0}(V) \to T_{k,0}(W)$ and $\mathbf{f}_* : \Lambda^k(V) \to \Lambda^k(W)$ and the pullbacks by $\mathbf{f}^* : T_{0,k}(W) \to T_{0,k}(V)$ and $\mathbf{f}^* : \Lambda^k(W) \to \Lambda^k(V)$.

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 10. Until then, we do not use the star notation.

6.4 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[1] 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 6.25. (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 3.45 (this isomorphism is discussed in the proof of the referenced theorem).

Definition 6.26. (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,

$$V_1^*\widetilde{\otimes}...\widetilde{\otimes}V_k^* = \mathcal{L}(V_1 \times ... \times V_k \to K).$$

In analogy to the above, we also define $V_1\widetilde{\otimes}...\widetilde{\otimes}V_k := \mathcal{L}(V_1^* \times ... \times V_k^* \to K)$.

Theorem 6.27. (Dual distributes over $\widetilde{\otimes}$).

Let V and W be finite-dimensional vector spaces. Then we have the natural isomorphism

$$(V\widetilde{\otimes}W)^* \cong V^*\widetilde{\otimes}W^*$$

---

[1] $V_1^* \otimes ... \otimes V_k^* \cong (V_1 \otimes ... \otimes V_k)^* = \mathcal{L}(V_1 \otimes ... \otimes V_k \to K) \cong \mathcal{L}(V_1 \times ... \times V_k \to K)$.

Proof. By taking the dual of each side, we see that an equivalent fact is $(V^*\widetilde{\otimes}W^*)^* \cong V\widetilde{\otimes}W$. This is indeed the case, as we have $(V^*\widetilde{\otimes}W^*)^* = \mathcal{L}(V \times W \to K)^* \cong \mathcal{L}(V \otimes W \to K)^* = (V \otimes W)^{**} \cong (V^* \otimes W^*)^* = \mathcal{L}(V^* \otimes W^* \to K) \cong \mathcal{L}(V^* \times W^* \to K) \cong V\widetilde{\otimes}W$. $\qquad\square$

Theorem 6.28. (Natural isomorphisms relating $\otimes$ and $\widetilde{\otimes}$).

For finite-dimensional vector spaces V and W over a field K, we have $V^*\widetilde{\otimes}W^* \cong V^* \otimes W^*$ naturally. It follows from taking the dual of each side of this natural isomorphism and applying that $V\widetilde{\otimes}W \cong V \otimes W$ naturally. Induction then gives that $V^{\widetilde{\otimes}k} \cong V^{\otimes k}$.

Proof. We prove the first statement: we have $V^*\widetilde{\otimes}W^* = \mathcal{L}(V \times W \to K) \cong \mathcal{L}(V \otimes W \to K) = (V \otimes W)^* \cong V^* \otimes W^*$. $\qquad\square$

Definition 6.29. ((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 6.30. (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 $(\phi^1\widetilde{\otimes}...\widetilde{\otimes}\phi^k)^\sigma = \phi^{\sigma(1)}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\sigma(k)}$.

Derivation 6.31. (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 multilinear alternating function because $\wedge$ is antisymmetric and appears multilinear.

Definition 6.32. (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 $V_1\widetilde{\wedge}...\widetilde{\wedge}V_k := \mathcal{L}(V_1^* \times ... \times V_k^* \to K)$.

Theorem 6.33. (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. It follows from taking the dual of each side of this natural isomorphism 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. $\qquad\square$

Definition 6.34. (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 multilinear alternating functions to produce an multilinear alternating 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 6.35. (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^*. Since the pushforward on the top exterior power is multiplication by the determinant (recall Theorem 6.20), then we have

$$\phi^1 \wedge ... \wedge \phi^n = \det(\mathbf{g})\phi^{\mathbf{e}_1} \wedge ... \wedge \phi^{\mathbf{e}_n} \text{ for all } \phi^1, ..., \phi^n \in V^*$$

where $\mathbf{g} : V^* \to V^*$ is the linear function that is defined on basis vectors by $\mathbf{g}(\phi^{\mathbf{e}_i})) = \phi^i$. The matrix $[\mathbf{f}(E^*)]_{E^*}$ of $\mathbf{f}$ relative to E^* and E^* is $\left([\phi^1]_{E^*} \quad \cdots \quad [\phi^1]_{E^*}\right)$, so we have

$$\phi^1 \wedge ... \wedge \phi^n = \det\left([\phi^1]_{E^*} \quad \cdots \quad [\phi^1]_{E^*}\right)\phi^{\mathbf{e}_1} \wedge ... \wedge \phi^{\mathbf{e}_n}$$
$$= \det(\phi^i(\mathbf{e}_j))\phi^{\mathbf{e}_1} \wedge ... \wedge \phi^{\mathbf{e}_n} \text{ for all } \phi^1, ..., \phi^n \in V^*.$$

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 4.16).

Lemma 6.36. (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

$$(\phi^{\mathbf{e}_1} \widetilde{\wedge} ... \widetilde{\wedge} \phi^{\mathbf{e}_n})(\mathbf{e}_1, ..., \mathbf{e}_n) = \left(\sum_{\sigma \in S_n} \text{sgn}(\sigma)\phi^{\mathbf{e}_{\sigma(1)}} \widetilde{\otimes} ... \widetilde{\otimes} \phi^{\mathbf{e}_{\sigma(n)}}\right)(\mathbf{e}_1, ..., \mathbf{e}_n) = \sum_{\sigma \in S_n} \left(\text{sgn}(\sigma)(\phi^{\mathbf{e}_{\sigma(1)}} \widetilde{\otimes} ... \widetilde{\otimes} \phi^{\mathbf{e}_{\sigma(n)}})(\mathbf{e}_1, ..., \mathbf{e}_n)\right).$$

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

$$(\phi^{\mathbf{e}_1} \widetilde{\wedge} ... \widetilde{\wedge} \phi^{\mathbf{e}_n})(\mathbf{e}_1, ..., \mathbf{e}_n) = \text{sgn}(i) \cdot 1 = 1 \cdot 1 = 1.$$

$\square$

Theorem 6.37. (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

$$\boxed{(\phi^1 \widetilde{\wedge} ... \widetilde{\wedge} \phi^n)(\mathbf{v}_1, ..., \mathbf{v}_n) = \det(\phi^i(\mathbf{v}_j)) \text{ for all } \mathbf{v}_1, ..., \mathbf{v}_n \in V}$$

137

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[2]. Thus, when $\mathbf{v}_1,...,\mathbf{v}_n$ are linearly independent, we can interpret $\phi^i(\mathbf{v}_j)$ in the style of Lemma 6.35 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:

$$(\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n)(\mathbf{v}_1,...,\mathbf{v}_n) = \det\left([\phi^1]_{E^*} \quad ... \quad [\phi^1]_{E^*}\right) \text{ when } \mathbf{v}_1,...,\mathbf{v}_n \text{ are linearly independent.}$$

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. $\qquad\square$

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 6.38. (Pushforward on $\widetilde{T}_{k,0}(V)$)

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Definition 6.23 defined the pushforward $\otimes_{k,0}\mathbf{f} : T_{k,0}(V) \to T_{k,0}(W)$ by

$$\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k \stackrel{\widetilde{\otimes}_{k,0}\mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \otimes ... \otimes \mathbf{f}(\mathbf{v}_k).$$

The natural isomorphisms $T_{k,0}(V) \cong \widetilde{T}_{k,0}(V)$ and $T_{k,0}(W) \cong \widetilde{T}_{k,0}(W)$ induce a pushforward map $\widetilde{\otimes}_{k,0}\mathbf{f} : \widetilde{T}_{k,0}(V) \to \widetilde{T}_{k,0}(W$ defined by

$$\mathbf{v}_1\widetilde{\otimes}...\widetilde{\otimes}\mathbf{v}_k \stackrel{\widetilde{\otimes}_{k,0}\mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1)\widetilde{\otimes}...\widetilde{\otimes}\mathbf{f}(\mathbf{v}_k).$$

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

$$\mathbf{v}_1\widetilde{\otimes}...\widetilde{\otimes}\mathbf{v}_k \stackrel{\widetilde{\otimes}_{k,0}\mathbf{f}}{\longmapsto} \mathbf{f}^{\widetilde{\otimes}k}(\mathbf{v}_1,...,\mathbf{v}_k).$$

Derivation 6.39. (Pullback on $\widetilde{T}_{0,k}(W)$).

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Definition 6.23 defined the pullback $\otimes_{k,0}\mathbf{f}^* : T_{0,k}(W) \to T_{0,k}(V)$ by

$$\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k \stackrel{\Lambda^k\mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1)\widetilde{\otimes}...\widetilde{\otimes}\mathbf{f}^*(\psi^k).$$

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 $\widetilde{\otimes}_{0,k}\mathbf{f}^* : \widetilde{T}_{0,k}(W) \to \widetilde{T}_{0,k}(V)$ defined by

$$\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k \stackrel{\widetilde{\otimes}_{0,k}\mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1)\widetilde{\otimes}...\widetilde{\otimes}\mathbf{f}^*(\psi^k) = (\psi^1 \circ \mathbf{f})\widetilde{\otimes}...\widetilde{\otimes}(\psi^k \circ \mathbf{f})$$

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

$$\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k \stackrel{\widetilde{\otimes}_{0,k}\mathbf{f}^*}{\longmapsto} (\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k) \circ \mathbf{f}.$$

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

---

[2]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.

$$\mathbf{T} \in \widetilde{T}_{0,k}(W) \overset{\widetilde{\otimes}_{0,k}\mathbf{f}^*}{\longmapsto} \mathbf{T} \circ \mathbf{f} \in \widetilde{T}_{0,k}(V) \text{ for all } \mathbf{T} \in \widetilde{T}_{0,k}(W).$$

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 6.40. (Pushforward on $\widetilde{\Lambda}^k(V)$).
Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Definition 6.23 defined the pushforward $\Lambda^k\mathbf{f} : \Lambda^k(V) \to \Lambda^k(W)$ 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,

$$\mathrm{alt}(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k) \overset{\Lambda^k\mathbf{f}}{\longmapsto} \mathrm{alt}(\otimes_{k,0}\mathbf{f}(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k)).$$

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

$$\mathbf{v}_1\widetilde{\wedge}...\widetilde{\wedge}\mathbf{v}_k = \widetilde{\mathrm{alt}}(\mathbf{v}_1\widetilde{\otimes}...\widetilde{\otimes}\mathbf{v}_k) \overset{\widetilde{\Lambda}^k\mathbf{f}}{\longmapsto} \widetilde{\mathrm{alt}}(\widetilde{\otimes}_{k,0}\mathbf{f}(\mathbf{v}_1\widetilde{\otimes}...\widetilde{\otimes}\mathbf{v}_k)) = \mathbf{f}^{\widetilde{\wedge}k}(\mathbf{v}_1, ..., \mathbf{v}_k).$$

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 6.41. (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 6.23 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,

$$\mathrm{alt}(\psi^1 \otimes ... \otimes \psi^k) \overset{\Lambda^k\mathbf{f}^*}{\longmapsto} \mathrm{alt}(\otimes_{0,k}\mathbf{f}^*(\psi^1 \otimes ... \otimes \psi^k)).$$

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

$$\psi^1\widetilde{\wedge}...\widetilde{\wedge}\psi^k = \widetilde{\mathrm{alt}}(\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k) \overset{\widetilde{\Lambda}^k\mathbf{f}^*}{\longmapsto} \widetilde{\mathrm{alt}}(\widetilde{\otimes}_{0,k}\mathbf{f}^*(\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k))$$
$$= \widetilde{\mathrm{alt}}((\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k) \circ \mathbf{f}) = \widetilde{\mathrm{alt}}(\psi^1\widetilde{\otimes}...\widetilde{\otimes}\psi^k) \circ \mathbf{f} = (\psi^1\widetilde{\wedge}...\widetilde{\wedge}\psi^k) \circ \mathbf{f}.$$

(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 6.39. 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 multilinear alternating functions $\mathbf{T}$, while $\widetilde{\otimes}_{0,k}\mathbf{f}^*$ acts on multilinear (not necessarily alternating) functions $\mathbf{T}$.

6.5 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 $\mathbf{v} \times \mathbf{w} = \Big(([\mathbf{v}]_E)^2([\mathbf{w}]_E)^3 - ([\mathbf{v}]_E)^3([\mathbf{w}]_E)^2\Big) - \Big(([\mathbf{v}]_E)^1([\mathbf{w}]_E)^3 - ([\mathbf{v}]_E)^3([\mathbf{w}]_E)^1\Big) + \Big(([\mathbf{v}]_E)^1([\mathbf{w}]_E)^2 - ([\mathbf{v}]_E)^2([\mathbf{w}]_E)^1\Big)).$

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 6.42. (Cross product).

In Section 2.4, 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 6.43. (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.

$\square$

The below is in-progress

Definition 6.44. (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 6.45. (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. $\qquad\square$

Derivation 6.46. (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 6.47. (Cyclic pattern with the cross product).

$\mathbf{e}_{\sigma(1)} \times \mathbf{e}_{\sigma(2)} = \operatorname{sgn}(\sigma)\mathbf{e}_{\sigma(3)}$

so $\hat{\mathbf{\varepsilon}}_1 \times \hat{\mathbf{\varepsilon}}_2 = \hat{\mathbf{\varepsilon}}_3$, $\hat{\mathbf{\varepsilon}}_2 \times \hat{\mathbf{\varepsilon}}_3 = \hat{\mathbf{\varepsilon}}_1$, $\hat{\mathbf{\varepsilon}}_3 \times \hat{\mathbf{\varepsilon}}_1 = \hat{\mathbf{\varepsilon}}_2$

Derivation 6.48. (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 $\perp: \Lambda^2(\mathbb{R}^3) \to \mathbb{R}^3$ is defined by $\perp (\hat{\mathbf{\varepsilon}}_{\sigma(1)} \wedge \hat{\mathbf{\varepsilon}}_{\sigma(2)}) = \operatorname{sgn}(\sigma)\hat{\mathbf{\varepsilon}}_{\sigma(3)}$. We call $\perp$ the *Hodge dual.*

$$\perp (\mathbf{v}_1 \wedge \mathbf{v}_2) = \perp \Big(\sum_{i=1}^{n} \big(([\mathbf{v}_1]_{\hat{\mathbf{\mathfrak{E}}}})_i \hat{\mathbf{\varepsilon}}_i\big) \wedge \sum_{i=1}^{n} \big(([\mathbf{v}_2]_{\hat{\mathbf{\mathfrak{E}}}})_i \hat{\mathbf{\varepsilon}}_i\big) \Big) = ... = \mathbf{v}_1 \times \mathbf{v}_2$$

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 6.49. (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[3]. Thus, to show the condition, it suffices to show that the two functions agree on a particular input. We show $\perp (\hat{\mathbf{\varepsilon}}_1, ..., \hat{\mathbf{\varepsilon}}_{n-1}) \cdot \mathbf{v} = \det(\hat{\mathbf{\varepsilon}}_1, ..., \hat{\mathbf{\varepsilon}}_{n-1}, \mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^n$. This condition is logically equivalent to the condition "$\hat{\mathbf{\varepsilon}}_n \cdot \mathbf{v} = \det(\hat{\mathbf{\varepsilon}}_1, ..., \hat{\mathbf{\varepsilon}}_{n-1}, \mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^n$", which is in turn logically equivalent to "$\hat{\mathbf{\varepsilon}}_n \cdot \mathbf{v} = \det(\hat{\mathbf{\varepsilon}}_1, ..., \hat{\mathbf{\varepsilon}}_{n-1}, \mathbf{v})$ for all $\mathbf{v} \in \hat{\mathbf{\mathfrak{E}}}$, where $\hat{\mathbf{\mathfrak{E}}}$ is the standard basis for $\mathbb{R}^n$". When $\mathbf{v} = \hat{\mathbf{\varepsilon}}_i \in \hat{\mathbf{\mathfrak{E}}}$, we have $\hat{\mathbf{\varepsilon}}_n \cdot \hat{\mathbf{\varepsilon}}_i = \delta^i{}_n = \det(\hat{\mathbf{\varepsilon}}_1, ..., \hat{\mathbf{\varepsilon}}_{n-1}, \hat{\mathbf{\varepsilon}}_i)$, so this last condition (and thus all of them) are true. $\qquad\square$

---

[3]In Section 5, we proved that the determinant is the unique alternating linear k-form function sending $\hat{\mathbf{\mathfrak{E}}} \mapsto 1$. An arbitrary alternating linear k-form $f : (K^n)^{\times n} \to K$ is such that $\frac{1}{c}f$ satisfies $(\frac{1}{c}f)(\hat{\mathbf{\varepsilon}}_1, ..., \hat{\mathbf{\varepsilon}}_n) = 1$, so such an f is $f = c \det$.

