Tensors, Differential Forms, and Computer Graphics

Ross Grogan-Kaylor

September 2020

Acknowledgements

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

To do

Checklist for turn in

Defns

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

Lin alg

- replace "matrix relative to E and E" with "matrix relative to E"
- check that there's enough stars in lin alg. add stars to other review sections?
- fix lin alg injective iff surjective. only works when dimensions the same. look at relevant thms about linear isos, main dim thm, and fix
- remove finite-dimensionality from motivated intro and other sections when unnecessary
- make sure "bilinear" said instead of "multilinear" when applicable
- new style: write overall hypotheses for each section. add redundant hypotheses for each theorem in small text. don't worry about this till ends
- use "has as a basis" instead of "and let E be a basis for V"?
- say "over the same field" when no field is explicitly mentioned, when necessary
- add matrix addition, etc.
- Do I emphasize the equivalence between primitive matrix and the characterizing property of $[\mathbf{f}(E)]_F$ enough? Maybe a remark on these two ways to interpret things would be good.
- Add direct sums and $\dim(W_1 + W_2)$ theorem to lin alg. Look in Halmos for good proof on direct sum condition.

Proofs

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

$\binom{p}{q}$ tensors

- make sure things of the form $\cup_{p,q \in \mathbb{N}} T^p_q(V)$ are used instead of $T^p_q(V)$
- decide on when to use ϵ^i vs. $\phi^{\mathbf{e}_i}$
- When defining maps on elementary tensors, is the elementary tensor of the form $\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_p \otimes \phi^1 \otimes ... \otimes \phi^q$?

Misc.

- add more subsections, and explanatory text in-between theorems in these subsections
- add text in-between proofs regardless!
- make sure sum parenthetical conventions are consistent
- Are all instances of lists such as $x_1 ... x_r ... x_n$ replaced with $x_1 ... \cancel{x_r} ... x_n$?
- make sure $\{1, ..., n\} - i$ is replaced with $\{1, ..., n\} - \{i\}$ in sum subscripts
- standardize the way I type "pull back", "pullback", "pull-back"
- emphasize "for all" as encompassing entire proof, like brackets in programming.
- replace $\widetilde{\hat{\mathbf{u}}}_i$ with $\widetilde{\hat{\mathbf{u}}_i}$

Calc

- add technical hypotheses to calc chapter
- proof FTC via Hubbard

Notation

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

Outline

Equivalent definitions of "tensor"

- Original def: tensor product spaces.
- Multilinear map def. (Example: bilinear forms)
  - Pull-push forward and pull-back of tensors in this interpretation.
    Push-forward...
    Pull-back...
    Originally was $(\otimes_q \mathbf{f}^*)(\phi^1 \otimes ... \otimes \phi^q) = \mathbf{f}^*(\phi^1) \otimes ... \otimes \mathbf{f}^*(\phi^q)$, extended linearly. Now have $(\overline{\otimes}_q \mathbf{f}^*)(\phi^1 \otimes ... \otimes \phi^q) = (\mathbf{f}^*(\phi^1) \overline{\otimes} ... \overline{\otimes} \mathbf{f}^*(\phi^q)) = ((\phi^1 \circ \mathbf{f}) \overline{\otimes} ... \overline{\otimes} (\phi^q \circ \mathbf{f})) = ((\phi^1 \circ \mathbf{f}) \overline{\otimes} ... \overline{\otimes} (\phi^q \circ \mathbf{f})) = (\phi^1 \circ \overline{\otimes} ... \overline{\otimes} \phi^q) \circ \mathbf{f}$. Check this last equality by evaluating both sides on $(\mathbf{v}_1, ..., \mathbf{v}_q)$. Extending linearly, we see $\overline{\otimes}_q \mathbf{T} = \mathbf{T} \circ \mathbf{f} = \mathbf{f}^*(\mathbf{T})$. That is, $(\overline{\otimes}_q \mathbf{T})(\mathbf{v}_1, ..., \mathbf{v}_q) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_q))$.
- Highly optional
  - Bonet and Wood linear map recursive def.
  - Def as multidimensional matrices satisfying change of basis condition
- For each definition, present the manifestation of $\otimes$.
- Add $\overline{\otimes}$, $\overline{\wedge}$ to notation page.