Remark 6.50. (Interpretation of k-wedges as oriented volumes).

We can actually define the Hodge dual in a more general context.

Definition 6.51. (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

$$\perp (\hat{\mathbf{u}}_{i_1} \wedge ... \wedge \hat{\mathbf{u}}_{i_k}) := \text{sgn}(\sigma)\ \hat{\mathbf{u}}_{i_{k+1}} \wedge ... \wedge \hat{\mathbf{u}}_{i_{k+(n-k)}},$$
$$\text{where } \{i_{k+1}, ..., i_{k+(n-k)}\} = \{1, ..., n\} - \{i_1, ..., i_k\} \text{ and } \sigma = (i_1, ..., i_n).$$

[Extending with multilinearity and alternatingness], we see that

$$\perp (\mathbf{v}_{i_1} \wedge ... \wedge \mathbf{v}_{i_k}) = \text{sgn}(\sigma) \det\left((g(\mathbf{v}_{i_k}, \mathbf{v}_{j_\ell})) \right) \mathbf{v}_{i_{k+1}} \wedge ... \wedge \mathbf{v}_{i_n},$$
$$\text{where } \{i_{k+1}, ..., i_n\} = \{1, ..., n\} - \{i_1, ..., i_k\} \text{ and } \sigma = (i_1, ..., i_n).$$

In the above, $\det\left((\langle \mathbf{v}_{i_k}, \mathbf{v}_{j_\ell} \rangle) \right)$ 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

$$\perp \mathbf{v} = \perp (\mathbf{v}_{i_1} \wedge ... \wedge \mathbf{v}_{i_k}) = \text{sgn}(\sigma)\widetilde{g}(\mathbf{v}, \mathbf{v})\ \mathbf{v}_{i_{k+1}} \wedge ... \wedge \mathbf{v}_{i_n},$$
$$\text{where } \{i_{k+1}, ..., i_n\} = \{1, ..., n\} - \{i_1, ..., i_k\} \text{ and } \sigma = (i_1, ..., i_n).$$

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 6.52. 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 6.53. (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 $\perp \mathbf{v} = \text{sgn}(\sigma)g(\mathbf{v}, \mathbf{v})\mathbf{v}_{i_{k+1}} \wedge ... \wedge \mathbf{v}_{i_n}$, where $\sigma = (i_1, ..., i_n)$. Applying $\perp$ again, we have $\perp\perp \mathbf{v} = \text{sgn}(\tau)\text{sgn}(\sigma)\mathbf{v}_{i_1} \wedge ... \wedge \mathbf{v}_{i_k} = \text{sgn}(\tau)\text{sgn}(\sigma)\mathbf{v}$, where $\tau = (i_{k+1}, ..., i_n, i_1, ..., i_k)$. The key is to notice that $\text{sgn}(\tau) = (-1)^{k(n-k+1)}\text{sgn}(\sigma)$, 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 $\perp\perp \mathbf{v} = (-1)^{k(n-k+1)}\text{sgn}(\sigma)^2\hat{\mathbf{v}} = (-1)^{k(n-k+1)}\hat{\mathbf{v}}$. $\square$

Theorem 6.54. 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 6.55. 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

$$\omega \wedge (\perp \eta) = \widetilde{g}(\omega, \eta)\omega_{\text{vol}},$$

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

7

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.

7.1   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 7.1. (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 7.2. (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.

<div align="center">

Leibniz notation

$$\frac{df}{dt} := f'$$

$$\left.\frac{df}{dt}\right|_{t_0} = f'(t_0)$$

Operator notation

$$\frac{d}{dt}f := f'$$

$$\frac{d}{dt}f(t) := \left.\left(\frac{d}{ds}f\right)\right|_{s=t} = f'(t)$$

$$\frac{df(t)}{dt} := \frac{d}{dt}f(t)$$

</div>

Remark 7.3. ($\frac{df}{dt}$ 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 7.4. (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

$$\frac{df}{dx}\bigg|_s := \frac{df}{dt}\bigg|_{t=x(s)} = f'(x(s)).$$

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:

$$\frac{d(f \circ x)}{dt}\bigg|_t = \frac{df(u)}{du}\frac{du(t)}{dt}, \text{ where } u = x(t).$$

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 7.5. (Letters in the denominator).

The definition $\frac{df}{dt} := f'$ from Definition 7.2 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!

7.2 Multivariable calculus

Lemma 7.6. (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

$$\frac{d(f \circ \mathbf{x})}{dt}\bigg|_{t_0} = \frac{\partial f(\mathbf{x})}{\partial x_1}\bigg|_{\mathbf{p}} \frac{dx_1}{dt}\bigg|_{t_0} + ... + \frac{\partial f(\mathbf{x})}{\partial x_n}\bigg|_{\mathbf{p}} \frac{dx_n}{dt}\bigg|_{t_0}.$$

In other words,

$$\frac{d(f \circ \mathbf{x})}{dt}\bigg|_{\mathbf{p}} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}(t)}{dt}\bigg|_{t_0},$$

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 7.4. 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

$$\boxed{\frac{d(f \circ \mathbf{x})}{dt} = (\nabla_{\mathbf{x}} f) \cdot \frac{d\mathbf{x}(t)}{dt}}$$

Definition 7.7. (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

$$\frac{\partial f}{\partial \mathbf{v}}\bigg|_{\mathbf{p}} := \frac{d(f \circ \mathbf{x})}{dt}\bigg|_{\mathbf{p}} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}(t)}{dt}\bigg|_{t_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v}.$$

Therefore the directional derivative is expressed as

$$\boxed{\begin{aligned} \frac{\partial f}{\partial \mathbf{v}}\bigg|_{\mathbf{p}} &= (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} \\ \frac{\partial f}{\partial \mathbf{v}} &= \nabla f \cdot \mathbf{v} \end{aligned}}$$

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 7.8. [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 7.9. (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

$$\frac{\partial f}{\partial \mathbf{v}}\bigg|_{\mathbf{p}} = \frac{d(f \circ \mathbf{x})(t)}{dt}\bigg|_{t_0} = \frac{df(\mathbf{p}+\mathbf{v}t)}{dt}\bigg|_{t_0} = \lim_{h \to 0} \left(\frac{f(\mathbf{p}+\mathbf{v}(t+h)) - f(\mathbf{p}+vt)}{h} \right)\bigg|_{t_0}$$

$$= \lim_{h \to 0} \left(\left(\frac{f(\mathbf{p}+\mathbf{v}(t+h)) - f(\mathbf{p}+vt)}{h} \right)\bigg|_{t_0} \right) = \lim_{h \to 0} \frac{f(\mathbf{p}+h\mathbf{v}) - f(\mathbf{v})}{h}.$$

Therefore

$$\frac{\partial f}{\partial \mathbf{v}}\bigg|_{\mathbf{p}} = \lim_{h \to 0} \frac{f(\mathbf{p}+h\mathbf{v}) - f(\mathbf{v})}{h}.$$

Many authors define the directional derivative using this formula.

Theorem 7.10. (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 $(\nabla_{\mathbf{x}}f)|_{\mathbf{p}} \cdot \mathbf{v} = ||(\nabla_{\mathbf{x}}f)|_{\mathbf{p}}|| \, ||\text{proj}(\mathbf{v} \to (\nabla_{\mathbf{x}}f)|_{\mathbf{p}})||$. 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 $(\nabla_{\mathbf{x}}f)|_{\mathbf{p}} \cdot \mathbf{v} = ||(\nabla_{\mathbf{x}}f)|_{\mathbf{p}}|| \, ||\text{proj}(\mathbf{v} \to (\nabla_{\mathbf{x}}f)|_{\mathbf{p}})||$. When $||\mathbf{v}|| = 1$, this reduces to $||(\nabla_{\mathbf{x}}f)|_{\mathbf{p}}||$. $\qquad\square$

Remark 7.11. (Directional derivative simplifies to partial derivative).

We have $\frac{\partial}{\partial \hat{\mathbf{e}}_i} = \frac{\partial}{\partial x_i}$.

Lemma 7.12. (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 $\mathbf{p} = \mathbf{x}(t_0)$.

$$\frac{d(\mathbf{f} \circ \mathbf{x})(t)}{dt}\bigg|_{\mathbf{p}} = \begin{pmatrix} \frac{d}{dt}f_1(\mathbf{x}(t)) \\ \vdots \\ \frac{d}{dt}f_m(\mathbf{x}(t)) \end{pmatrix} = \begin{pmatrix} (\nabla_{\mathbf{x}}f_1)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}}{dt}\big|_{t_0} \\ \vdots \\ (\nabla_{\mathbf{x}}f_m)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}}{dt}\big|_{t_0} \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \end{pmatrix}\bigg|_{\mathbf{p}} \frac{d\mathbf{x}}{dt}\big|_{t_0}.$$

In terms of functions, we have

$$\frac{d(\mathbf{f} \circ \mathbf{x})(t)}{dt} = \begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \end{pmatrix} \frac{d\mathbf{x}}{dt}$$

Recall from Derivation **??** and Theorem 2.89 that a matrix-vector product can be expressed as either a linear combination of column vectors or as a vector of dot products. We have already seen the second expression; here is the first:

$$\begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \end{pmatrix} \frac{d\mathbf{x}}{dt} = \begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \cdot \frac{d\mathbf{x}}{dt} \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \cdot \frac{d\mathbf{x}}{dt} \end{pmatrix} = \sum_{i=1}^{n} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_i} \frac{dx_i(t)}{dt}.$$

Definition 7.13. (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 7.4), we define the *Jacobian matrix* $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ to be

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} := \begin{pmatrix} \nabla_{\mathbf{x}}(f_1) \\ \vdots \\ \nabla_{\mathbf{x}}(f_m) \end{pmatrix}$$

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

$$\frac{d(\mathbf{f} \circ \mathbf{x})}{dt} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt}$$

Definition 7.14. (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

$$\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\Big|_{\mathbf{p}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{p}} \frac{d\mathbf{x}}{dt}\Big|_{t_0}$$
$$\frac{\partial \mathbf{f}}{\partial \mathbf{v}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt}$$

Remark 7.15. (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 7.16. (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.

8

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 8.1. (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 8.2. (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 8.3. (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). Since an arbitrary union of open sets in $\mathbb{R}$ is open, then A must be open. $\qquad \square$

Next, see what the interior point characterization of open sets implies for complements of open sets.

Derivation 8.4. (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!

$$U \text{ is open}$$
$$U = \text{int}(U) \iff U \subseteq \text{int}(U) \iff \forall x \in U \ x \in \text{int}(U)$$
$$\forall x \in U \ \exists \text{open } U_x \ni x \ U_x \subseteq U$$
$$\forall x \in U \ \exists \text{open } U_x \ni x \ U_x \cap U^c = \emptyset$$
$$\forall x \ x \notin U^c \implies \text{not}(\forall \text{open } U_x \ni x \ \ U_x \cap U^c \neq \emptyset)$$
$$\forall x \ x \notin U^c \implies \text{not}(\forall \text{open } U_x \ni x \ \ U_x \cap U^c - \{x\} \neq \emptyset)$$
$$\forall x \ x \notin U^c \implies \ x \notin (U^c)'$$
$$\forall x \ x \in (U^c)' \implies \ x \in U^c$$
$$(U^c)' \subseteq U^c.$$

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 $\forall \text{open } U_x \ \ U_x \cap U^c - \{x\} \neq \emptyset$. That is, $x \in \mathbb{R}$ is a *limit point of $A \subseteq \mathbb{R}$* iff every open set containing x intersects A at a point other than x. The set of limit points of A is denoted A'.

The above shows that U is open iff $(U^c)' \subseteq U^c$. In words, a set is open iff its complement contains all of its limit points. For this reason, we define a *closed set in $\mathbb{R}$* to be any set which is the complement of an open set, or, equivalently, any set which contains all of its limit points. We call the fact that C is closed iff $C' \subseteq C$ the *limit point characterization of closed sets*. We repeat these definitions and facts below.

Definition 8.5. (Limit point in $\mathbb{R}$).

Let $A \subseteq \mathbb{R}$ be any subset of $\mathbb{R}$. We say $x \in \mathbb{R}$ is a *limit point of $A \subseteq \mathbb{R}$* iff every open set containing x intersects A at a point other than x. Symbolically, $x \in \mathbb{R}$ is a *limit point of $A \subseteq \mathbb{R}$* iff $\forall \text{open } U_x \ \ U_x \cap U^c - \{x\} \neq \emptyset$.

The set of limit points of A is denoted A'.

Definition 8.6. (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 $\mathbb{R}$.

Theorem 8.7. (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 8.8. (Characterizations of open sets and closed sets in $\mathbb{R}$).

We have seen that open sets in $\mathbb{R}$ are characterized by the "interior point characterization of open sets", while closed sets in $\mathbb{R}$ are characterized by the "limit point characterization of closed sets".

Remark 8.9. (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).

8.1 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 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 8.10. (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 $\{C_\alpha\}_{\alpha \in I} \subseteq \mathcal{C}$. (Here, I denotes an indexing set).

We can now define topological spaces.

Definition 8.11. (Topological space).

Let (X, τ) be a tuple, where X is any set and τ is a set of subsets of X. Let us to refer to elements of τ as "open sets".

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.

2. Finite unions of closed sets are closed.

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 8.12. ($\mathbb{R}^n$ with the standard topology).

In the section "Topology on $\mathbb{R}$", we were really doing topology on $(\mathbb{R}, \tau_{\text{std}})$, where τ_{std} is the *standard topology on $\mathbb{R}$* generated by $\mathcal{B}_{\text{std}} = \{$open intervals in $\mathbb{R}\}$.

With the help of the following lemma, the above definition of a topology can be seen to be equivalent to the following most common characterization, which makes no mention of a generating cover.

Lemma 8.13. (Logical equivalences regarding open sets).

Let (X, τ) be a tuple, where X is any set and τ is a set of subsets of X. Let us refer to elements of τ as "open sets", and suppose $\mathcal{B}$ is a cover of X. The following statements are equivalent:

1. $\mathcal{B}$ generates τ.

2. Arbitrary unions of open sets are open.

3. The interior point characterization of open sets holds: U is an open set iff every $x \in U$ is contained in some open set that is a subset of U.

Proof.

- $((1) \iff (2))$.

  - $((1) \implies (2))$. Suppose there is a cover $\mathcal{B}$ of X that generates τ. Consider an arbitrary union $U = \cup_\alpha U_\alpha$ of open sets. Since $\mathcal{B}$ generates τ, then each U_α is of the form $U_\alpha = \cup_{\alpha\beta} B_{\alpha\beta}$, where $\{B_{\alpha\beta} \mid \beta \in I\} \subseteq \mathcal{B}$. (Here, I denotes an indexing set). Thus $\cup_\alpha U_\alpha = \cup_\alpha (\cup_\beta B_{\alpha\beta}) = \cup_{\alpha,\beta} B_{\alpha\beta}$, so the union of open sets is in fact a union of sets from the cover $\mathcal{B}$. Since the cover generates τ, any union of sets from the cover $\mathcal{B}$ must be elements of τ, i.e., must be open.

  - $((2) \implies (1))$. Suppose arbitrary unions of open sets are open. We can then take τ to be a cover that generates itself: every element of τ is obviously a union of elements of τ.

- $((2) \iff (3))$. This proof is essentially the same as the proof of Theorem 8.3.

$\square$

Theorem 8.14. (Most common characterization of a topological space).

Let (X, τ) be a tuple, where X is any set and τ is a set of subsets of X. Let us to refer to elements of τ as "open sets". (X, τ) is a topological space iff

3. Arbitrary unions of open sets are open.

4. Finite intersections of open sets are open.

Proof. (3) holds because of the previous lemma. (4) holds because we have the following equivalence due to DeMorgan's laws: "finite unions of open sets are open $\iff$ finite intersections of closed sets are closed". $\square$

Remark 8.15. ($\emptyset, X \in \tau$).

Most definitions of a topological space explicitly require that the empty set and X are open subsets of X. This requirement is implied by the above definition. 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).

Topological 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 8.16. (Refining with interior points).
If $\mathcal{B}$ is a cover of some set X, we say that $\mathcal{B}$ *refines with interior points* iff
$$\forall B_1, B_2 \in \mathcal{B} \quad x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B} \ x \in B_3.$$

Theorem 8.17. (Finite intersections of open sets are open iff the cover "refines with interior points").
Let (X, τ) be a tuple, where X is any set and τ is a set of subsets of X. Let us refer to elements of τ as "open sets".
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.
($\implies$). Assume that finite intersections of open sets are open. We must show that $\mathcal{B}$ refines with interior points. Since elements of $\mathcal{B}$ are open, then if $B_1, B_2 \in \mathcal{B}$, the intersection $B_1 \cap B_2$ 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, $U = \cup_{i \in \{3,4,\dots\}} B_i$. Since x is in U, it must be in B_j for some $j \in \{3, 4, \dots\}$. Relabel B_j to be B_3 to match the syntax of the theorem.