Differential forms

- Don't have to use $\overline{\wedge}$ (interpretation of diff form at a point as actual alteranting multilinear function), can use $\wedge$ (interpretation of diff form at a point as alternating tensor)!
  - Proof that $\det(d\mathbf{f}_{\mathbf{x}})$ is involved in the change of variables theorem.
    By Lemma [...] we have $\det((d\mathbf{f}_{\mathbf{x}})^*) = \det(\delta^i(\mathbf{e}_j))$. Then $\delta^i(\mathbf{e}_j) = (\epsilon^i \circ d\mathbf{f}_{\mathbf{x}})(\mathbf{e}_j) = \epsilon^i(d\mathbf{f}_{\mathbf{x}}(\mathbf{e}_j))$. But $d\mathbf{f}_{\mathbf{x}}(\mathbf{e}_j)$ is the directional derivative of $\mathbf{f}$ in the $\mathbf{e}_j$ direction, $d\mathbf{f}_{\mathbf{x}}(\mathbf{e}_j) = \frac{\partial \mathbf{f}(x_1, ..., x_j, ..., x_k)}{\partial x_j}$. We also have $\epsilon^i(\mathbf{v}) = ([\mathbf{v}]_E)^i$, so $\delta^i(\mathbf{e}_j) = ([\frac{\partial \mathbf{f}(x_1, ..., x_j, ..., x_k)}{\partial x_j}]_E)^i$. If $\mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_k(\mathbf{x}) \end{pmatrix}$, then by definition of partial derivative of a vector-valued function, we have $([\frac{\partial \mathbf{f}(x_1, ..., x_j, ..., x_k)}{\partial x_j}]_E)^i = \frac{\partial f_i(\mathbf{x})}{\partial x_j}$. Therefore $\det((d\mathbf{f})^*) = \det(\frac{\partial f_i(\mathbf{x})}{\partial x_j})$.
  - $\Omega^k \mathbf{f}^*$ and $\overline{\Omega^k \mathbf{f}^*}$ are nonstandard notation; people denote $\overline{\Omega^k} \mathbf{f}^*$ by $\mathbf{f}^*$, even though this can be confused with the dual transformation.

Manifolds
Tensors in physics and engineering. (Put in Appendix?)

- Outer product
- If we have identified $V \cong V^*$, then every linear transformation $V \to W$ has a matrix of the form $\sum_k \mathbf{w}_k \mathbf{v}_k^\top = \mathbf{W}^\top \mathbf{V}$.

– By the discussion on rank 1 tensors above, every element of $\text{Hom}(V, W)$ has a matrix of the form $\sum_k [\mathbf{w}_k]_F [\mathbf{v}_k]_E^\top$, where $[\mathbf{v}_k]_F = (v_{k1}, ..., v_{kn})^\top, [\mathbf{w}_k]_E = (w_{k1}, ..., w_{kn})^\top$. The matrix $\sum_k [\mathbf{w}_k]_F [\mathbf{v}_k]_E^\top$ has ij entry $\sum_k w_{ki} v_{kj}$. Let $\mathbf{W}$ be the matrix whose ith column is $[\mathbf{w}_i]_F$ and $\mathbf{V}$ be the matrix whose ith column is $[\mathbf{v}_i]_E$. Then the ij entry of $\sum_k [\mathbf{w}_k]_F [\mathbf{v}_k]_E^\top$ is $\sum_k w_{ki} v_{kj} = \mathbf{W}_{:i} \cdot \mathbf{V}_{:j} = \mathbf{W}_{i:}^\top \cdot \mathbf{V}_{:j} = ij$ entry of $\mathbf{W}^\top \mathbf{V}$. Therefore $\sum_k [\mathbf{v}_k]_E [\mathbf{w}_k]_F^\top = \mathbf{W}^\top \mathbf{V}$.

References

- Books
  - Introduction to Smooth Manifolds by John Lee
  - Solution manual to Lee: `https://wj32.org/wp/wp-content/uploads/2012/12/Introduction-to-Sm` `pdf`
  - Chapter 4 of Differential Topology by Victor Guillemin and Alan Pollack
  - Chapter 7 of Mathematical Methods of Classical Mechanics by Vladimir Arnold
  - Mathematics for Physics by Michael Stone and Paul Goldbart
    - Look at Ch 11!
  - Vector Calculus, Linear Algebra, and Differential Forms by John Hamal Hubbard and Barbara Burke Hubbard
- Slanted indices (`https://math.stackexchange.com/questions/73171/index-notation-for-tensors-is`
- Exterior derivative
  - `https://mathoverflow.net/questions/10574/how-do-i-make-the-conceptual-transition-from`
  - `https://mathoverflow.net/questions/21024/what-is-the-exterior-derivative-intuitively`
  - `https://math.stackexchange.com/questions/1270673/geometric-intuition-about-the-exteri`
  - `https://math.stackexchange.com/questions/209241/exterior-derivative-vs-covariant-deri`
  - `https://math.stackexchange.com/questions/1908008/why-isnt-there-a-contravariant-deriv`
- Exterior powers (`https://kconrad.math.uconn.edu/blurbs/linmultialg/extmod.pdf`)

References not used

- Linear Algebra via Exterior Products (`https://www.google.com/books/edition/Linear_Algebra_Via_Exterior_Products/G1lpaPlErAIC?hl=en&gbpv=1&printsec=frontcover`)
- Good explanation of changing bases for (p, q) tensors (Ricci's transformation law) begins at bottom of p. 549 `https://cseweb.ucsd.edu/~gill/CILASite/Resources/15Chap11.pdf`

About this book

Main goals

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

Tentative prerequisites and reading advice

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

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

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

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

On the prominence of algebraic structure

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

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

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

Other

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

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

Contents

Notation

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

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

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

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

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

1

Review of logic, proofs, and functions

1.1 Propositions and predicates

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

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

1.1.1 Logical operators

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

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

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

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

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

P	not P
T	F
F	T

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

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

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

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

1.2 Quantificational logic

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

1.2.1 Quantifiers

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

$\forall x\ P(x)$	is the predicate which states "for all x, $P(x) \cong T$"
$\exists x\ \ Q(x)$	is the predicate which states "there exists an x such that $Q(x) \cong T$".

The predicate $\forall x\ P(x)$ statement is true exactly when $P(x)$ is is true for all x, and the predicate $\exists x\ Q(x)$ is true exactly when $Q(x)$ is true for one or more x. In this sense, $\forall$ is similar to *and* and $\exists$ is similar to *or*. Here's an example: the predicate $(\forall x\ x > 3)$ is false, while the predicate $(\exists x\ x > 3)$ is true[1].

Nested quantifers

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

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

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

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

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

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

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

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

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

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

Negating quantifiers

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

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

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

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

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

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

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

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

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

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

1.3 Implications with sets

1.3.1 Sets

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

1.3.2 Constructing sets

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

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

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

1.3.3 Implications

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

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

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

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

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

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

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

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

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

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

"Necessary" and "sufficient"

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

Quantifiers with set membership

We define

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

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

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

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

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

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

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

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

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

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

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

Common shorthand

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

Converses

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

1.3.4 "If and only if"

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

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

Here's the truth table for $\iff$.

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

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

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

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

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

1.3.5 The contrapositive and proof by contradiction

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

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

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

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

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

Proof by contradiction

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

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

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

1.4 Sets

1.4.1 Set equality, subsets

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

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

1.4.2 Common sets

We define notation for many common infinite sets.

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

1.4.3 Indexing sets

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

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

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

1.4.4 Union and intersection

We define the following operations on sets.

1.4.5 There is no universal set in ZFC

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

Other versions of set theory can handle

1.4.6 Set difference and set complement

DeMorgan's laws for sets

1.4.7 Ordered pairs

1.4.8 Cartesian product

1.5 Relations

inverse relations

inverse functions <=> bijection

1.5.1 Equivalence relations

Definition 1.1. (Equivalence relation).

Definition 1.2. (Quotient set).

1.5.2 Functions

Latin meanings of injective and surjective.

Definition 1.3. (Function).

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

domain, codomain, range

Definition 1.4. (Uniqueness).

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

Remark 1.5. (Well-definedness and uniqueness).
What well-definedness refers to
Any property which is unique is the output of a well-defined function

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

Definition 1.7. (Onto, surjection).

Definition 1.8. (Bijection).

left-inverse, right-inverse, relation to -jectivities
addition of fns, scaling of fns?, image-set
preimages (not neccess fns)

Definition 1.9. (Inverse function).

Theorem 1.10. (Invertible iff bijection).
Every one-to-one function has an invertible restriction.

1.5.3 Cardinality of sets

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

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

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

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

Definition 1.12. (Proof by induction).
Defn?

2

Review of linear algebra

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

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

This review chapter does not present “eigenstuff.” There are two core concepts taught in an introductory linear algebra class that do not appear in this chapter. The first of these core concepts, determinants, is treated extensively in Ch. 6. The second core concept, “eigenstuff,” is not treated because knowing it is not necessary for understanding the main content of this book. Traditional linear algebra texts explain “eigenstuff” quite well- consult one of the linear algebra texts mentioned in the preface (“About this book”) if you are interested.

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

2.1 Vector spaces, span, and linear independence

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.1.1 Span and linear independence

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

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

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

where the c_i's are some scalars in K.

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

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

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

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

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

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

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

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

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

2.1.2 Bases and dimension of vector spaces

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

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

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

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

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

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

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

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

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

Proof. We prove the lemma by induction on k.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.2 Linear functions

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

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

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

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

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

$$\mathbf{f}(\mathbf{v} + \mathbf{w}) = \mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w})$$
$$\mathbf{f}(c\mathbf{v}) = c\mathbf{f}(\mathbf{v}).$$

Another equivalent condition is

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

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

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

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

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

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

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

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

2.2.1 Kernel and image of a linear function

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

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

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

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

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

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

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

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

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

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

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

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

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

If V is finite-dimensional, then $\mathbf{f}^{-1}(\mathbf{0})$ is also finite-dimensional since $\mathbf{f}^{-1}(\mathbf{0}) \subseteq V$. Choose a basis $\{\mathbf{e}_1, ..., \mathbf{e}_k\}$ for the kernel. Then add vectors 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 $\dim(\mathbf{f}(V)) = \dim(V) - \dim(\mathbf{f}^{-1}(\mathbf{0})) = n - k$.

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

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 $\dim(\mathbf{f}(V)) = \dim(V) - \dim(\mathbf{f}^{-1}(\mathbf{0})) = n - k$, which is what we want.