($\impliedby$). Assume $\mathcal{B}$ is a cover that refines with interior points and let $V = \cap_{i=1}^n U_i$ be a finite intersection of open sets. We want to show that V is open. First, assume that V is nonempty- if V is empty then it's open, and we're done.

Because of Lemma 8.13, we can show that V is open by showing that the interior point characterization of open sets holds. So, consider $x \in V$. Then $x \in U_i$ for all i. Each U_i is a union of elements of $\mathcal{B}$, 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 $B_x \in \mathcal{B}$ such that $B_x \subseteq \cap_{i=1}^n B_i$ and $x \in B_x$. Note that since $B_x \in \mathcal{B}$, then B_x is open. Overall, we have shown "$x \in V$ implies that $x \in B_x$, where B_x is open", so the interior point characterization of open sets holds. $\qquad\square$

Definition 8.18. (Basis for a topological space).
A cover $\mathcal{B}$ for X is called a *basis* iff $\mathcal{B}$ refines with interior points, that is, iff
$$\forall B_1, B_2 \in \mathcal{B} \quad x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B} \ x \in B_3.$$
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 8.19. 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. $\qquad\square$

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 8.20. (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 8.21. (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 8.22. (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 8.23. (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 8.24. (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 8.25. (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 8.26. (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 8.27. 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.

$(\Longrightarrow)$. 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.

$(\Longleftarrow)$. 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. $\qquad \square$

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 8.28. (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. $\qquad\square$

Definition 8.29. (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 8.30. (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 8.31. (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 8.32. (Boundary).
Let (X, τ) be a topological space. The *boundary* ∂A of $A \subseteq X$ is defined to be $\partial A := \text{cl}(A) - \text{int}(A)$.

8.2 Continuous functions and homeomorphisms

Limits and continuity for functions $A \subseteq \mathbb{R} \to \mathbb{R}$

Definition 8.33. (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 8.34. (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 8.35. (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 8.36. (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 $f(x_0) = f(x_0)$. $\qquad \square$

Limits and continuity for functions on subsets of topological spaces

Definition 8.37. (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 8.38. (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 8.39. (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 8.40. (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 $f(x_0) \in V$. $\qquad \square$

We now present the classic topological interpretation of a continuous function.

Definition 8.41. (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 8.42. (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. $\qquad \square$

Homeomorphisms

Definition 8.43. (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 8.44. (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.

Theorem 8.45. (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. $\square$

Hausdorff spaces

We need to know what Hausdorff spaces are because manifolds are special types of Hausdorff topological spaces.

Definition 8.46. (Hausdorff space).

Let (X, τ) be a topological space. We say that (X, τ) *is Hausdorff* iff $\forall x \in X \ \forall y \in X \ \exists$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

9

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 9.43 and 9.44, 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.

9.1 Introduction to manifolds

Definition 9.1. [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 9.2. [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 9.3. [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 8.1 of Chapter 8), 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 8.1 of Chapter 8 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 9.4. [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 9.5. [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 8.1 of Chapter 8). 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 9.6. [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.

9.2 Coordinatizing manifolds

Definition 9.7. [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 9.8. [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 9.9. [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 $\widetilde{\mathbf{F}}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{F} : M \to \mathbb{R}^k$.

9.3 Smooth manifolds

Definition 9.10. [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 9.11. [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

$$\mathbf{F} : M \to \mathbb{R}^k \text{ is smooth relative to some chart } (U, \mathbf{x})$$
$$\implies$$
$$\mathbf{F} \text{ is smooth relative to all other charts } (V, \mathbf{y}) \text{ that overlap } (U, \mathbf{x}), \, U \cap V \neq \emptyset.$$

In our present situation, this is not the case. If $(U, \mathbf{x})$ and $(V, \mathbf{y})$ intersect, $U \cap V \neq \emptyset$, 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 9.12. (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 9.13. (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 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^∞. 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 9.14. [Lee, p. 20] Graphs 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 9.15. [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 9.16. [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 9.17. [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 9.18. [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 $\mathbf{p}$.

Remark 9.19. [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.

9.4 Tangent vectors

In multivariable calculus, one can consider tangent vectors that reside in the tangent plane 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 derivative functions" 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 9.20. [Lee, p. 51] (Vectors "anchored" at a point).
Let M be a n-manifold. 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 9.21. (Geometric tangent vectors are naturally isomorphic to "directional derivative functions").
Let $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}^n$, and let $\frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}}$ denote the so-called *directional derivative function* $C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}$ that sends $f \mapsto \frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}}(f)$. For each $\mathbf{p} \in \mathbb{R}^n$, the directional derivative function $\frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}}$ is a natural linear isomorphism between $\mathbb{R}^n_{\mathbf{p}}$ and the vector space of directional derivative functions at $\mathbf{p} \in \mathbb{R}^n$. That is, for each $\mathbf{p} \in \mathbb{R}^n$ we have

$$\mathbb{R}^n_{\mathbf{p}} \cong \{\text{directional derivative functions at } \mathbf{p}\} = \left\{ \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}} \mid \mathbf{v} \in \mathbb{R}^n \right\}.$$

Proof. First, recall from calculus that $\frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}}$ is linear with respect to $\mathbf{v}$ (see Remark 7.15). 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 $\left\{ \frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}} \mid \mathbf{v} \in \mathbb{R}^n \right\}$, the space of directional derivative functions at $\mathbf{p}$. Thus, the vector 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 $\mathbf{v}_{\mathbf{p}} \mapsto \frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}}$ 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 that this implies $\mathbf{v}_{\mathbf{p}} = \mathbf{0}_{\mathbf{p}}$.

At $\mathbf{p} \in \mathbb{R}^n$, the ith coordinate of the zero function is $\frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}}(x^i)$, where $x^i : \mathbb{R}^n \to \mathbb{R}$ denotes the ith coordinate function[3] on $\mathbb{R}^n$. 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 7.15), we have

$$0 = \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}(x^i) = \frac{\partial}{\partial (\sum_j [\mathbf{v}]^j_E \hat{\mathfrak{e}}j)}\Big|_{\mathbf{p}}(x^i) = \sum_j [\mathbf{v}]^j_E \left(\frac{\partial}{\partial \hat{\mathfrak{e}}_j}\Big|_{\mathbf{p}} \right)(x^i) = \sum_j [\mathbf{v}]^j_E \delta^i{}_j = [\mathbf{v}]^i_E.$$

We see $[\mathbf{v}]^i_E = 0$ for all i, so $\mathbf{v} = \mathbf{0}$, which means $\mathbf{v}_{\mathbf{p}} = \mathbf{0}_{\mathbf{p}}$. $\qquad\square$

Now, we investigate the vector space of "directional derivative functions" at a $\mathbf{p} \in \mathbb{R}^n$. While such directional derivative functions can be specified concretely by specifying the vector $\mathbf{v} \in \mathbb{R}^n$ involved in the expression $\frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}}$, 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 function.

Definition 9.22. [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,

$$v_{\mathbf{p}}(fg) = v_{\mathbf{p}}(f)g(\mathbf{p}) + f(\mathbf{p})v_{\mathbf{p}}(g) \text{ for all } f, g \in C^\infty(\mathbb{R}^n \to \mathbb{R}).$$

---

[2]We adapt the argument from [Lee, p. 53] (which is intended for a slightly different purpose in that book).

[3]You might notice that the ith coordinate function on $\mathbb{R}^n$ is also $\phi^{\hat{\mathfrak{e}}_i}$! As there is no need to emphasize dual spaces here, we opt for the x^i notation because it is simpler.

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 function. First, though, we need the following lemma.

Lemma 9.23. [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:

1. If f is a constant function, then $v_{\mathbf{p}}(f) = 0$.

2. If $f(\mathbf{p}) = g(\mathbf{p}) = 0$, then $v_{\mathbf{p}}(fg) = 0$.

Proof.

1. 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.

2. Use the product rule.

$\square$

Theorem 9.24. (Every directional derivative function is a derivation, and every derivation is a directional derivative function).

The set of directional derivative functions at $\mathbf{p} \in \mathbb{R}^n$ is equal to the set of derivations at $\mathbf{p}$:

$$\left\{ \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}} \mid \mathbf{v} \in \mathbb{R}^n \right\} = \{\text{derivations at } \mathbf{p} \in \mathbb{R}^n\}.$$

Proof.

($\subseteq$). Showing that every directional derivative function 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 function at $\mathbf{p}$; we must 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

$$v_{\mathbf{p}}(f) = \left(\frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}} \right)(f) \text{ for all } f \in C^\infty(\mathbb{R}^n \to \mathbb{R}).$$

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 $\mathbf{v_p}$. We set $\mathbf{v_p} = \sum_{i=1}^n v_{\mathbf{p}}(x^i)\hat{\mathfrak{e}}_i$, where here $x^i : \mathbb{R}^n \to \mathbb{R}$ denotes the ith coordinate function on $\mathbb{R}^n$. This choice of $\mathbf{v_p}$ makes more sense in hindsight after reading Theorem 9.44. (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

$$v_{\mathbf{p}}(f) = v_{\mathbf{p}}\left(\sum_{i,j=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} (x^i - p^i) \right).$$

(Notice that we are abusing notation slightly, since we are writing the argument of $v_{\mathbf{p}}$ on the right side so that it appears to be the evaluation of a function at $\mathbf{x}$ rather than the function itself. Thus, the x^i on the right side still represents a coordinate function and not a real number).

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

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}(v_{\mathbf{p}}(x^i) - v_{\mathbf{p}}(p^i)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} v_{\mathbf{p}}(x^i) = (\nabla \mathbf{F})_{\mathbf{p}} \cdot \mathbf{v_p} = \frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}}.$$

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 function. $\qquad\square$

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 9.25. ($\mathbb{R}^n_{\mathbf{p}} \cong \{$directional derivative functions at $\mathbf{p}\} = \{$derivations at $\mathbf{p} \in \mathbb{R}^n\}$).
For any $\mathbf{p} \in \mathbb{R}^n$, we have

$$\boxed{\mathbb{R}^n_{\mathbf{p}} \quad \overset{\mathbf{v_p} \mapsto \frac{\partial}{\partial \mathbf{v}}\big|_{\mathbf{p}},\ \mathbf{v} \in \mathbb{R}^n}{\cong} \quad \{\text{directional derivative functions at } \mathbf{p}\} = \{\text{derivations at } \mathbf{p} \in \mathbb{R}^n\}}$$

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 9.21 and Theorem 9.24 $\qquad\square$

Definition 9.26. (Tangent space to $\mathbb{R}^n$ at a point).
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 9.27. [Lee, p. 54] (A basis of $T_{\mathbf{p}}(\mathbb{R}^n)$).
The directional derivative functions $\left\{ \frac{\partial}{\partial x^i}\big|_{\mathbf{p}} \right\}_{i=1}^{n} = \left\{ \frac{\partial}{\partial \hat{\mathfrak{e}}^i}\big|_{\mathbf{p}} \right\}_{i=1}^{n}$ form a basis for $T_{\mathbf{p}}(\mathbb{R}^n)$.

Proof. We proved this in the proof of Theorem 9.21. $\qquad\square$

Tangent vectors on manifolds

Since the notion "tangent vector to $\mathbf{p} \in \mathbb{R}^n$" does not rely upon coordinates, we can extend the notion to smooth manifolds.

Definition 9.28. [Lee, p. 54] (Tangent space to a manifold at a point).
Let M be a smooth n-manifold. 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 following 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 vector space over $\mathbb{R}$ 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 $\mathbf{p}$)*.

Remark 9.29. (On interface vs. implementation; geometric interpretation of tangent vectors).
A helpful perspective to remember is that the particular details what exactly a tangent vector $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ is are just that- *implementation* details. There are in fact other implementations; there are other definitions of $T_{\mathbf{p}}(M)$, different from the "linear derivation at $\mathbf{p}$" definition we gave, that establish vector spaces isomorphic to our $T_{\mathbf{p}}(M)$!

When thinking intuitively about what a tangent vector *is*, the *interface* to the tangent vector notion that is the interpretation of elements of $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ as an arrow anchored at $\mathbf{p}$ is what's most relevant.

Theorem 9.30. [Lee, p. 54] (Properties of derivations at $\mathbf{p} \in M$).
Let M be a smooth n-manifold, 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 9.23. $\square$

Definition 9.31. [War, p. 14, 15] (Basis of $T_{\mathbf{p}}(M)$ and its abuse of notation).
Let M be a smooth n-manifold, let $(U, \mathbf{x})$ be a smooth chart on M, and let $\mathbf{p} \in M$. In an abuse of notation, we define $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}$ to be the derivations at $\mathbf{p} \in U \subseteq M$ for which

$$\frac{\partial}{\partial x^i}\bigg|_{\mathbf{p}}(f) := \frac{\partial \widetilde{f}_{(U,\mathbf{x})}}{\partial x^i}\bigg|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(f \circ \mathbf{x}^{-1})}{\partial x^i}\bigg|_{\mathbf{x}(\mathbf{p})}.$$

(In the above, $\widetilde{f}_{(U,\mathbf{x})}$ 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 side are directional derivatives taking in smooth functions on $\mathbb{R}^n$ as their arguments.

We also define the notation

$$\frac{\partial f}{\partial x^i} := \frac{\partial}{\partial x^i}\bigg|_{\mathbf{p}}(f) = \frac{\partial(f \circ \mathbf{x}^{-1})}{\partial x^i}\bigg|_{\mathbf{x}(\mathbf{p})}.$$

Proof. We need to show (1) that the $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}$ of the left side of the above definition are derivations at $\mathbf{p} \in M$ and (2) that $\left\{ \frac{\partial}{\partial x^i}\big|_{\mathbf{p}} \right\}_{i=1}^n$ is a basis of $T_{\mathbf{p}}(M)$.

1. We will put an "LS" 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 side of the condition in the above definition. All other occurrences of $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}$ in this proof are partial derivatives, $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}} = \frac{\partial}{\partial \hat{\mathfrak{e}}_i}\big|_{\mathbf{p}}$.

   We need to show that $\underbrace{\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}}_{\text{LS}}$ 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:

$$\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LS}}(fg) = \frac{\partial(fg \circ \mathbf{x}^{-1})}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\Big|_{\mathbf{x}^{-1}(\mathbf{x}(\mathbf{p}))} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$$

$$= \left(\frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial g}{\partial x^i}\Big|_{\mathbf{p}}\right)\frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$$

$$= \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial g}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$$

$$= \frac{\partial f}{\partial x^i}\Big|_{\mathbf{x}^{-1}(\mathbf{x}(\mathbf{p}))} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial g}{\partial x^i}\Big|_{\mathbf{x}^{-1}(\mathbf{x}(\mathbf{p}))} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$$

$$= \frac{\partial(f \circ \mathbf{x}^{-1})}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial(g \circ \mathbf{x}^{-1})}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$$

$$= \left(\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LS}}\right)(f)\, g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p})) \left(\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LS}}\right)(g).$$

In the above, we used the chain and product rules and then reversed the product and chain rules.

2. Now we show $\left\{\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LS}}\right\}_{i=1}^n \subseteq T_{\mathbf{p}}(M)$ is a basis of $T_{\mathbf{p}}(M)$. Linear independence of this set follows from

linear independence of $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}_{i=1}^n \subseteq T_{\mathbf{p}}(\mathbb{R}^n)$. To show the set spans $T_{\mathbf{p}}(M)$, consider $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. Recall that $T_{\mathbf{p}}(M)$ is a subset of functions $C^\infty(M \to \mathbb{R}) \to \mathbb{R}$, so $v_{\mathbf{p}}$ is such a function; $v_{\mathbf{p}} : C^\infty(M \to \mathbb{R}) \to \mathbb{R}$. The homeomorphism $\mathbf{x} : M \to \mathbb{R}^n$ induces a $\widetilde{v}_{\mathbf{x}(\mathbf{p})} : C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}$ defined by $\widetilde{v}_{\mathbf{x}(\mathbf{p})}(\widetilde{f}) = v_{\mathbf{p}}(\widetilde{f} \circ \mathbf{x})$. It's straightforward to check that $\widetilde{v}_{\mathbf{x}(\mathbf{p})}$ is a derivation at $\mathbf{x}(\mathbf{p}) \in \mathbb{R}^n$, so $\widetilde{v}_{\mathbf{x}(\mathbf{p})} \in T_{\mathbf{p}}(\mathbb{R}^n)$. Since $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}\right\}_{i=1}^n$ spans $T_{\mathbf{x}(\mathbf{p})}(\mathbb{R}^n)$, we have $\widetilde{v}_{\mathbf{x}(\mathbf{p})} = \sum_{i=1}^n c^i \frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$ for some $c^i \in \mathbb{R}$. Evaluating both sides of this equation on $\widetilde{f} \in C^\infty(\mathbb{R}^n \to \mathbb{R})$ and recalling the definition of $\widetilde{v}_{\mathbf{x}(\mathbf{p})}$, we have

$$\widetilde{v}_{\mathbf{x}(\mathbf{p})}(\widetilde{f}) = v_{\mathbf{p}}(\widetilde{f} \circ \mathbf{x}) = \sum_{i=1}^n c^i \frac{\partial \widetilde{f}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}.$$

Substitute $f := \widetilde{f} \circ \mathbf{x}^{-1} : M \to \mathbb{R}^n$ to obtain

$$v_{\mathbf{p}}(f) = \sum_{i=1}^n c^i \frac{\partial(f \circ \mathbf{x}^{-1})}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \sum_{i=1}^n c^i \underbrace{\frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LS}}.$$

Thus

$$v_{\mathbf{p}} = \sum_{i=1}^n c^i \underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LS}},$$

so $\left\{\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LS}}\right\}_{i=1}^n$ spans $T_{\mathbf{p}}(M)$.