So, 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 linearly dependent, i.e., that $d_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + ... + d_n\mathbf{f}(\mathbf{e}_n) = \mathbf{0}$ for some d_i's not all zero. By the linearity of $\mathbf{f}$, this is equivalent with $\mathbf{f}(d_{k+1}\mathbf{e}_{k+1} + ... + d_n\mathbf{e}_n) = \mathbf{0}$ for some d_i's not all zero. Thus $d_{k+1}\mathbf{e}_{k+1} + ... + d_n\mathbf{e}_n \in \mathbf{f}^{-1}(\mathbf{0}) = \text{span}(\mathbf{e}_1, ..., \mathbf{e}_k)$, which means $d_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + ... + d_n\mathbf{f}(\mathbf{e}_n) = q_1\mathbf{e}_1 + ... + q_k\mathbf{e}_k$ for some d_i's and q_i's not all zero. Then $d_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + ... + d_n\mathbf{f}(\mathbf{e}_n) = q_1\mathbf{e}_1 + ... + q_k\mathbf{e}_k$ for some d_i's and q_i's not all zero, i.e., $-(q_1\mathbf{e}_1 + ... + q_k\mathbf{e}_k) + d_{k+1}\mathbf{f}(\mathbf{e}_{k+1}) + ... + d_n\mathbf{f}(\mathbf{e}_n) = 0$ for some d_i's and q_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. □

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

Let V, W be vector spaces over a field K. If $\mathbf{f} : V \to W$ is a bijective linear function, then it is called a *linear isomorphism* or an *isomorphism (of vector spaces)*. (In the terminology of abstract algebra, a linear function is a *homomorphism* of vector spaces. An isomorphism is in general a bijective homomorphism).

When we have a linear isomorphism $\mathbf{f} : V \to W$, then, roughly speaking, all elements in V "interact" in the same way as their coorresponding elements in W, so V and W are in some sense the same vector space. For this reason, we often say that an element $\mathbf{v} \in V$ can be *identified* with an element $\mathbf{w} \in W$.

Specifically, the "interaction" among elements of V is mirrored by an "interaction" among elements of W as follows. If $\mathbf{f}(\mathbf{v}_1) = \mathbf{w}_1, \mathbf{f}(\mathbf{v}_2) = \mathbf{w}_2$, and c_1, c_2 are scalars in K, then, because $\mathbf{f}$ is linear, $\mathbf{f}$ sends the vector $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \in V$ to $c_1\mathbf{w}_1 + c_2\mathbf{w}_2 \in W$.

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

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

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

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

Proof.

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

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

Lemma 2.34. (Linear isomorphisms provide bases).

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

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

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

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

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

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

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

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

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

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

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

Thus, $\mathbf{f}$ is one-to-one iff $\dim(V) = \dim(\mathbf{f}(V))$. To complete the proof, we show that if $\mathbf{f}$ is one-to-one, $\dim(V) = \dim(W)$, and $\dim(V) = \dim(\mathbf{f}(V))$, then $\mathbf{f}$ is a linear isomorphism. So, assume these hypotheses are the case. Then $\dim(\mathbf{f}(V)) = \dim(W)$ and $\mathbf{f}(V) \subseteq W$, so it is straightforwardly shown that $\mathbf{f}(V) = W$. Thus $\mathbf{f}$ is onto, so it is a linear isomorphism. □

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

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

2.2.2 Inverse of a linear function

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

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

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

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

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

Theorem 2.40. (Dimensions of spaces must be equal when linear function invertible).

Let V and W be finite-dimensional, and consider an invertible linear function $\mathbf{f} : V \to W$. Then $\dim(V) = \dim(W)$.

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

The remaining theorems of this subsection are not necessary to review to understand the later content in this book, and are only presented for completeness.

- only invertible linear functions preserve linear independence
  - implies columns of a linear function's matrix must be linearly independent
- identity matrix
- inverse matrix as matrix of inverse
- RREF is identity iff invertible. proof: row operations correspond to invertible linear functions. since columns are LI iff invertible and (only) linear functions preserve LI, RREF must be identity iff invertible.

2.2.3 Matrices and coordinatization

$\star$ Coordinates relative to a basis $\star$

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.

Definition 2.41. (Coordinates of a vector relative to a basis). Given a vector $\mathbf{v} \in V$, we define $[\mathbf{v}]_E$ to be the vector in $K^{\dim(V)}$ that stores the *coordinates of* $\mathbf{v}$ *relative to the basis* E. Formally, $[\mathbf{v}]_E$ is the tuple of scalars

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

for which

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

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

Definition 2.42. (Linear function acting on a list of vectors). Consider a linear function $\mathbf{f} : V \to K^m$. We define the notation

$$\mathbf{f}(E) := \Big(\mathbf{f}(\mathbf{e}_1) \quad \dots \mathbf{f}(\mathbf{e}_n) \Big).$$

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

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

$$\mathbf{I}_V(E) = \Big(\mathbf{e}_1 \quad \dots \mathbf{e}_n \Big).$$

Theorem 2.43. Taking coordinates relative to a basis is an invertible linear operation. Put differently, $[\cdot]_E$ is an invertible linear function.

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

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

Now we show $[c\mathbf{v}]_E = c[\mathbf{v}]_E$. If $[\mathbf{v}]_E = \begin{pmatrix} ([\mathbf{v}]_E)_1 \\ \vdots \\ ([\mathbf{v}]_E)_n \end{pmatrix}$, then $\mathbf{v} = \sum_{i=1}^{n} ([\mathbf{v}]_E)_i \mathbf{e}_i$, so $c\mathbf{v} = c \sum_{i=1}^{n} ([\mathbf{v}]_E)_i \mathbf{e}_i = \sum_{i=1}^{n} c([\mathbf{v}]_E)_i \mathbf{e}_i$. Thus by definition of $[\cdot]_E$ we see $[\cdot]_E(c\mathbf{v}) = c[\mathbf{v}]_E$. Therefore $[\cdot]_E$ is linear.

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

Matrices as representative of linear functions

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

Derivation 2.44. ($\star$ Primitive matrix of a linear function $V \to K^m$ $\star$).

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

To start, consider a linear function $\mathbf{f} : V \to K^m$. Then from the definition of $[\cdot]_E$ (see Definition 2.41) we have $\mathbf{v} = \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{e}_i$, so

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

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

Why not just specify what $\mathbf{f}$ is by storing the transformed basis vectors? This exactly what we will do. We define the *matrix-vector product* between a *matrix* $\mathbf{A} = (a_{ij})$, which is a two-dimensional grid of scalars from K whose ij entry is denoted a_{ij}, and a *column vector* $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in K^{\dim(V)}$.

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

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

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

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

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

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

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

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

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

Derivation 2.45. ($\star$ Matrix of a linear function relative to bases $\star$).

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

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

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

We call this matrix *the matrix of* $\mathbf{f}$ *relative to the bases* E *and* F. This matrix is $[\mathbf{f}(E)]_F$, where we have used Definition 2.42 to define $[\mathbf{f}(E)]_F := [\cdot]_F(E)$. That is,

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

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

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

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

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

Definition 2.46. (Matrix of a linear function $V \to V$ relative to one basis).

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

Remark 2.47. (Matrix-vector product pedagogy). All the linear algebra texts I have read always present the relationship between linear functions and matrices in the following way: first define matrices as grids of scalars (often in the context of systems of linear equations), then define a linear function as satisfying the second condition of Definition 2.23, and then prove that each linear function has a matrix. This is bad pedagogy; there should be no need to conjecture and prove that a matrix-vector product corresponds to the action of a linear function, because this fact is apparent from Derivation 2.44.

It's important to emphasize that definition of matrix-vector product is therefore really no more than a definition. We decided that writing a grid of scalars next to a vector in the specific way of Derivation 2.44 should produce the output of Derivation 2.44 because this is what formalizes the correspondence between linear functions and the lists of their transformed basis vectors. Now, it is true that this definition of matrix-vector product leads to somewhat complicated formulas for the ith entry of a matrix-vector product and for the ij entry of a matrix-matrix product (these formulas will be presented in Theorems 2.54 and 2.55). Upon seeing these index-notation formulas, remember that they are consequences, and are not "just the way things are"!

Remark 2.48. (Primitive matrix as special case of matrix relative to bases). The primitive matrix of a linear function $\mathbf{f} : V \to K^m$ relative to E is the matrix $[\mathbf{f}(E)]_{\hat{\mathfrak{E}}} = \mathbf{f}(E)$ of $\mathbf{f} : V \to K^m$ relative to the bases E and $\hat{\mathfrak{E}}$, where $\hat{\mathfrak{E}}$ is the standard basis for K^m, and where $\mathbf{f}(E)$ is the matrix resulting from a linear function $V \to K^m$ acting on a list of vectors (see Definition 2.42). The key fact here is that $\left[\begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}\right]_{\hat{\mathfrak{E}}} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$ for $\begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} \in K^m$.

Theorem 2.49. (Matrices with respect to bases are also primitive matrices).

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

Let V and W be a finite-dimensional vector spaces with bases E and F, respectively. We know from Derivation 2.45 that every linear function $\mathbf{f} : V \to W$ corresponds to the matrix $[\mathbf{f}(E)]_F$ of $\mathbf{f}$ relative to E and F, where $[\mathbf{f}(E)]_F$ is characterized by the equation $[\mathbf{f}(\mathbf{v})]_F = [\mathbf{f}(E)]_F[\mathbf{v}]_E$. Rephrase this equation as $([\cdot]_F \circ \mathbf{f})(\mathbf{v}) = [\mathbf{f}(E)]_F[\mathbf{v}]_E$ and set $\mathbf{c} = [\mathbf{v}]_E$ to obtain

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

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

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

so that the above rephrases as

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

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

Looking at the characterizing property of a primitive matrix (see the last box of Derivation 2.44), we see that $[\mathbf{f}(E)]_F$ is the primitive matrix of $\mathbf{f}_{E,F} : K^{\dim(V)} \to K^{\dim(W)}$ relative to $\hat{\mathfrak{E}}$. Recalling from the previous remark that we can notate the primitive matrix of $\mathbf{f}_{E,F}$ relative to $\hat{\mathfrak{E}}$ by $\mathbf{f}_{E,F}(\hat{\mathfrak{E}})$, this fact can be stated as

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

Remark 2.50. ($\mathbf{f}_{E,F}$ as an induced function).

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

$$\begin{tikzcd} V \arrow[r, "\mathbf{f}"] \arrow[d, "{[\cdot]_E}"'] & W \arrow[d, "{[\cdot]_F}"] \\ K^{\dim(V)} \arrow[r, "\mathbf{f}_{E,F}"'] & K^{\dim(W)} \end{tikzcd}$$

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

Definition 2.51. ($\star$ Composition of linear functions that map into K^m and K^p $\star$).