$\square$

The differential

Now that we have defined tangent vectors to a manifold, we can define a generalization of the directional derivative of a function $\mathbb{R}^n \to \mathbb{R}^m$: the *differential of a function between smooth manifolds*.

Definition 9.32. (Differential of a smooth function).

Let M and N be smooth n- and m- manifolds and let $(U, \mathbf{x})$ and $(V, \mathbf{y})$ be smooth charts on M and N. Consider a smooth function $\mathbf{F} : U \subseteq M \to V \subseteq N$. We define the *differential $d\mathbf{F_p} : T_\mathbf{p}(M) \to T_{\mathbf{F(p)}}(N)$ of $\mathbf{F}$ at* $\mathbf{p} \in U$ to be the function $T_\mathbf{p}(M) \to T_{\mathbf{F(p)}}(N)$ whose matrix relative to the bases $\left\{ \frac{\partial}{\partial x^i} \big|_\mathbf{p} \right\}_{i=1}^n$ and $\left\{ \frac{\partial}{\partial y^i} \big|_{\mathbf{F(p)}} \right\}_{i=1}^n$ for $T_\mathbf{p}(M)$ and $T_{\mathbf{F(p)}}(N)$ is the Jacobian matrix of the coordinate representation $\widetilde{\mathbf{F}}_{(U,\mathbf{x}),(V,\mathbf{y})} : \mathbb{R}^n \to \mathbb{R}^m$ of $\mathbf{F}$:

$$\boxed{\begin{array}{c} \text{For all smooth charts } (U, \mathbf{x}) \text{ and } (V, \mathbf{y}) \text{ and all } \mathbf{p} \in U, \text{ the matrix of } d\mathbf{F_p} \text{ is} \\[2mm] \dfrac{\partial \widetilde{\mathbf{F}}_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial \mathbf{x}} \Big|_\mathbf{p} = \Big(\dfrac{\partial \widetilde{F}^i_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^j} \Big|_\mathbf{p} \Big) \end{array}}$$

Remark 9.33. (Is the differential well-defined?).

It is not obvious that $d\mathbf{F_p}$ is the unique function that satisfies the above definition. Fortunately, we will be able to conclude this after the following derivation and theorem.

Derivation 9.34. (Coordinate-free formula for the differential).

Assume the hypotheses of the previous definition. To determine a coordinate-free[4] formula for $d\mathbf{F_p}$, we will consider how $d\mathbf{F_p}$ acts on a basis tangent vector $\frac{\partial}{\partial x^i} \big|_\mathbf{p}$ and extend whatever result we get with the linearity of $d\mathbf{F_p}$.

Recalling from Theorem 2.109 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

$$d\mathbf{F_p} \Big(\frac{\partial}{\partial x^i} \Big|_\mathbf{p} \Big) = \sum_{j=1}^m a_i^j \frac{\partial}{\partial y^j} \Big|_{\mathbf{F(p)}},$$

where a_j^i is the $\frac{i}{j}$ entry of the matrix of $d\mathbf{F_p}$ relative to the bases $\left\{ \frac{\partial}{\partial x^i} \big|_\mathbf{p} \right\}$ and $\left\{ \frac{\partial}{\partial y^i} \big|_{\mathbf{F(p)}} \right\}$. Since we imposed $a_j^i = \frac{\partial \widetilde{F}^i_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^j} \Big|_\mathbf{p}$, then

$$d\mathbf{F_p} \Big(\frac{\partial}{\partial x^i} \Big|_\mathbf{p} \Big) = \sum_{j=1}^m a_i^j \frac{\partial}{\partial y^j} \Big|_{\mathbf{F(p)}} = \sum_{j=1}^m \frac{\partial \widetilde{F}^j_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^i} \Big|_\mathbf{p} \frac{\partial}{\partial y^j} \Big|_{\mathbf{F(p)}}.$$

So far, we have shown

$$d\mathbf{F_p} \Big(\frac{\partial}{\partial x^i} \Big|_\mathbf{p} \Big) = \sum_{j=1}^m \frac{\partial \widetilde{F}^j_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^i} \Big|_\mathbf{p} \frac{\partial}{\partial y^j} \Big|_{\mathbf{F(p)}}.$$

At the beginning, we claimed that $d\mathbf{F_p} : \mathbf{T}_p(M) \to \mathbf{T}_{\mathbf{F(p)}}(N)$; that is, we claimed that $d\mathbf{F_p}$ 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 $d\mathbf{F_p} \left(\frac{\partial}{\partial x^i} \right)$, we must apply this new derivation to a smooth function $f \in C^\infty(V \subseteq N \to \mathbb{R})$:

$$d\mathbf{F_p} \Big(\frac{\partial}{\partial x^i} \Big)(f) = \Big(\sum_{j=1}^m \frac{\partial \widetilde{F}^j_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^i} \Big|_\mathbf{p} \frac{\partial}{\partial y^j} \Big|_{\mathbf{F(p)}} \Big)(f) = \sum_{j=1}^m \Big[\Big(\frac{\partial \widetilde{F}^j_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^i} \Big|_\mathbf{p} \frac{\partial}{\partial y^j} \Big|_{\mathbf{F(p)}} \Big)(f) \Big]$$

$$= \sum_{j=1}^m \frac{\partial \widetilde{F}^j_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^i} \Big|_\mathbf{p} \frac{\partial f}{\partial y^j} \Big|_{\mathbf{F(p)}} = \sum_{j=1}^m \frac{\partial \widetilde{F}^j_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^i} \Big|_\mathbf{p} \frac{\partial \widetilde{f}_{(V,\mathbf{y})}}{\partial y^j} \Big|_{\mathbf{F(p)}}.$$

---

[4] By "coordinate-free", we mean "independent of a choice of bases for $T_\mathbf{p}(M)$ and $T_\mathbf{p}(N)$".

In this last equation, since $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}$ and $\frac{\partial}{\partial y^i}\big|_{\mathbf{F(p)}}$ are operating on functions $\mathbb{R}^n \to \mathbb{R}^m$, they are the usual partial derivatives $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}} = \frac{\partial}{\partial \hat{\mathbf{c}}_i}\big|_{\mathbf{p}}$ and $\frac{\partial}{\partial y^i}\big|_{\mathbf{F(p)}} = \frac{\partial}{\partial \hat{\mathbf{c}}_i}\big|_{\mathbf{F(p)}}$ (recall Definition 9.31). So, we can apply the chain rule:

$$\sum_{j=1}^{m} \frac{\partial \widetilde{F}^j_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^i}\bigg|_{\mathbf{p}} \frac{\partial \widetilde{f}_{(V,\mathbf{y})}}{\partial y^j}\bigg|_{\mathbf{F(p)}} = \sum_{j=1}^{m} \frac{\partial \widetilde{f}_{(V,\mathbf{y})}}{\partial y^j}\bigg|_{\mathbf{F(p)}} \frac{\partial \widetilde{F}^j_{(U,\mathbf{x}),(V,\mathbf{y})}}{\partial x^i}\bigg|_{\mathbf{p}} = \frac{\partial}{\partial x^i}\bigg|_{\mathbf{p}}(\widetilde{f}_{(V,\mathbf{y})} \circ \widetilde{\mathbf{F}}_{(U,\mathbf{x}),(V,\mathbf{y})}).$$

The above shows

$$d\mathbf{F}_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}(\widetilde{f}_{(V,\mathbf{y})} \circ \widetilde{\mathbf{F}}_{(U,\mathbf{x}),(V,\mathbf{y})}).$$

We have $\widetilde{f}_{(V,\mathbf{y})} \circ \widetilde{\mathbf{F}} = (f \circ \mathbf{y}^{-1}) \circ (\mathbf{y} \circ \mathbf{F} \circ \mathbf{x}^{-1}) = f \circ \mathbf{F} \circ \mathbf{x}^{-1} = (f \circ \mathbf{F})_{(U,\mathbf{x})}$, so

$$d\mathbf{F}_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}((f \circ \mathbf{F})_{(U,\mathbf{x})}) = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}(f \circ \mathbf{F}).$$

In all,

$$d\mathbf{F}_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}(f \circ \mathbf{F}).$$

(This last $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}$ is no longer one of the usual partial derivatives $\frac{\partial}{\partial \hat{\mathbf{c}}^i}\big|_{\mathbf{p}}$; recall from Definition 9.31) that it is defined by $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}}(g) = \frac{\partial \widetilde{g}_{(U,\mathbf{x})}}{\partial x^i}\big|_{\mathbf{p}})$.

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 $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$:

$$d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})(f) = v_{\mathbf{p}}(f \circ \mathbf{F}).$$

This characterization of the differential is coordinate-free, and therefore provides an easy way to define the differential without mentioning any smooth charts. We do this in the next definition.

Theorem 9.35. (Formula for the differential of a smooth function).

Let M and N be smooth n- and m- dimensional manifolds, and let $\mathbf{p} \in M$. The differential $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{\mathbf{F(p)}}(N)$ satisfies

$$\boxed{d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})(f) = v_{\mathbf{p}}(f \circ \mathbf{F}) \text{ for all } v_{\mathbf{p}} \in T_{\mathbf{p}}(M) \text{ and } f \in C^{\infty}(N \to \mathbb{R})}$$

Theorem 9.36. (The differential is well-defined).

The function $f \mapsto v_{\mathbf{p}}(f \circ \mathbf{F})$ is obviously well-defined for all $\mathbf{p} \in M$, so the function $v_{\mathbf{p}} \mapsto d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})$ is well-defined for all $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. Since $dF_{\mathbf{p}}(v_{\mathbf{p}})$ is well defined for all $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$, then the function $d\mathbf{F}_{\mathbf{p}}$, which sends $v_{\mathbf{p}} \mapsto dF_{\mathbf{p}}(v_{\mathbf{p}})$, is well-defined.

Theorem 9.37. [Lee, p. 55] (Properties of differentials). Let M, N, and P be smooth manifolds, 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). $d(\mathbf{G} \circ \mathbf{F})_{\mathbf{p}} = d\mathbf{G}_{\mathbf{F(p)}} \circ d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{(\mathbf{G} \circ \mathbf{F})(\mathbf{p})}(P)$.

- (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.

- (Inverse differential). If $\mathbf{F}$ is a diffeomorphism, then $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{\mathbf{F(p)}}(N)$ is a linear isomorphism, and $(d\mathbf{F}_{\mathbf{p}})^{-1} = d(\mathbf{F}^{-1})_{\mathbf{F(p)}}$.

Proof. See [Lee] for the proof. $\qquad \square$

Theorem 9.38. [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:

$$
\begin{array}{ccc}
T_{\mathbf{p}}(M) & \xrightarrow{\ v_{\mathbf{p}} \mapsto df_{\mathbf{p}}(v_{\mathbf{p}})\ } & T_{f(\mathbf{p})}(\mathbb{R}) \\
& \searrow{\scriptstyle \widetilde{d}f_{\mathbf{p}}} & \downarrow{\scriptstyle \left(w = \sum_{i=1}^{1} w(x^i)\hat{\imath}_i\right) \mapsto w(x^1)} \\
& & \mathbb{R}
\end{array}
$$

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

$$\widetilde{d}f_{\mathbf{p}}(v_{\mathbf{p}}) = df_{\mathbf{p}}(v_{\mathbf{p}})(x^1) = v_{\mathbf{p}}(x^1 \circ f).$$

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 $\mathbb{R}$. This final identification [5] gives

$$\widetilde{d}f_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f).$$

In practice, we write $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ to mean $\widetilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$. So, the above characterizing condition of the differential of $f : M \to \mathbb{R}$ *that is induced by the identification* $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ is restated as

$$\boxed{df_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)}$$

Theorem 9.39. [Lee, p. 281] (Differential of a smooth function $M \to \mathbb{R}$ in a smooth chart).

Let M be a smooth n-manifold and let $(U, \mathbf{x})$ be a smooth chart on M. 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

$$\boxed{df_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\bigg|_{\mathbf{p}} dx^i_{\mathbf{p}} \ \text{ for all } \mathbf{p} \in U}$$

Recall from Definition 9.31 that in the above, $\frac{\partial f}{\partial x^i}\big|_{\mathbf{p}}$ means $\frac{\partial \widetilde{f}_{(U,\mathbf{x})}}{\partial \hat{\imath}^i}\big|_{\mathbf{p}}$.

Proof. Notice that $(\mathbf{v} \mapsto df_{\mathbf{p}}(\mathbf{v})) = df_{\mathbf{p}}$ is an element of $T_{\mathbf{p}}(M)^*$. (We will explore this fact more later, in Section 9.5). Since $T_{\mathbf{p}}(M)$ is an n-dimensional vector space, then $T_{\mathbf{p}}(M)^*$ is also n-dimensional, and we can express $df_{\mathbf{p}}$ as

$$df_{\mathbf{p}} = \sum_{i=1}^{n} C_i(\mathbf{p}) \lambda^i_{\mathbf{p}}$$

where each $C_i : U \to \mathbb{R}$ is a smooth function and $\lambda^i \in T_{\mathbf{p}}(M)^*$.

To determine the C_i, recall that Theorem 4.16 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

---

[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.

$$C_i(\mathbf{p}) = df_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big) = \Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}.$$

Thus $df_{\mathbf{p}}$ is

$$df_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \lambda^i_{\mathbf{p}}.$$

It remains to determine the $\lambda^i_{\mathbf{p}}$. To do so, consider the above and substitute the coordinate function $x^i : U \to \mathbb{R}$ in for f. We have

$$dx^i_{\mathbf{p}} = \sum_{j=1}^{n} \frac{\partial x^i}{\partial x^j}\Big|_{\mathbf{p}} \lambda|^j_{\mathbf{p}} = \sum_{j=1}^{n} \frac{\partial}{\partial \hat{\mathfrak{e}}^j}(x^i \circ \mathbf{x}^{-1})\Big|_{\mathbf{p}} \lambda|^j_{\mathbf{p}} = \sum_{j=1}^{n} \frac{\partial \widetilde{x}^i}{\partial \hat{\mathfrak{e}}^j}\Big|_{\mathbf{p}} \lambda^j|_{\mathbf{p}},$$

where $\widetilde{x}^i : \mathbb{R}^n \to \mathbb{R}$ is the ith coordinate function on $\mathbb{R}^n$. Since $\frac{\partial \widetilde{x}^i}{\partial \hat{\mathfrak{e}}^j}\Big|_{\mathbf{p}} = \delta^i{}_j$, the sum simplifies to $\lambda^i_{\mathbf{p}}$, and we get $dx^i_{\mathbf{p}} = \lambda^i_{\mathbf{p}}$. $\qquad\square$

Remark 9.40. (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 $T_{\mathbf{p}}(\mathbb{R}^n) \cong \mathbb{R}^n_{\mathbf{p}}$ 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 $dx^i_{\mathbf{p}}$ is the ith coordinate function on $T_{\mathbf{p}}(M)$, it follows[6] that $\widetilde{d}x^i_{\mathbf{p}}$ 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

$$\widetilde{d}f_{\mathbf{p}}(\mathbf{v}_{\mathbf{p}}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} \widetilde{d}x^i(\mathbf{v}_{\mathbf{p}}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} [\mathbf{v}_{\mathbf{p}}]^i_{\hat{\mathfrak{e}}} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v}_{\mathbf{p}} = \frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}},$$

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 9.32) 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 7.13 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 $m = 1$).