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

The composition $\mathbf{g} \circ \mathbf{f} : V \to K^p$ is also a linear function (prove this as an exercise). Since $\mathbf{g} \circ \mathbf{f}$ is a linear function $V \to K^p$, its primitive matrix relative to E (see Derivation 2.44) is

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

But $\mathbf{f} : V \to K^m$ and $\mathbf{g} : K^m \to K^p$, so $\mathbf{f}$ and $\mathbf{g}$ also have primitive matrices. The primitive matrix of $\mathbf{f}$ relative to E is $[\mathbf{f}(E)]_{\hat{\mathfrak{E}}} = \mathbf{f}(E)$ and the primitive matrix of $\mathbf{g}$ relative to $\hat{\mathfrak{E}}$ is $[\mathbf{g}(\hat{\mathfrak{E}})]_{\hat{\mathfrak{F}}} = \mathbf{g}(\hat{\mathfrak{E}})$ (see Remark 2.48). The above then becomes

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

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

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

$$\boxed{\mathbf{BA} := \left(\mathbf{Ba}_1 \quad \dots \quad \mathbf{Ba}_n\right)}$$

so that

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

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

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

Remark 2.52. (Compatibility of matrices for matrix-matrix products). The composition of two linear functions is only defined when the output space of one is the entire input space of the other. Thus, the matrix-matrix product $\mathbf{BA}$ of an $m \times n$ matrix with an $r \times s$ matrix $\mathbf{B}$ is only defined when $r = n$.

Theorem 2.53. (Matrix-matrix products of linear functions mapping into finite-dimensional vector spaces).

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

The matrix of $\mathbf{g} \circ \mathbf{f}$ relative to E and G is the same as the primitive matrix for $(\mathbf{g} \circ \mathbf{f})_{E,G}$ relative to E (see Theorem 2.49). We will therefore compute this later matrix. We have

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

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

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

Therefore

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

Theorem 2.54. ($\star$ ith entry of matrix-vector product $\star$).

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix with entries in a field K and let $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in K^n$ be a column vector. Referring to the definition of matrix-vector product in Derivation 2.44, we see the matrix-vector product $\mathbf{Ac}$ has the following ith entry:

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

Therefore,

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

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

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

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

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

Theorem 2.55. ($\star$ ij entry of matrix-matrix product $\star$).

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix with entries in a field K and $\mathbf{B} = (b_{ij})$ be an $n \times p$ matrix with entries in K. Then the ij entry of the matrix-matrix product $\mathbf{BA}$ can be computed using the definition of matrix-matrix product (Theorem 2.53) and the previous theorem, which gives a formula for the ith entry of a vector:

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

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

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

Similarly as in the previous theorem, what we mean by "ith row of $\mathbf{B}$" in the boxed equation is "column vector that contains entries of ith row of $\mathbf{A}$."

Change of basis

Theorem 2.56. ($\star$ Change of basis for vectors $\star$).

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

To start, consider the special case $V = K^n$. Let $\mathbf{E}$ and $\mathbf{F}$ be the matrices with ith columns $\mathbf{e}_i$ and $\mathbf{f}_i$, respectively. Then for $\mathbf{c} \in K^n$ we have $\mathbf{F}[\mathbf{c}]_F = \mathbf{c} = \mathbf{E}[\mathbf{c}]_E$, so $\mathbf{c} = \mathbf{F}^{-1}\mathbf{E}[\mathbf{c}]_E$. By the definition of matrix-matrix multiplication (Definition 2.51), we have

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

where we have defined the notation $[\mathbf{E}]_F := [\cdot]_F(E)$. (The notation $[\cdot]_F(E)$ was defined in Definition 2.42).

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

The matrix $[\cdot]_F(E) = [\mathbf{E}]_F$ was the centerpiece of the above argument. To generalize, we notice that the primitive matrix of $[\cdot]_F$ relative to E is $[\cdot]_F(E) = [\mathbf{E}]_F$. Tthe characterizing property of primitive matrices (see the very end of Derivation 2.44) then implies that for any $\mathbf{v} \in V$, we have

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

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

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

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

Theorem 2.57. ($\star$). Let V be a vector space. The identity function $\mathbf{I}_V : V \to V$ on V satisfies $(\mathbf{I}_V)_{E,F}^{-1} = (\mathbf{I}_V)_{F,E}$. As a corollary, we have $[\mathbf{E}]_F^{-1} = [\mathbf{F}]_E$.

Proof. Given any bases E, F of V, Theorem 2.49 defines $\mathbf{f}_{E,F} := [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}$. Therefore $(\mathbf{I}_V)_{E,F} = [\cdot]_F \circ [\cdot]_E^{-1}$. Since the definition of $\mathbf{f}_{E,F}$ holds for any two bases of V, we can switch E and F to obtain $(\mathbf{I}_V)_{F,E} = [\cdot]_E \circ [\cdot]_F^{-1}$. The claim follows.