---

[6]Let V_1, V_2, W be finite-dimensional vector spaces and suppose $\mathbf{g} : V_1 \to W$ is a linear isomorphism. Then the matrix of a linear function $\mathbf{f} : V_1 \to W$ relative to a basis E of V_1 is the same as the matrix of the induced function $(\widetilde{\mathbf{f}} = \mathbf{f} \circ \mathbf{g}^{-1}) : V_2 \to W$ relative to the basis $\mathbf{g}(E)$ of V_2: the two matrices are $\mathbf{f}(E)$ and $(\mathbf{f} \circ \mathbf{g}^{-1})(\mathbf{g}(E))$, respectively. We are in the situation $V_1 = T_{\mathbf{p}}(\mathbb{R}^n)$, $V_2 = \mathbb{R}^n_{\mathbf{p}}$, $W = \mathbb{R}$, $\mathbf{f} = dx^i_{\mathbf{p}}$, $E = \{(\partial/\partial x^i)_{\mathbf{p}}\}^n_{i=1}$, $(\partial/\partial x^i)_{\mathbf{p}} \overset{\mathbf{g}}{\mapsto} \hat{\mathfrak{e}}^i_{\mathbf{p}}$.

9.5 Cotangent vectors

Remark 9.41. (Intuition for differentials).

Let M be an smooth n-manifold, let $(U, \mathbf{x})$ be a smooth chart on M, and let $\mathbf{p} \in U$. When we proved Theorem 9.39, we quickly applied the fact that we can compute the ith coordinate of $v_{\mathbf{p}}$ relative to the basis $\left\{ \left. \frac{\partial}{\partial x^i} \right|_{\mathbf{p}} \right\}$ for $T_{\mathbf{p}}(M)$ as

$$[v_{\mathbf{p}}]^i_{\left\{ \left. \frac{\partial}{\partial x^i} \right|_{\mathbf{p}} \right\}_{i=1}^n} = dx^i_{\mathbf{p}}(v_{\mathbf{p}}).$$

Let us think about this fact for a second. Recall that in calculus, the expression "dx^i" is thought of as an infinitesimal displacement in the $\hat{\mathfrak{e}}_i$ direction. On manifolds, we think of tangent vectors- elements of $T_{\mathbf{p}}(M)$- as representing the infinitesimal displacements; a tangent vector at $\mathbf{p}$ is thought of as an infinitesimal displacement originating at $\mathbf{p}$. The differential $dx^i_{\mathbf{p}}$ then formalizes the idea of an infinitesimal displacement in the direction of the ith basis tangent vector, since, according to the above, applying $dx^i_{\mathbf{p}}$ to a tangent vector $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ gives the component of that tangent vector in the direction of the ith basis tangent vector.

Derivation 9.42. (Cotangent space).

Let M be a smooth manifold, consider a smooth chart $(U, \mathbf{x})$ on M, and let $\mathbf{p} \in U$. What set do the differentials $dx^i_{\mathbf{p}}$ reside in? Well, $dx^i_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$, so $dx^i_{\mathbf{p}} \in T_{\mathbf{p}}(M)^*$.

Additionally, notice that since applying $dx^i_{\mathbf{p}}$ to $v_{\mathbf{p}}$ yields the ith coordinate of $v_{\mathbf{p}}$ in the $\left\{ \left. \frac{\partial}{\partial x^i} \right|_{\mathbf{p}} \right\}_{i=1}^n$ basis, the set $\{dx^i_{\mathbf{p}}\}_{i=1}^n$ must[7] be the dual basis for $T_{\mathbf{p}}(M)^*$ that is induced by the basis $\left\{ \left. \frac{\partial}{\partial x^i} \right|_{\mathbf{p}} \right\}_{i=1}^n$ for $T_{\mathbf{p}}(M)$.

The set $T_{\mathbf{p}}(M)^*$ is called the *cotangent space to M at $\mathbf{p}$*. We denote it by $T^*_{\mathbf{p}}(M)$ for convenience.

9.6 Coordinates of tangent and cotangent vectors

Theorem 9.43. (Induced bases in a chart).

Let M be a smooth n-manifold 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 $\left\{ \left. \frac{\partial}{\partial x^i} \right|_{\mathbf{p}} \right\}$.

2. The dual basis for $T^*_{\mathbf{p}}(M) = T_{\mathbf{p}}(M)^* = C^\infty(U \subseteq M \to \mathbb{R})^{**}$ induced by $\left\{ \left. \frac{\partial}{\partial x^i} \right|_{\mathbf{p}} \right\}$ 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 9.38 to see happens to d when we identify $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$).

Proof. Recall from the proof of Theorem 9.38 that when using the notation $\left. \frac{\partial}{\partial x^j} \right|_{\mathbf{p}} := \left. \frac{\partial (\cdot \circ \mathbf{x}^{-1})}{\partial \hat{\mathfrak{e}}^j} \right|_{\mathbf{p}}$, we have $\left. \frac{\partial x^i}{\partial x^j} \right|_{\mathbf{p}} = \delta^i{}_j$.

1. We have $\left(\left. \frac{\partial}{\partial x^i} \right|_{\mathbf{p}} \right)(x^j) = \left. \frac{\partial x^i}{\partial x^j} \right|_{\mathbf{p}} = \delta^i{}_j$.

2. We know $df_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)$ from Theorem 9.38 so $dx^i_{\mathbf{p}} \left(\left. \frac{\partial}{\partial x^j} \right|_{\mathbf{p}} \right) = \left. \frac{\partial x^i}{\partial x^j} \right|_{\mathbf{p}} = \delta^i{}_j$.

$\square$

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.

---

[7]If two functions send $v_{\mathbf{p}}$ to its ith coordinate, then, since these functions are linear, they have the same matrix and are thus the same function. The functions that are $dx^i_{\mathbf{p}}$ and the ith dual basis vector for $T_{\mathbf{p}}(M)^*$ are the same for this reason.

Theorem 9.44. (Coordinates of tangent vectors and cotangent vectors).

Theorem 4.16 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 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

$$\left([v_{\mathbf{p}}]_{\left\{ \frac{\partial}{\partial x^j} \big|_{\mathbf{p}} \right\}_{j=1}^n}\right)^i = v_{\mathbf{p}}(x^i) = \phi_{x^i}(v_{\mathbf{p}})$$

$$\left([\phi_{\mathbf{p}}]_{\left\{ dx^j |_{\mathbf{p}} \right\}_{j=1}^n}\right)_i = \phi_{\mathbf{p}}\left(\frac{\partial}{\partial x^i} \Big|_{\mathbf{p}} \right)$$

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 3.28 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 $dx_{\mathbf{p}}^i$ are "the same": $dx_{\mathbf{p}}^i = (\phi_{x^i})|_{\mathbf{p}}$.

Remark 9.45. (Covariance and contravariance with tangent vectors and tangent covectors).

Recall the convention of "covariance and contravariance" from Definition 3.22. 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 $dx_{\mathbf{p}}^i$ 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 $dx_{\mathbf{p}}^i$, all is well and good- an upper index is used on a tangent covector- so no retroactive argument is necessary.

Remark 9.46. (Uses of the notation x^i).

Let M be a smooth n-manifold, and consider a smooth chart $(U, \mathbf{x})$ about $\mathbf{p} \in M$. You have likely noticed that the collection of symbols "x^i" appears in notation in various ways. A summary of all the ways x^i is used is as follows. When x^i appears in a "numerator" (e.g. $\frac{\partial x^i}{\partial x^j}$), as an argument of a differential (e.g. dx^i), or "by itself" (e.g. x^i), then x^i is the ith coordinate function of $\mathbf{x}$. 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 $T_{\mathbf{p}}(M)$.

Theorem 9.47. [Lee, p. 64] (Change of coordinates for tangent vectors in terms of basis vectors of $T_{\mathbf{p}}(M)$).

Theorem 2.109 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

$$[\mathbf{F}]_E = \left(\left[\left. \frac{\partial}{\partial y^1} \right|_{\mathbf{p}} \right]_E \quad \cdots \quad \left[\left. \frac{\partial}{\partial y^n} \right|_{\mathbf{p}} \right]_E \right) = \left(\begin{pmatrix} \left. \frac{\partial}{\partial y^1} \right|_{\mathbf{p}} (x^1) \\ \vdots \\ \left. \frac{\partial}{\partial y^1} \right|_{\mathbf{p}} (x^n) \end{pmatrix} \quad \cdots \quad \begin{pmatrix} \left. \frac{\partial}{\partial y^n} \right|_{\mathbf{p}} (x^1) \\ \vdots \\ \left. \frac{\partial}{\partial y^n} \right|_{\mathbf{p}} (x^n) \end{pmatrix} \right) = \begin{pmatrix} \left. \frac{\partial x^1}{\partial y^1} \right|_{\mathbf{p}} & \cdots & \left. \frac{\partial x^1}{\partial y^n} \right|_{\mathbf{p}} \\ \vdots & & \vdots \\ \left. \frac{\partial x^n}{\partial y^1} \right|_{\mathbf{p}} & \cdots & \left. \frac{\partial x^n}{\partial y^n} \right|_{\mathbf{p}} \end{pmatrix}.$$

Don't forget that $\left. \frac{\partial x^i}{\partial y^j} \right|_{\mathbf{p}} = \left. \frac{\partial (x^i \circ \mathbf{y}^{-1})}{\partial y^j} \right|_{\mathbf{p}}$, where y^i is the ith coordinate function of $\mathbf{y}$.

Applying the fact $\mathbf{f}_i = \sum_{j=1}^n ([\mathbf{F}]_E)_{ji} \mathbf{e}_j$ from above, we have

$$\left. \frac{\partial}{\partial y^i} \right|_{\mathbf{p}} = \sum_{j=1}^n \left. \frac{\partial x^j}{\partial y^i} \right|_{\mathbf{p}} \left. \frac{\partial}{\partial x^j} \right|_{\mathbf{p}}$$

(If you are careful to remember that $\left. \frac{\partial}{\partial y^i} \right|_{\mathbf{p}} = \frac{\partial (\cdot \circ \mathbf{y}^{-1})}{\partial y^i}$, then you can prove the above just by using the chain rule).

[... obtain result for general $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ ...]

9.7   Vector fields and covector fields

The tangent, cotangent, and tensor bundles

Definition 9.48. [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[8] union.

Theorem 9.49. [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 9.50. [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 9.51. [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 9.52. [Lee, p. 316, 317] (Tangent bundle and cotangent bundle special cases of the tensor bundle).
We have $T_0^p(T(M)) = \bigsqcup_{\mathbf{p} \in M} (T_{\mathbf{p}}(M))^{\otimes p}$ and $T_q^0(T(M)) = \bigsqcup_{\mathbf{p} \in M} (T^*_{\mathbf{p}}(M))^{\otimes p}$. So in particular, $T_0^1(T(M)) = T(M)$, $T_1^0(T(M)) = T^*(M)$.

Definition 9.53. [Lee, p. 174, 276, 316, 317] (Vector fields, covector fields, and (p, q) tensor fields).
Let M be a smooth n-manifold 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.

---

[8]Note that a disjoint union is *not* a "pairwise disjoint union". We say $A = \bigsqcup_\alpha A_\alpha$ iff $A = \bigcup_\alpha A_\alpha$ and $\cap_\alpha A_\alpha = \emptyset$.

Frames and coframes

Definition 9.54. [Lee, p. 178, 278] (Frames and coframes).

Let M be a smooth n-manifold and let $(U, \mathbf{x})$ be a smooth chart on 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.

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 9.55. [Lee, p. 176, 278] (Coordinate vector fields and coordinate covector fields).

Let M be a smooth n-manifold and let $(U, \mathbf{x})$ be a smooth chart on M.

The smooth[9] 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 $\{dx_\mathbf{p}^i\}_{i=1}^n$ 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[10] covector field defined by $\mathbf{p} \mapsto dx_\mathbf{p}^i$ is denoted (in an abuse of notation) by dx^i and is called the *ith local coordinate covector field (on* $(U, \mathbf{x})$*)*.

Definition 9.56. [Lee, p. 178, 278] (Coordinate frames and coordinate coframes).

Let M be a smooth manifold and let $(U, \mathbf{x})$ be a smooth chart on M.

The smooth[11] 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[12] 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 9.57. [Lee, p. 278] (Duality of frames and coframes).

Let M be a smooth n-manifold 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.

Ratio of differentials

At this point we are able to formalize the idea of taking the ratio of differentials- one of the most fundamental intuitions of single-variable calculus.

Lemma 9.58. (When M is one-dimensional, the differential becomes the single-variable derivative).

Let M be a smooth one-dimensional manifold and let (U, x) be a smooth chart on M. If $f : M \to \mathbb{R}$ is a smooth function, then for every $\mathbf{p} \in M$ the cotangent space $T_\mathbf{p}^*(M)$ is one-dimensional, so $df_\mathbf{p} = \sum_{i=1}^1 \frac{\partial f}{\partial x^i}\big|_\mathbf{p} dx_\mathbf{p}^i = \tilde{f}_{(U,\mathbf{x})}\,'(x(\mathbf{p}))dx_\mathbf{p}$ can be identified with the scalar $\tilde{f}_{(U,\mathbf{x})}\,'(x(\mathbf{p})) \in \mathbb{R}$. In other words, the induced differential $\widetilde{d}$ under these identifications is $\widetilde{d} = D(\cdot \circ x^{-1})$.

Theorem 9.59. (Ratio of differentials).

Just noticed that $\frac{\partial f}{\partial x^i} = \frac{\partial \tilde{f}}{\partial \hat{\mathbf{e}}_i}$ in the above and thus $\widetilde{d} = D(\cdot \circ \mathbf{x}^{-1})$. The below proof assumes $\widetilde{d} = D$.]

Let M be a smooth one-dimensional manifold, let (U, x) be a smooth chart on M, and suppose we have a smooth function $f : U \to \mathbb{R}$. Consider the coordinate representation $\tilde{f}_{(U,x)} = f \circ x^{-1} : \mathbb{R} \to \mathbb{R}$, denoted as $\tilde{f} := \tilde{f}_{(U,x)}$ for brevity. Using d to denote $\widetilde{d} = D$, since $f = \tilde{f} \circ x$ we have $df_\mathbf{p} = d\tilde{f}_{x(\mathbf{p})}dx_\mathbf{p}$ for every $\mathbf{p} \in U$.

Then, since $T_\mathbf{p}^*(M)$ is one-dimensional for every $\mathbf{p} \in M$, we can divide by $dx_\mathbf{p}$:

$$\frac{df_\mathbf{p}}{dx_\mathbf{p}} = \frac{d\tilde{f}_{x(\mathbf{p})}dx_\mathbf{p}}{dx_\mathbf{p}} = d\tilde{f}_{x(\mathbf{p})} = D\tilde{f}_{x(\mathbf{p})} = \tilde{f}\,'(x(\mathbf{p})) \text{ for all } \mathbf{p} \in U.$$

---

[9]The ith coordinate vector field is smooth because its component functions are constants.

[10]The ith coordiante covector field is smooth because its component functions are constants.

[11]The local coordinate frame is smooth because the ith local coordinate vector field is smooth.

[12]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.

Thus $\frac{df}{dx}\big|_{\mathbf{p}} = (\tilde{f}\,' \circ x)(\mathbf{p})$ for every $\mathbf{p} \in U$. Using more suggestive notation for function composition, we have

$$\frac{df}{dx} = \tilde{f}\,'(x).$$

Lemma 9.60 (TO-DO. Supposedly it's possible to obtain the same result when M is a manifold of arbitrary dimension. The claim is that the chain rule $d(g \circ f)_{\mathbf{p}} = dg_{f(\mathbf{p})}df_{\mathbf{p}}$ holds when $f : M \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$). Let M be a smooth n-manifold and let $(U, \mathbf{x})$ be a smooth chart on M. Consider a smooth function $f : M \to \mathbb{R}$, and define $\tilde{f} := \widetilde{f}_{(U,\mathbf{x})}$ for brevity. For all $\mathbf{p} \in U$, we have

$$d(g \circ f)_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial(g \circ f)}{\partial x^i}\Big|_{\mathbf{p}} dx^i_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial(g \circ f \circ \mathbf{x}^{-1})}{\partial \hat{\mathfrak{e}}^i}\Big|_{\mathbf{p}} dx^i_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial(g \circ \tilde{f})}{\partial \hat{\mathfrak{e}}^i}\Big|_{\mathbf{p}} dx^i_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial g}{\partial \hat{\mathfrak{e}}^i}\Big|_{\widetilde{f}(\mathbf{p})} \frac{\partial \tilde{f}}{\partial \hat{\mathfrak{e}}_i}\Big|_{\mathbf{p}} dx^i_{\mathbf{p}}$$

$$= ..\text{this might not actually work}.. = \Big(\sum_{i=1}^{n} \frac{\partial g}{\partial \hat{\mathfrak{e}}^i}\Big|_{\widetilde{f}(\mathbf{p})} dx^i_{f(\mathbf{p})}\Big)\Big(\sum_{j=1}^{n} \frac{\partial \tilde{f}}{\partial \hat{\mathfrak{e}}^j}\Big|_{\mathbf{p}} dx^j_{\mathbf{p}}\Big) = dg_{\widetilde{f}(\mathbf{p})} d\tilde{f}_{\mathbf{p}}.$$

9.8 Oriented manifolds and their oriented boundaries

In Section 5.1, we defined what it means to orient a finite-dimensional vector space. The key result we motivated was Definition 5.47, 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 6.22 that choosing an element of $\Lambda^{\dim(V)}(V)$ determines an orientation on V. We now extend the idea of orientation to manifolds.

Definition 9.61. [Lee, p. 380] (Oriented manifolds).
  Let M be an n-manifold. We define the notion of orientation on M 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 9.62. [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 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 9.63. 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 9.64. [Lee, p. 118] (Inward- and outward- pointing tangent vectors).
  Let M be a smooth n-dimensional manifold and 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 9.65. [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 9.66. [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[13]) 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.

---

[13]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.

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 $\frac{(\mathbf{N}'|_{\mathbf{p}})_n}{(\mathbf{N}_{\mathbf{p}})_n}$. For $i > 1$, the ith column of the change of basis matrix is $\hat{\mathbf{\epsilon}}_i$. The change of basis matrix is therefore upper triangular, so its determinant is the product of the diagonal entries, i.e., the determinant is $\frac{(\mathbf{N}'|_{\mathbf{p}})_n}{(\mathbf{N}_{\mathbf{p}})_n} > 0$. **this doesn't make any sense; fix it**

$\square$

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 9.67. (Notation for n-parallelapipeds).
Let M be a smooth n-manifold 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 $\mathbf{x}(v_1|_{\mathbf{p}}), ..., \mathbf{x}(v_n|_{\mathbf{p}}) \in \mathbb{R}^n_{\mathbf{p}}$. **should $d\mathbf{x}(v_i|_{\mathbf{p}})$ be used here instead of $\mathbf{x}(v_1)|_{\mathbf{p}}$?** Additionally, if $v \in T_{\mathbf{p}}(M)$ is identified with $\mathbf{v} \in \mathbb{R}^n_{\mathbf{p}}$ under the isomorphism $\frac{\partial}{\partial x^i}\big|_{\mathbf{p}} \mapsto \hat{\mathbf{\epsilon}}_i$, 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 9.68. [HH, p. 542 - 544] (Oriented boundary of an n-parallelapiped).
Let M be an n-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

$$\partial P_{\mathbf{p}}(v_1, ..., v_n) = \sum_{i=1}^{k} (-1)^{i-1} \Big(P_{\mathbf{p}+v_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n) - P_{\mathbf{p}}(v_1, ..., \cancel{v_i}, ..., v_n) \Big)$$

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 9.66, 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 $\omega_P(-v_i|_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., \cancel{v_i}|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}}) = -\omega_{\partial P}$. $\square$

10

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.

10.1 Differential forms

Definition 10.1. [Lee, p. 360] (Differential k-form).
Let M be a smooth n-manifold. 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 $\Omega^k(M)$.

The *degree* of a differential k-form is k.

Theorem 10.2. [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 10.3. [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 10.4. 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 10.5. (Differential forms treated as functions).
Let M be a smooth manifold. 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 $\widetilde{\Omega}^k(M) := \{$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 $T_\mathbf{p}(M)$.

10.2 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 10.6. [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, and let $\mathbf{F} : U \subseteq M \to V \subseteq N$ be a smooth function. Consider a differential k-form $\omega \in \Omega^k(V)$, $k \le n$, on $V \subseteq N$.

In Definition **??**, we showed how when V and W are finite-dimensional vector spaces and $\mathbf{f} : V \to W$ is a linear function, it is possible to use the dual $\mathbf{f} : W^* \to V^*$ to "pull back" an element of $\Lambda^k(W^*)$ and obtain an element of $\Lambda^k(V^*)$.

$T_{\mathbf{p}}(M)$ and $T_{\mathbf{F(p)}}(N)$ are finite-dimensional vector spaces and $d\mathbf{F_p} : T_{\mathbf{p}}(M) \to T_{\mathbf{F(p)}}(N)$ is a linear function, so it is possible to "pull back" an element of $\Lambda^k(T_{\mathbf{F(p)}}(N))$ to obtain an element of $\Lambda^k(T_{\mathbf{p}}(M))$.

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

$$\left((\Omega^k \mathbf{F}^*)(\omega) \right)_{\mathbf{p}} := \begin{cases} \left(\Lambda^k (d\mathbf{F_p})^* \right)(\omega_{\mathbf{F(p)}}) & \omega \text{ is a differential } k\text{-form on } V, \; k \ge 1 \\ (f \circ \mathbf{F})(\mathbf{p}) & \omega = f \text{ is a differential 0-form on } V \text{ (i.e. } \omega \text{ is a function } V \to \mathbb{R}) \end{cases},$$

where we recall that for a linear function $\mathbf{f} : T_{\mathbf{p}}(M) \to T_{\mathbf{F(p)}}(N)$, the dual $\mathbf{f}^* : T^*_{\mathbf{F(p)}}(V) \to T^*_{\mathbf{p}}(U)$ is characterized by $\mathbf{f}^*(w_{\mathbf{F(p)}}) = w_{\mathbf{F(p)}} \circ \mathbf{f}$, and the pullback map $\Lambda^k \mathbf{f}^* : T^*_{\mathbf{F(p)}}(V) \to T^*_{\mathbf{p}}(U)$ induced by the dual is characterized by

$$\left(\Lambda^k \mathbf{f}^* \right)(dy^1|_{\mathbf{F(p)}} \wedge ... \wedge dy^n|_{\mathbf{F(p)}}) = \mathbf{f}^*(dy^1|_{\mathbf{F(p)}}) \wedge ... \wedge \mathbf{f}^*(dy^n|_{\mathbf{F(p)}}).$$

Theorem 10.7. (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 $\omega_{\mathbf{p}} \in T_{\mathbf{p}}(V)$ for every $\mathbf{p} \in N$, and we have $(\Omega^k \mathbf{F}^*)(\omega)_{\mathbf{p}} = (d\mathbf{F_p})^*(\omega)$, where $(d\mathbf{F_p})^*$ is the dual of the linear map $d\mathbf{F_p}$.

2. If $f : V \to \mathbb{R}$ and ω is a differential form on N, then $(\Omega^k \mathbf{F}^*)(f\omega) = (\Omega^k \mathbf{F}^*)(f) \, (\Omega^k \mathbf{F}^*)(\omega)$.

3. If ω and η 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 $\Omega^k \mathbf{F}^*$. $\qquad\square$

Theorem 10.8. (The differential commutes with the pullback of a diffeomorphism).

Let M and N be smooth manifolds, 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, d is the differential which results from identifying $T_{\mathbf{p}}(\mathbb{R}) \cong \mathbb{R}$ (recall Theorem 9.38).

Proof. Let $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. We will show $\Omega^k \mathbf{F}^*(df_{\mathbf{p}})(v_{\mathbf{p}}) = d(\Omega^k \mathbf{F}^*(f))_{\mathbf{p}}$.

Note: in the below proof, the "d" in "$df_{\mathbf{p}}$" is the the differential induced by the identification $T_{\mathbf{p}}(\mathbb{R}) \cong \mathbb{R}$ and is characterized by $df_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)$ (recall Theorem 9.38). On the other hand, the "d" in "$d\mathbf{F_p}$" involves no such implicit identifications and is characterized by $d\mathbf{F_p}(v_{\mathbf{p}})(f) = v_{\mathbf{p}}(f \circ \mathbf{F})$ (recall Theorem 9.35).

By property (1) of the previous theorem, $\Omega^k \mathbf{F}^*(df_{\mathbf{p}})(v_{\mathbf{p}}) = (d\mathbf{F_p})^*(df_{\mathbf{p}})(v_{\mathbf{p}})$. Then $(d\mathbf{F_p})^*(df_{\mathbf{p}})(v_{\mathbf{p}}) = (df_{\mathbf{p}} \circ d\mathbf{F_p})(v_{\mathbf{p}}) = df_{\mathbf{p}}(d\mathbf{F_p}(v_{\mathbf{p}}))$. Since $df_{\mathbf{p}}(u_{\mathbf{p}}) = u_{\mathbf{p}}(f)$, we use $u_{\mathbf{p}} = d\mathbf{F_p}(v_{\mathbf{p}})$ to obtain $df_{\mathbf{p}}(d\mathbf{F_p}(v_{\mathbf{p}})) = d\mathbf{F_p}(v_{\mathbf{p}})(f)$. By the definition of the differential $d\mathbf{F_p}$ (recall Definition 9.35), $d\mathbf{F_p}(v_{\mathbf{p}})(f) = v_{\mathbf{p}}(f \circ \mathbf{F})$. Since f is a 0-form, $f \circ \mathbf{F} = \Omega^k \mathbf{F}^*(f)$ by definition of $\Omega^k \mathbf{F}^*$. So $v_{\mathbf{p}}(f \circ \mathbf{F}) = v_{\mathbf{p}}(\Omega^k \mathbf{F}^*(f))$. Lastly, since $v_{\mathbf{p}}(g) = dg_{\mathbf{p}}(v_{\mathbf{p}})$, we use $g = \Omega^k \mathbf{F}^*(f)$ to obtain $v_{\mathbf{p}}(\Omega^k \mathbf{F}^*(f)) = d(\Omega^k \mathbf{F}^*(f))_{\mathbf{p}}(v_{\mathbf{p}})$. $\qquad\square$

Theorem 10.9. [Lee, p. 361] (Pullback of a top degree differential form).

Let M and N be smooth n-manifolds, 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. If $f\ dy^1 \wedge ... \wedge dy^n$ is a differential n-form on $V \subseteq N$ (we say it is a *top degree* differential form on N since its degree is the same as the dimension of N) then the pullback of this differential form is

$$\boxed{\Omega^k\mathbf{F}^*(f\ dy^1 \wedge ... \wedge dy^n) = \det(d\mathbf{F})\ (f \circ \mathbf{F})\ dx^1 \wedge ... \wedge dx^n}$$

Here $d\mathbf{F}$ denotes the map $\mathbf{p} \mapsto d\mathbf{F_p}$.

Proof. Take the pullback $\Omega^k\mathbf{F}^*$ of $f\ dy^1 \wedge ... \wedge dy^n$ at $\mathbf{p} \in U$ and apply the facts

- $(\Omega^k\mathbf{F}^*)(f\omega) = (\Omega^k\mathbf{F}^*)(f)\ (\Omega^k\mathbf{F}^*)(\omega)$

- $\Omega^k\mathbf{F}^*(f)_\mathbf{p} = (f \circ \mathbf{F})(\mathbf{p})$

- $\Omega^k\mathbf{F}^*(dy^i)_\mathbf{p} = \Lambda^1(d\mathbf{F_p})^*(dy^i_\mathbf{p}) = (d\mathbf{F_p})^*(dy^i_\mathbf{p})$

to see

$$\left(\Omega^k\mathbf{F}^*(f\ dy^1 \wedge ... \wedge dy^n)\right)_\mathbf{p} = (f \circ \mathbf{F})(\mathbf{p})\ (d\mathbf{F_p})^*(dy^1_\mathbf{p}) \wedge ... \wedge (d\mathbf{F_p})^*(dy^n_\mathbf{p}).$$

Since the pushforward on the top exterior power is multiplication by the determinant (Theorem 6.21), the above is equal to

$$\det(d\mathbf{F_p^*})\ (f \circ \mathbf{F})(\mathbf{p})\ dx^1_\mathbf{p} \wedge ... \wedge dx^n_\mathbf{p}.$$

Recall from Theorem that if $\mathbf{f} : V \to V$ is a linear function on a finite-dimensional vector space V, then $\det(\mathbf{f}^*) = \det(\mathbf{f})$. Thus $\det(d\mathbf{F_p^*}) = \det(d\mathbf{F_p})$, so the result is proven. $\square$

Remark 10.10. (Star notation for the pullback of a top degree differential form).

In Remark 6.24, we introduced the practice of using $\mathbf{F}^*$ to denote a pullback. From now on, we denote the pullback $\Omega^k\mathbf{F}^* : T_{\mathbf{F(p)}}(N) \to T_\mathbf{p}(M)$ by $\mathbf{F}^* : T_{\mathbf{F(p)}}(N) \to T_\mathbf{p}(M)$.

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 10.11. [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 10.12. [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, $\mathrm{supp}(\omega) := \mathrm{cl}(M - \omega^{-1}(\mathbf{0}))$. Iff $\mathrm{supp}(\omega) \subseteq A$, then we say ω *is supported in A*. We say ω is *compactly supported* iff $\mathrm{supp}(\omega)$ 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 10.13. [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 $\mathbf{F} : \mathrm{cl}(D) \to \mathrm{cl}(E)$ is a smooth map that restricts to a diffeomorphism from D to E, then, for every continuous function $f : \mathrm{cl}(E) \to \mathbb{R}$, we have

$$\int_E f = \int_D (f \circ \mathbf{F}) \Big| \det\Big(\frac{\partial F^i}{\partial \hat{\mathfrak{e}}^j}\Big)\Big|.$$

The matrix argument of the determinant is equal to

$$\Big(\frac{\partial \widetilde{F}^i_{(D,\widetilde{\mathbf{x}}),(E,\widetilde{\mathbf{y}})}}{\partial \hat{\mathfrak{e}}^j}\Big),$$

where $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ are the identity functions on D and E and $\widetilde{\mathbf{F}}^i_{(D,\widetilde{\mathbf{x}}),(E,\widetilde{\mathbf{y}})}$ is the coordinate representation of $\mathbf{F}$ relative to the smooth charts $(D,\widetilde{\mathbf{x}})$ and $(E,\widetilde{\mathbf{y}})$ **can we conclude these charts are smooth?**. This is the matrix of $d\mathbf{F}_{\mathbf{p}}$ relative to the bases $\Big\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big\}$ and $\Big\{\frac{\partial}{\partial y^i}\Big|_{\mathbf{F}(\mathbf{p})}\Big\}$ for $\mathbf{p} \in D$. Thus, using $d\mathbf{F}$ to denote $\mathbf{p} \mapsto d\mathbf{F}_{\mathbf{p}}$, the change of variables theorem can be stated as

$$\int_E f = \int_D (f \circ \mathbf{F}) |\det(d\mathbf{F})|.$$

Notice that, when $\det(d\mathbf{F}) > 0$, it is *almost* true that the pullback of the integrand on the left side is equal to the integrand on the right side, since $f \, dy^1 \wedge ... \wedge dy^n$ pulls back to $(f \circ \mathbf{F}) \, \det(d\mathbf{F}) \, dx^1 \wedge ... \wedge dx^n$. That is, the pullback of the differential form corresponding to the integrand of the left side is equal to the differential form corresponding to the integrand of the right 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 10.14. [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 10.15. (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 side in terms of the right side, you might think of it as giving algebraic meaning to the old placeholder notation of the right side.

Remark 10.16. 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 10.17. [Lee, p. 403] (Change of variables theorem in light of the pullback, part 2).
Consider the hypotheses of the previous derivation, Derivation 10.13. *In the case that $\mathbf{F}$ is orientation-preserving or orientation-reversing*, the change of variables theorem can be restated in terms of the previous definitions as

$$\int_E f\ dy^1 \wedge ... \wedge dy^n = \pm \int_D \mathbf{F}^*(f\ dy^1 \wedge ... \wedge dy^n),$$
where $\pm$ is $+$ iff $\mathbf{F}$ is orientation-preserving and $\pm$ is $-$ iff $\mathbf{F}$ is orientation-reversing.

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 10.18. [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 $\mathrm{cl}(E)$. Given the above definition, the restatement of the change of variables theorem presented at the end of the previous derivation is further restated as

$$\boxed{\begin{array}{l} \int_E \omega = \pm \int_D \mathbf{F}^*(\omega), \\[2mm] \text{where } \pm \text{ is } + \text{ iff } \mathbf{F} \text{ is orientation-preserving and } \pm \text{ is } - \text{ iff } \mathbf{F} \text{ is orientation-reversing} \end{array}}$$

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 10.19. [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 $\mathrm{supp}(\omega)$. This definition does not depend on the choice of D; see [Lee, p. 403] for the details. The right 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 10.20. [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

$$\boxed{\begin{aligned} \int_V \omega = \pm \int_U \mathbf{F}^*(\omega), \\ \text{where } \pm \text{ is } + \text{ iff } \mathbf{F} \text{ is orientation preserving and } - \text{ if } \mathbf{F} \text{ is orientation reversing} \end{aligned}}$$

Proof. The proof is a matter of using technical properties of diffeomorphisms. See [Lee, p. 404]. $\square$

As promised, we are now ready to define integration on smooth manifolds.

Definition 10.21. [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

$$\int_M \omega := \pm \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega),$$

where $\pm$ is $+$ iff $(U, \mathbf{x})$ is positively oriented and $-$ if $(U, \mathbf{x})$ is negatively oriented.