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

Theorem 2.58. (Change of basis for linear functions). Let V and W be finite-dimensional vector spaces. Let E, G be bases of V, let F, H be bases of W, and consider a linear function $\mathbf{f} : V \to W$. Then $\mathbf{f}_{E,F}$ and $\mathbf{f}_{G,H}$ are related by

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

This is because $\mathbf{f}_{E,F}$ was defined as $\mathbf{f}_{E,F} := [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}$. (See Theorem 2.49). But $[\cdot]_H \circ [\cdot]_{F^{-1}} = (\mathbf{I}_W)_{F,H}$ and $[\cdot]_E \circ [\cdot]_{G^{-1}} = (\mathbf{I}_V)_{G,F}$, so

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

We now translate the above equation into a statement about primitive matrices. Since the primitive matrix of a composition of functions is the product of matrices taken relative to the appropriate bases (see Theorem 2.53), we have

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

The last equality follows from the previous theorem.

Theorem 2.59. (Change of basis for linear functions for a common special case).

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

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

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

Theorem 2.60. ($\star$ Change of basis in terms of basis vectors $\star$).

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

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

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

Remark 2.61. (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 2.56. This approach for proving Theorem 2.56 was not used because it involves quite a bit more matrix algebra than the approach supplied in this text. However, it good to know that these theorems are equivalent.

2.3 The dot product and cross product

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

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

The cross product also comes with two common pedagogical problems. The first is that the complicated algebraic formula for the cross product is rarely explained. The second problem is that the "right hand rule" is never explicitly formalized. One common "explanation" for the right hand rule goes as follows: "you can use a 'left hand rule" if you want to, but then you'll have to account for a minus sign". This is a true statement, but it only relates the "right hand rule" with the "left hand rule"- it does not explain the fundamental reason why a right hand rule or left hand rule would emerge in the first place. The "right hand rule" is really a consequence of the concept of *orientation*, which is discussed in Chapter 6.

2.3.1 The dot product on $\mathbb{R}^n$

Definition 2.62. (Length of a vector in $\mathbb{R}^n$). Let V be an n-dimensional vector space, and let $\hat{\mathfrak{E}}$ be the standard basis for $\mathbb{R}^n$. In analogy to the Pythagorean theorem, we define the *length* of a vector $\mathbf{v} \in \mathbb{R}^n$ to be $||\mathbf{v}|| := \sqrt{\sum_{i=1}^n([\mathbf{v}]_{\hat{\mathfrak{E}}})_i^2}$.

Definition 2.63. (Unit vector hat notation). We define $\wedge : \mathbb{R}^n \to \mathbb{R}^n$ to be the function $\wedge(\mathbf{v}) = \frac{\mathbf{v}}{||\mathbf{v}||}$, and denote $\wedge(\mathbf{v}) := \hat{\mathbf{v}}$.

Definition 2.64. (Unsigned angle between vectors in $\mathbb{R}^2$).

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

Definition 2.65. ($\perp$ operator on $\mathbb{R}^2$). We define $\perp\colon \mathbb{R}^2 \to \mathbb{R}^2$ to be the rotation that rotates a vector $\mathbf{v}$ counterclockwise by $\frac{\pi}{2}$ radians. Specifically, the primitive matrix of $\perp$ relative to E is

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

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

Remark 2.66. When we consider $\wedge : \mathbb{R}^2 \to \mathbb{R}^2$ and $\perp\colon \mathbb{R}^2 \to \mathbb{R}^2$, we have $\perp \circ \wedge = \wedge \circ \perp$, so writing $\hat{\mathbf{v}}_\perp$ is unambiguous when $\mathbf{v} \in \mathbb{R}^2$.

Definition 2.67. (Orthogonal linear function on $\mathbb{R}^n$).
We say that a linear function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is *orthogonal* iff it preserves length and angle, i.e., iff

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

Example 2.68. Rotations are orthogonal linear functions.

Definition 2.69. (Vector projection). Let V be a vector space over K, and consider vectors $\mathbf{v}_1, \mathbf{v}_2, (\mathbf{v}_2)_\perp \in V$.
The *vector projection* of $\mathbf{v}_1$ onto $\mathbf{v}_2$ is the unique vector $\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2) := (v_1)_{||}\hat{\mathbf{v}}_2$ such that $\mathbf{v}_1 = (v_1)_{||}\hat{\mathbf{v}}_2 + (v_1)_\perp(\hat{\mathbf{v}}_2)_\perp$, where $(v_1)_{||}, (v_1)_\perp \in K$.

Remark 2.70. Note that, on $\mathbb{R}^2$, we have $\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2) = \text{proj}(\mathbf{v}_1 \to \hat{\mathbf{v}}_2)$ because $\hat{\hat{\mathbf{v}}}_2 = \hat{\mathbf{v}}_2$.

Lemma 2.71. (Rotated projection is projection of rotated vectors). Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$. If $\mathbf{f}$ is a rotation, then $||\mathbf{f}(\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = ||\text{proj}(\mathbf{f}(\mathbf{v}_1) \to \mathbf{f}(\mathbf{v}_2))||$.