(Recall from Definition 9.62 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 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 10.22. (Pullback of a differential k-form, interpreted as a pointwise function, by a diffeomorphism).

The pullback of a "regular" differential form defined in Definition 10.6 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 10.6, we see the induced pullback $\widetilde{\Omega}^k \mathbf{F}^*$ is defined by:

$$\left((\widetilde{\Omega}^k \mathbf{F}^*)(\widetilde{\omega}) \right)_{\mathbf{p}} := \begin{cases} \left(\widetilde{\Lambda}^k (d\mathbf{F}_{\mathbf{p}})^* \right)(\widetilde{\omega}_{\mathbf{F}(\mathbf{p})}) & \widetilde{\omega} \text{ is a differential } k\text{-form on } V, \, k \geq 1 \\ (f \circ \mathbf{F})(\mathbf{p}) & f \text{ is a differential } 0\text{-form on } V \text{ (i.e. } f \text{ is a function } V \to \mathbb{R}) \end{cases}$$

Since $\mathbf{T} \in \Lambda^k(W^*) \overset{\widetilde{\Lambda}^k \mathbf{f}^*}{\longmapsto} \mathbf{T} \circ \mathbf{f} = \mathbf{f}^*(\mathbf{T}) \in \Lambda^k(V^*)$ (see the equation in Theorem 6.41 right above the boxed equation), the above becomes

$$\begin{cases} (d\mathbf{F}_{\mathbf{p}})^*(\widetilde{\omega}_{\mathbf{F}(\mathbf{p})}) = \widetilde{\omega}_{\mathbf{F}(\mathbf{p})} \circ d\mathbf{F}_{\mathbf{p}} & \widetilde{\omega} \text{ is a differential } k\text{-form on } V, \, k \geq 1 \\ (f \circ \mathbf{F})(\mathbf{p}) & f \text{ is a differential } 0\text{-form on } V \text{ (i.e. } f \text{ is a function } V \to \mathbb{R}) \end{cases}.$$

(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 6.41 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 10.23. [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 $U = \mathbf{y}(V)$, $\mathbf{x} = \mathbf{y}^{-1}$, and let $\widetilde{x}^i_{(V,\mathbf{y})}$ denote the ith coordinate function of $\mathbf{x}$ relative to the chart $(V, \mathbf{y})$. Then

$$\boxed{\begin{array}{l} \displaystyle\int_M \omega = \pm \int_M \widetilde{\omega}\Big(\frac{\partial \mathbf{x}}{\partial \widetilde{x}^1_{(V,\mathbf{y})}}, ..., \frac{\partial \mathbf{x}}{\partial \widetilde{x}^n_{(V,\mathbf{y})}}\Big), \\[4mm] \text{where } \pm \text{ is } + \text{ iff } (V, \mathbf{y}) \text{ is positively oriented and } - \text{ if } (V, \mathbf{y}) \text{ is negatively oriented.} \end{array}}$$

Proof. By Definition 10.21, 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 $\mathbf{x}^*$ to denote the pullback $\Omega^k \mathbf{x}^*$.

Since $\mathbf{x}^*(dy^i) = dy^i \circ \mathbf{x} = \mathbf{x}^i_{(V,\mathbf{y})}$, the above becomes

$$\pm \int_U \mathbf{x}^*(f\, dy^1 \wedge ... \wedge dy^n) = \pm \int_U (f \circ \mathbf{x})\, d\widetilde{x}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{x}^n_{(V,\mathbf{y})}.$$

Here, $\widetilde{x}^i_{(V,\mathbf{y})}$ denotes the ith coordinate function of $\mathbf{x}$ relative to the chart $(V, \mathbf{y})$. (Sidenote: $\mathbf{x}$ relative to $(V, \mathbf{y})$ is the identity on U: $\mathbf{x}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{x} = \mathbf{x}^{-1} \circ \mathbf{x} = \mathbf{I}_U$).

Definition 10.14 says that the integral of a differential form over a subset of $\mathbb{R}^n$ or $\mathbb{H}^n$ is computed by "erasing" the $d\widetilde{x}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{x}^n_{(V,\mathbf{y})}$. Thus, we have

$$\pm \int_U (f \circ \mathbf{x})\, d\widetilde{x}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{x}^n_{(V,\mathbf{y})} = \pm \int_U f \circ \mathbf{x}.$$

We now consider the differential form $\mathbf{x}^*(\widetilde{\omega}) = (f \circ \mathbf{x})\, d\widetilde{x}^1_{(V,\mathbf{y})} \widetilde{\wedge}...\widetilde{\wedge} d\widetilde{x}^n_{(V,\mathbf{y})} \in \widetilde{\Omega}^k(U)$ that corresponds to the "regular" differential form[1] $\mathbf{x}^*(\omega) = (f \circ \mathbf{x})\, d\widetilde{x}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{x}^n_{(V,\mathbf{y})} \in \Omega^k(U)$. Applying Lemma 6.36 at each point in U, we have

$$d\widetilde{x}^1_{(V,\mathbf{y})} \widetilde{\wedge}...\widetilde{\wedge} d\widetilde{x}^n_{(V,\mathbf{y})}\Big(\frac{\partial}{\partial \widetilde{x}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{x}^n_{(V,\mathbf{y})}}\Big) = 1.$$

Thus

$$\pm \int_U f \circ \mathbf{x} = \pm \int_U \underbrace{(f \circ \mathbf{x}) d\widetilde{x}^1_{(V,\mathbf{y})} \widetilde{\wedge}...\widetilde{\wedge} d\widetilde{x}^n_{(V,\mathbf{y})}}_{\mathbf{x}^*(\widetilde{\omega})}\Big(\frac{\partial}{\partial \widetilde{x}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{x}^n_{(V,\mathbf{y})}}\Big) = \pm \int_U \mathbf{x}^*(\widetilde{\omega})\Big(\frac{\partial}{\partial \widetilde{x}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{x}^n_{(V,\mathbf{y})}}\Big).$$

---

[1] "Regular" here means "is not treated as a function when evaluated at a point".

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

$$\pm \int_U \mathbf{x}^*(\widetilde{\omega}) \Big(\frac{\partial}{\partial \widetilde{x}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{x}^n_{(V,\mathbf{y})}} \Big) = \pm \int_U \widetilde{\omega} \Big(d\mathbf{x} \Big(\frac{\partial}{\partial \widetilde{x}^1_{(V,\mathbf{y})}} \Big), ..., d\mathbf{x} \Big(\frac{\partial}{\partial \widetilde{x}^1_{(V,\mathbf{y})}} \Big) \Big)$$

$$= \pm \int_U \widetilde{\omega} \Big(\frac{\partial \mathbf{x}}{\partial \widetilde{x}^1_{(V,\mathbf{y})}}, ..., \frac{\partial \mathbf{x}}{\partial \widetilde{x}^n_{(V,\mathbf{y})}} \Big) = \pm \int_M \widetilde{\omega} \Big(\frac{\partial \mathbf{x}}{\partial \widetilde{x}^1_{(V,\mathbf{y})}}, ..., \frac{\partial \mathbf{x}}{\partial \widetilde{x}^n_{(V,\mathbf{y})}} \Big).$$

$\square$

Theorem 10.24. (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 $\mathrm{supp}(\omega) \subseteq \bigcup_{i=1}^k U_i$, then

Since $\mathrm{supp}(\omega) \subseteq \bigcup_{i=1}^k U_i$, then $\mathrm{supp}((\mathbf{x}^{-1})^*) \subseteq \bigcup_{i=1}^k \mathbf{x}(U_i)$. 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 7.16), 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. $\square$

The previous theorem is generalized by the following definition, which finalizes the definition of integration of differential forms.

Definition 10.25. [Lee, p. 408] (Integral of a top degree differential form on a smooth manifold).

Let M be an oriented smooth n-manifold and let ω be a compactly supported n-form on M. Since $\mathrm{supp}(\omega)$ 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 $\mathrm{supp}(\omega)$, 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 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.

Theorem 10.26. (Integral of a differential form on a smooth manifold is linear).

Let M be an oriented smooth n-manifold, 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

$$\int_M c_1 \omega + c_2 \eta = c_1 \int_M \omega + c_2 \int_M \eta.$$

---

[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.

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 10.21- 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 $\mathbb{R}^n \to \mathbb{R}$. $\square$

10.3 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 10.27. [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:

$$\mathrm{div}(\mathbf{V}) := \lim_{|V| \to 0} \frac{1}{|V|} \int_{\partial V} \mathbf{V} \cdot \hat{\mathbf{n}} dS.$$

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 curl$(\mathbf{V})$ is defined as work per "infinitesimal" surface area:

$$\mathrm{curl}(\mathbf{V}) \cdot \hat{\mathbf{n}} := \lim_{A(C) \to 0} \frac{1}{A(C)} \int_C \mathbf{V} \cdot d\mathbf{r} = \lim_{A(C) \to 0} \frac{1}{A(C)} \int_C \mathbf{V} \cdot \frac{d\mathbf{r}}{dt} dt.$$

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 $T_{\mathbf{p}}(M)$ spanned by $\mathbf{v}_1|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. 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 10.28. [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 9.38).

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 7.9, 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 $\{(V_{\mathrm{f}i}^h, \mathbf{y}_{\mathrm{f}i})\}_{i=1}^{k+1}$ that parameterize the $k+1$ faces on the "front" of this $(k+1)$-parallelapiped and charts $\{(V_{\mathrm{b}i}^h, \mathbf{y}_{\mathrm{b}i})\}_{i=1}^{k+1}$ that parameterize the remaining $k+1$ faces on its "back".

   Specifically, let $\mathbf{v}_i|_{\mathbf{p}} \in \mathbb{R}_{\mathbf{p}}^n$ be the image of $v_i|_{\mathbf{p}}$ under the isomorphism $T_{\mathbf{p}}(P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})) \cong \mathbb{R}_{\mathbf{p}}^n$ that sends $\frac{\partial}{\partial y^i} \mapsto \hat{\boldsymbol{\varepsilon}}_i$. We define, for $\mathbf{q} \in [0, h]^k \subseteq \mathbb{R}^k$,

$$\mathbf{y}_{\mathrm{f}i}^{-1}(\mathbf{q}) := q^1 \mathbf{v}_1|_{\mathbf{p}} + ... + q^{i-1}\mathbf{v}_{i-1}|_{\mathbf{p}} + q^i \cancel{\mathbf{v}_i|_{\mathbf{p}}} + q^{i+1}\mathbf{v}_{i+1}|_{\mathbf{p}} + ... + q^{k+1}\mathbf{v}_{k+1}|_{\mathbf{p}}$$

$$\mathbf{y}_{\mathrm{b}i}^{-1}(\mathbf{q}) := q^1 \mathbf{v}_1|_{\mathbf{p}} + ... + q^{i-1}\mathbf{v}_{i-1}|_{\mathbf{p}} + h\mathbf{v}_i|_{\mathbf{p}} + q^{i+1}\mathbf{v}_{i+1}|_{\mathbf{p}} + ... + q^{k+1}\mathbf{v}_{k+1}|_{\mathbf{p}}.$$

Following the conventions of Theorem 10.23, we define $U_{\mathrm{f}i}^h = \mathbf{y}(V_{\mathrm{f}i}^h)$, $\mathbf{x}_{\mathrm{f}i} = \mathbf{y}^{-1}$ and $U_{\mathrm{b}i}^h = \mathbf{y}(V_{\mathrm{b}i}^h)$, $\mathbf{x}_{\mathrm{b}i} = \mathbf{y}^{-1}$ so that $\{(U_{\mathrm{f}i}^h, \mathbf{x}_{\mathrm{f}i})\}_{i=1}^{k+1}$ and $\{(U_{\mathrm{b}i}^h, \mathbf{x}_{\mathrm{b}i})\}_{i=1}^{k+1}$ are charts in $\mathbb{R}^n$ or $\mathbb{H}^n$. Note that $\mathbf{x}_{\mathrm{b}i} = \mathbf{y}_{\mathrm{b}i}^{-1} = \mathbf{y}_{\mathrm{f}i}^{-1} + h\mathbf{v}_i = \mathbf{x}_{\mathrm{f}i} + h\mathbf{v}^i$. Additionally, since we required $\mathbf{q} \in [0, h]^k \subseteq \mathbb{R}^k$, then $U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h$ 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 10.23 to treat the integral as the integral of a differential form evaluated on tangent vectors. Also using Theorem 10.24 to break up the domain of integration and Theorem 9.68 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

$$\int_{\partial P_{\mathbf{p}}(\mathbf{v}_1, ..., \mathbf{v}_{k+1})} f\ d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k} = \underbrace{-\sum_{i=1}^{k+1}(-1)^{i-1} \int_{U_{\mathrm{f}i}^h} (f \circ \mathbf{x}_{\mathrm{f}i})\ d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k} \left(\frac{\partial \mathbf{x}_{\mathrm{f}i}}{\partial q^1}, ..., \cancel{\frac{\partial \mathbf{x}_{\mathrm{f}i}}{\partial q^i}}, ..., \frac{\partial \mathbf{x}_{\mathrm{f}i}}{\partial q^{k+1}} \right)}_{\text{sum of integrals over front faces}}$$

$$+ \underbrace{\sum_{i=1}^{k+1}(-1)^{i-1} \int_{U_{\mathrm{b}i}^h} (f \circ \mathbf{x}_{\mathrm{b}i})\ d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k} \left(\frac{\partial \mathbf{x}_{\mathrm{b}i}}{\partial q^1}, ..., \cancel{\frac{\partial \mathbf{x}_{\mathrm{b}i}}{\partial q^i}}, ..., \frac{\partial \mathbf{x}_{\mathrm{b}i}}{\partial q^{k+1}} \right)}_{\text{sum of integrals over back faces}}.$$

For $j \neq i$, we have $\frac{\partial \mathbf{x}_{\mathrm{f}i}}{\partial q^j} = \frac{\partial \mathbf{x}_{\mathrm{b}i}}{\partial q^j} = \mathbf{v}_j|_{\mathbf{p}}$. Since the argument of $d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k}$ in each integrand is the same, the above is

$$\sum_{i=1}^{k+1}(-1)^{i-1} \int_{U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h} (f \circ \mathbf{x}_{\mathrm{b}i} - f \circ \mathbf{x}_{\mathrm{f}i})\ d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k} (\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).$$

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 $\mathbf{x}_{\mathrm{f}i} : U_{\mathrm{f}i}^h \subseteq \mathbb{R}^n \to U$ and $\mathbf{x}_{\mathrm{b}i} : U_{\mathrm{b}i}^h \subseteq \mathbb{R}^n \to U$!). We now write $f : U \to \mathbb{R}$ as a Taylor polynomial centered at $\mathbf{p}_0 \in U$ with remainder:

$$T_f(\mathbf{p}) = T_f^0(\mathbf{p}) + T_f^1(\mathbf{p}) + R_f(\mathbf{p}) = f(\mathbf{p}_0) + df_{\mathbf{p}_0}(\mathbf{p}) + R_f(\mathbf{p}),$$

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 $f \circ \mathbf{x}_{\mathrm{f}i} - f \circ \mathbf{x}_{\mathrm{b}i}$ from the above integrand, we see

$$T_f \circ \mathbf{x}_{\mathrm{b}i} - T_f \circ \mathbf{x}_{\mathrm{f}i} = (T_f^0 - T_f^1 + R_f) \circ \mathbf{x}_{\mathrm{b}i} - (T_f^0 - T_f^1 + R_f) \circ \mathbf{x}_{\mathrm{f}i}$$

$$= (T_f^0 \circ \mathbf{x}_{\mathrm{b}i} - T_f^0 \circ \mathbf{x}_{\mathrm{f}i}) + (T_f^1 \circ \mathbf{x}_{\mathrm{b}i} - T_f^1 \circ \mathbf{x}_{\mathrm{f}i}) + (R_f \circ \mathbf{x}_{\mathrm{b}i} - R_f \circ \mathbf{x}_{\mathrm{f}i}).$$

Since T_f^0 is a constant function, the first term is zero: $T_f^0 \circ \mathbf{x}_{\mathrm{f}i} - T_f^0 \circ \mathbf{x}_{\mathrm{b}i} = 0$. 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

$$\int_{U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h} (T_f^1 \circ \mathbf{x}_{\mathrm{b}i} - T_f^1 \circ \mathbf{x}_{\mathrm{f}i}) d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k} (\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}})$$

$$+ \int_{U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h} (R_f \circ \mathbf{x}_{\mathrm{b}i} - R_f \circ \mathbf{x}_{\mathrm{f}i}) d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k} (\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).$$

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 $d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k}(\mathbf{v}_1|_\mathbf{p}, ..., \cancel{\mathbf{v}_i|_\mathbf{p}}, ..., \mathbf{v}_{k+1}|_\mathbf{p})$ are constant with respect to the limit, which takes $h \to 0$, then $|d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k}(\mathbf{v}_1|_\mathbf{p}, ..., \cancel{\mathbf{v}_i|_\mathbf{p}}, ..., \mathbf{v}_{k+1}|_\mathbf{p})|$ 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

$$\lim_{h \to 0} \frac{1}{h^{k+1}} \int_{\partial P_\mathbf{p}(hv_1|_\mathbf{p}, ..., hv_{k+1}|_\mathbf{p})} f \, dy^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dy^{i_k}$$
$$= \lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} \int_{U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h} (T_f^1 \circ \mathbf{x}_{\mathrm{b}i} - T_f^1 \circ \mathbf{x}_{\mathrm{f}i}) d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k}(\mathbf{v}_1|_\mathbf{p}, ..., \cancel{\mathbf{v}_i|_\mathbf{p}}, ..., \mathbf{v}_{k+1}|_\mathbf{p})$$

Observe that the term $T_f^1 \circ \mathbf{x}_{\mathrm{b}i} - T_f^1 \circ \mathbf{x}_{\mathrm{f}i}$ in the integrand is $T_f^1 \circ \mathbf{x}_{\mathrm{b}i} - T_f^1 \circ \mathbf{x}_{\mathrm{f}i} = hdf_{\mathbf{p}_0}(\mathbf{v}_i)$:

$$(T_f^1 \circ \mathbf{x}_{\mathrm{b}i} - T_f^1 \circ \mathbf{x}_{\mathrm{f}i})(\mathbf{p}) = df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{b}i}(\mathbf{p})) - df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{f}i}(\mathbf{p})) = df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{b}i}(\mathbf{p})) - df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{f}i}(\mathbf{p}))$$
$$= df_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{f}i}(\mathbf{p}) + h\mathbf{v}_i) - d_{\mathbf{p}_0}(\mathbf{x}_{\mathrm{f}i}(\mathbf{p})) = hdf_{\mathbf{p}_0}(\mathbf{v}_i).$$

Thus, the limit is

$$\lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} \int_{U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h} hdf_{\mathbf{p}_0}(\mathbf{v}_i) \, d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k}(\mathbf{v}_1|_\mathbf{p}, ..., \cancel{\mathbf{v}_i|_\mathbf{p}}, ..., \mathbf{v}_{k+1}|_\mathbf{p}).$$

(Here, $df_{\mathbf{p}_0}(\mathbf{v}_i) \in U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h$ denotes the constant map $\mathbf{q} \in \mapsto df_{\mathbf{p}_0}(\mathbf{v}_i)$). Since the integral is taken over points in $U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h$, then $hdf_{\mathbf{p}_0}(\mathbf{v}_i)$ is constant with respect to the integral, and the above is

$$\lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} hdf_{\mathbf{p}_0}(\mathbf{v}_i) \int_{U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h} d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k}(\mathbf{v}_1|_\mathbf{p}, ..., \cancel{\mathbf{v}_i|_\mathbf{p}}, ..., \mathbf{v}_{k+1}|_\mathbf{p}).$$

$U_{\mathrm{f}i}^h \sqcup U_{\mathrm{b}i}^h$ 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:

$$\lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} hdf_{\mathbf{p}_0}(\mathbf{v}_i) h^k \, d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k}(\mathbf{v}_1|_\mathbf{p}, ..., \cancel{\mathbf{v}_i|_\mathbf{p}}, ..., \mathbf{v}_{k+1}|_\mathbf{p})$$
$$= \sum_{i=1}^{k+1} (-1)^{i-1} df_{\mathbf{p}_0}(\mathbf{v}_i) \, d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k}(\mathbf{v}_1|_\mathbf{p}, ..., \cancel{\mathbf{v}_i|_\mathbf{p}}, ..., \mathbf{v}_{k+1}|_\mathbf{p})$$
$$= \mathrm{alt}(df \widetilde{\otimes} (d\widetilde{x}^{i_1} \widetilde{\otimes} ... \widetilde{\otimes} d\widetilde{x}^{i_k}))(\mathbf{v}_1|_\mathbf{p}, ..., \mathbf{v}_{k+1}|_\mathbf{p})$$
$$= df \widetilde{\wedge} d\widetilde{x}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{x}^{i_k}(\mathbf{v}_1|_\mathbf{p}, ..., \mathbf{v}_{k+1}|_\mathbf{p}).$$

$\square$

Theorem 10.29. [Lee, p. 366] (The exterior derivative commutes with the pullback of a smooth function $M \to N$).

Let M and N be smooth manifolds, 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 10.8. $\square$

196

Theorem 10.30. [Lee, p. 364] $(d \circ d = \mathbf{0})$.

Let M be a smooth manifold 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 side and use the fact that the exterior derivative commutes with the pullback of a diffeomorphism on the right 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 $d\omega = x d\omega$. Somehow, this implies that $d\omega = 0$. Thus, we have shown that all translation invariant differential forms ω satisfy $d\omega = 0$. $\qquad\square$

Theorem 10.31. [HH, p. 652 - 655] (Linearity of the exterior derivative).

Let M be a smooth manifold 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 10.28, 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 10.28, 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 10.32. [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 10.33. (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 $\sigma \in S_n$,

$$*(\mathbf{e}_{\sigma(1)} \wedge ... \wedge \mathbf{e}_{\sigma(k)}) = \text{sgn}(\sigma)(\mathbf{e}_{\sigma(k+1)} \wedge ... \wedge \mathbf{e}_{\sigma(n)}).$$

Theorem 10.34. (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 9.43 and 9.44).

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:
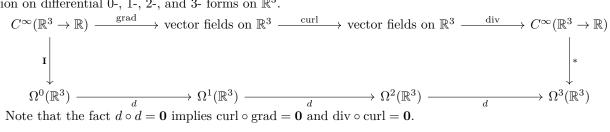
$$\boxed{\begin{aligned} \nabla f &= \mathbf{F}((df)^\sharp) \\ \text{curl}(\mathbf{V}) &= \mathbf{F}((*d(\mathbf{V}^\flat))^\sharp) \\ \text{div}(\mathbf{V}) &= *d * (\mathbf{V}^\flat) \end{aligned}}$$

Proof.

- From Theorem 9.39, 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 $\mathbf{F}((df)^{\sharp}) = \sum_{i=1}^{3} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \hat{\mathbf{e}}_i = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}}$.

- 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 $\hat{\mathbf{e}}_i$ 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 10.35. [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 $\mathbb{R}^3$.

$$
\begin{array}{ccccccc}
C^{\infty}(\mathbb{R}^3 \to \mathbb{R}) & \xrightarrow{\text{grad}} & \text{vector fields on } \mathbb{R}^3 & \xrightarrow{\text{curl}} & \text{vector fields on } \mathbb{R}^3 & \xrightarrow{\text{div}} & C^{\infty}(\mathbb{R}^3 \to \mathbb{R}) \\
\downarrow{\scriptstyle \mathbf{I}} & & & & & & \downarrow{\scriptstyle *} \\
\Omega^0(\mathbb{R}^3) & \xrightarrow{\quad d \quad} & \Omega^1(\mathbb{R}^3) & \xrightarrow{\quad d \quad} & \Omega^2(\mathbb{R}^3) & \xrightarrow{\quad d \quad} & \Omega^3(\mathbb{R}^3)
\end{array}
$$

Note that the fact $d \circ d = \mathbf{0}$ implies curl $\circ$ grad $= \mathbf{0}$ and div $\circ$ curl $= \mathbf{0}$.

10.4 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 10.36. [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

- $\mathrm{supp}(f_\alpha) \subseteq U_\alpha$ for all α

- $f_\alpha(\mathbf{p})$ is nonzero for only finitely many α, for each $\mathbf{p} \in M$

- $(\sum_\alpha f_\alpha(\mathbf{p}) = 1$ for all $\mathbf{p} \in M) \iff (\sum_\alpha f$ is the multiplicative identity of $\{$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 admits a smooth partition of unity ([Lee, p. 43]). Since every manifold has an open cover, this means that every smooth manifold admits a smooth partition of unity.

Theorem 10.37. [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 has a smooth partition of unity.

Theorem 10.38. (Integral of a differential n-form on a smooth n-manifold via partition of unity).
Let M be an oriented smooth n-manifold and let ω be a compactly supported n-form on M. Since $\mathrm{supp}(\omega)$ 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 $\mathrm{supp}(\omega)$, 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 10.21).

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 10.39. (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 10.25), as this definition is an intuitive starting point.

Theorem 10.40. [Lee, p. 407] (Properties of integrals of differential forms).

Theorem 10.41. [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 $\sum_{C \in D_N(\mathrm{cl}(\mathbb{H}^k))} \int_{\partial C} \omega = \sum_{C' \in D_N(\partial \mathrm{cl}(\mathbb{H}^k))} \int_{C'} \omega$, which is equal to $\int_{\partial \mathrm{cl}(\mathbb{H}^k)} \omega$ by Definition 10.25.

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 $|\int_U d\omega - \sum_{C_i \in D_N(\mathrm{cl}(\mathbb{H}^k))} d\omega(P_i)| < \epsilon$.

$(\sum_i d\omega(P_i) \approx \sum_i \int_{\partial P_i})$. 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 $C_i = P_{\mathbf{p}}(h\hat{\mathbf{e}}_1, ..., h\hat{\mathbf{e}}_k)$ for some $\mathbf{p}$, and where $h = 2^{-N}$. By the definition of the exterior derivative (Definition 10.27), 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 $\sum_{C \in D_N(\mathrm{cl}(\mathbb{H}^k))} \int_{\partial C} \omega$. Applying Theorem 9.68, which gives the oriented boundary of a k-parallelapiped, 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. Thus $\sum_{C \in D_N(\mathrm{cl}(\mathbb{H}^k))} \int_{\partial C} \omega = \sum_{C' \in D_N(\partial \mathrm{cl}(\mathbb{H}^k))} \int_{C'} \omega = \int_{\partial M} \omega$. $\square$

Remark 10.42. 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 10.43. [Lee, p. 419] (The generalized Stokes' theorem).

Let M be an oriented smooth manifold with corners, and let ω be a compactly supported smooth differential $(n-1)$-form on M. Then

$$\int_M d\omega = \int_{\partial M} \omega.$$

Proof. 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

$$\int_M d\omega = \int_M \sum_{i=1}^k f_i d\omega = \int_M \sum_{i=1}^k d(\underbrace{f_i \omega}_{\omega_i}) = \int_M \sum_{i=1}^k d\omega_i.$$

---

[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].

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 10.21 and 10.25, 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 10.29), we have

$$\sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(d\omega_i) = \sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} d\left((\mathbf{x}_i^{-1})^*(\omega_i)\right).$$

We apply the previous theorem- Stokes' theorem on a single smooth chart- to the integral inside the sum to obtain

$$\sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} d\left((\mathbf{x}_i^{-1})^*(\omega_i)\right) = \sum_{i=1}^k \int_{\partial \mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(\omega_i).$$

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 10.38. $\qquad\qquad\square$

Remark 10.44. ("∂ 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 **??**, 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 $([\omega]_{\{f_i\}_{i=1}^k})_i$.

Part IV

Appendices

11

Appendix

Function-input duality

Derivation 11.1. (Function-input duality).

Let X and Y be sets, let $x \in X$, let $f : X \to Y$ be a function, and consider the expression $f(x)$.

Typically, we think of f as being applied to x to produce the output $f(x)$. We can also think of x as being applied to f, in some sense, to produce $f(x)$. In intuitive notation, we have "$x(f)$" $= f(x)$.

To formalize the intuition, notice that for every $x \in X$ there is a function F_x that itself takes a function $X \to Y$ as input and applies it to x; so, $F_x(f) = f(x)$. Since the function sending $x \mapsto F_x$ is a bijection, we can identify x with F_x and then think of x as being applied to f via the action of F_x on f. The intuition thus formalizes as "$x(f)$" $= F_x(f) = f(x)$.

We've successfully formalized the idea of applying inputs to functions, but there is still more insight to be had. If we define $X^* := \{\text{functions } X \to Y\}$, then we have $f \in X^*$ and $F_x \in (X^*)^* = X^{**}$. We've seen that $x \mapsto F_x$ is a bijection $X \to X^{**}$. Thus, X and X^{**} are "the same", in some sense, and we write $X \cong X^{**}$. On the other hand, there is no natural bijection between X and X^*.

Overall, the relationship between X, X^*, and X^{**} is that elements of $X \cong X^{**}$ and X^* act on each other. Since elements of X^* cannot be naturally identified with elements X, yet can still act on elements of X, we call X^* the *dual set (to X)*. (We call X^{**} the *double dual set (to X)*).

Symmetric and orthogonal linear functions

The only results from this appendix that are necessary to know are the conditions (3) and (4) of Definition 11.5 satisfied by an orthogonal linear function.

Derivation 11.2. (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:

$$
\begin{array}{ccc}
V & \xrightarrow{\;\mathbf{f}^\dagger\;} & W \\
{\scriptstyle \flat_1}\downarrow & & \downarrow{\scriptstyle \flat_2} \\
W^* & \xrightarrow[\;\mathbf{f}^*\;]{} & V^*
\end{array}
$$

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 3.33), we have $\mathbf{f}^\dagger(\mathbf{v}_0)^{\flat_2} = \mathbf{v}_0^{\flat_1} \circ \mathbf{f}$. This is the same[1] 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:

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

---

[1] Here we've used $B(\cdot, \mathbf{v}_2)$ to denote the function $\mathbf{v}_1 \mapsto B(\mathbf{v}_1, \mathbf{v}_2)$.

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 11.3. (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 $\mathbf{f}^\dagger = \mathbf{f}^{-1}$.

Definition 11.4. (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 $a_{ij} = a_i^j$. 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. $\square$

Definition 11.5. (Orthogonal linear function).
Let V be a vector space over K (where[2] 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: $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_2) \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.

4. $\mathbf{f}$ preserves length: $\langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{f}(\mathbf{v}), \mathbf{f}(\mathbf{v}) \rangle$ for all $\mathbf{v} \in V$.

5. $\mathbf{f}$ preserves length and angle: $\langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{f}(\mathbf{v}), \mathbf{f}(\mathbf{v}) \rangle$ for all $\mathbf{v} \in V$ and $\theta(\mathbf{v}_1, \mathbf{v}_2) = \theta(\mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_2))$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.

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 $(1) \iff (2) \iff (3) \iff (6) \iff (7)$. The proof of $(3) \iff (4) \iff (5)$ is essentially the same as the proof of Lemma 2.136; just replace the dot product of that lemma with an inner product. So, it remains to show $(1) \iff (2) \iff (3) \iff (6) \iff (7)$:
$(1 \iff 2)$.

---

[2]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 $\iff$ 3). Setting $\mathbf{v}_3 := \mathbf{f}^{-1}(\mathbf{v}_2)$ shows that $(\langle \mathbf{f}(\mathbf{v}_1), \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{f}^{-1}(\mathbf{v}_2) \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V)$ $\iff$ $(\langle \mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_3) \rangle = \langle \mathbf{v}_1, \mathbf{v}_3 \rangle$ 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 $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$)·($j$th column of $[\mathbf{f}(\widehat{U})]_{\widehat{U}})\top = $ (ith row of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$)· ($j$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). $\qquad\square$

Bubble sort

Definition 11.6. (Bubble sort).

Given a permutation $\sigma \in S_n$, the *bubble sort algorithm* operates on the tuple $(\sigma(1), ..., \sigma(n))$, and produces the tuple $(1, ..., n)$ as output.

The following is a description of bubble sort that does not include the popular optimizations:

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

    - For each $j \in \{1, ..., n-1\}$, if $\sigma(j) \geq \sigma(j+1)$, then swap (i.e. apply a transposition to) $\sigma(j)$ and $\sigma(j+1)$.

This algorithm indeed sorts $(\sigma(1), ..., \sigma(n))$ because the ith greatest element of the tuple is in its final position after the ith pass of the outer loop. (The first-greatest element of the tuple, or maximum, is in its final position after the first pass of the loop, the second-greatest element of the tuple is in its final position after the second pass of the loop, and so on).

Bibliography

[BW97]   Javier Bonet and Richard D. Wood. <u>Nonlinear Continuum Mechanics for Finite Element Analysis</u>. 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. <u>Differential Topology</u>. Prentice-Hall, 1 edition, 1974.

[HH]   Barbara Burke Hubbard and John Hamal Hubbard. <u>Vector Calculus, Linear Algebra, And Differential Forms</u>. Prentice Hall, 1 edition.

[Lee]   John M. Lee. <u>Introduction to Smooth Manifolds</u>. Springer, 2 edition.

[War]   Frank W. Warner. <u>Foundations of Differentiable Manifolds and Lie Groups</u>. 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].