Proof. We have $\mathbf{v}_1 = (v_1)_{||}\hat{\mathbf{v}}_{12} + (v_1)_\perp(\hat{\mathbf{v}}_{12})_\perp$, so $\mathbf{f}(\mathbf{v}_1) = (v_1)_{||}\mathbf{f}(\hat{\mathbf{v}}_2) + (v_1)_\perp\mathbf{f}((\hat{\mathbf{v}}_2)_\perp)$. The claim follows if we show (1) that $\mathbf{f}(\hat{\mathbf{v}}_2) = \widehat{\mathbf{f}(\mathbf{v}_2)}$ and (2) that $\mathbf{f}((\hat{\mathbf{v}}_2)_\perp) = \widehat{\mathbf{f}(\mathbf{v}_2)}_\perp$. (1) is true because rotations are length-preserving. (2) is true because rotations commute with each other and because rotations are length-preserving. $\square$

Lemma 2.72. (Length of a projection). Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$. Then the length of the projection of $\mathbf{v}_1$ onto $\mathbf{v}_2$ is $||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = ||\mathbf{v}_1|| \cos(\theta)$, where θ is the unsigned angle between $\mathbf{v}_1$ and $\mathbf{v}_2$.

Proof. The lemma holds in the special case when $\mathbf{v}_2 = \hat{\mathfrak{e}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; draw a right triangle to see this. For the general case, consider the rotation $\mathbf{f}$ that satisfies $\mathbf{f}(\hat{\mathbf{v}}_2) = \hat{\mathfrak{e}}_1$, that is, $\mathbf{f}(\mathbf{v}_2) = ||\mathbf{v}_2||\hat{\mathfrak{e}}_1$. Then because rotations are length-preserving and with use of the previous lemma, $(v_1)_{||} = ||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = ||\mathbf{f}(\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2))|| = ||\text{proj}(\mathbf{f}(\mathbf{v}_1) \to \mathbf{f}(\mathbf{v}_2))||$. This is the same as $||\text{proj}(\mathbf{f}(\mathbf{v}_1) \to ||\mathbf{v}_2||\hat{\mathfrak{e}}_1)|| = ||\text{proj}(\mathbf{f}(\mathbf{v}_1) \to \hat{\mathfrak{e}}_1)|| = ||\mathbf{f}(\mathbf{v}_1)|| \cos(\phi)$, where ϕ is the unsigned angle between $\mathbf{f}(\mathbf{v}_1)$ and $\hat{\mathfrak{e}}_1$. We have $||\mathbf{f}(\mathbf{v}_1)|| = \mathbf{v}_1$ because rotations are length-preserving, and $\phi = \theta$, where θ is the unsigned angle from $\mathbf{v}_1$ to $\mathbf{v}_2$, because rotations preserve angle. Therefore $||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = ||\mathbf{v}_1|| \cos(\theta)$. $\square$

Remark 2.73. Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$. Since cos is an even function, the length of the projection of $\mathbf{v}_1$ onto $\mathbf{v}_2$ is also $||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = ||\mathbf{v}_1|| \cos(|\theta|)$.

Definition 2.74. (Geometric dot product on $\mathbb{R}^2$). The *geometric dot product on* $\mathbb{R}^2$ is the function $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined by

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

Why do we care about the geometric dot product? The primary reason is that $\mathbf{v}_1 \cdot \mathbf{v}_2 = \text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$ when $||\mathbf{v}_2|| = 1$, so investigating the geometric dot product can tell us more about vector projections (we will indeed find that the dot product tells us something about projections in Theorem 2.92!). In physics, the dot product is also used to define the work done by a force $\mathbf{F}$ along a displacement $\Delta\mathbf{x}$: (work done by $\mathbf{F}$) $:= \mathbf{F} \cdot \Delta\mathbf{x}$.

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

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

Remark 2.75. Vectors in $\mathbb{R}^2$ are perpendicular when their geometric dot product is zero, since $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ implies $\theta = \frac{\pi}{2}$.

Lemma 2.76. (Projection onto a vector is a linear function).

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

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

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

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

□

Theorem 2.77. The geometric dot product on $\mathbb{R}^2$ is a bilinear function. That is, $(\mathbf{v}_1, \mathbf{v}_2) \mapsto \mathbf{v}_1 \cdot \mathbf{v}_2$ is linear in the argument $\mathbf{v}_1$ when $\mathbf{v}_2$ is fixed, and is linear in the argument $\mathbf{v}_2$ when $\mathbf{v}_1$ is fixed.

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

Remark 2.78. Because projection in $\mathbb{R}^2$ can now be defined in terms of the geometric dot product on $\mathbb{R}^2$, we can note that $\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$ is linear in $\mathbf{v}_1, \mathbf{v}_2$ when $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$.

Lemma 2.79. Consider $\mathbf{v} \in \mathbb{R}^2$ with $||\mathbf{v}|| = 1$. Then, in applying the formula $\mathbf{v} \cdot \mathbf{v} = ||\mathbf{v}||^2 \cos(\theta)$, we have

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

Derivation 2.80. (Algebraic dot product on $\mathbb{R}^2$). We can now derive an "algebraic" formula for the dot product, using its bilinearity (Theorem 2.77) together with the previous lemma.

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

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

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

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

Therefore

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

Lemma 2.81. (Orthogonal linear functions on $\mathbb{R}^2$ preserve algebraic dot product).
Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ be an orthogonal linear function. Then $\mathbf{f}$ preserves the algebraic dot product on $\mathbb{R}^2$: $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{f}(\mathbf{v}_1) \cdot \mathbf{f}(\mathbf{v}_2)$.

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

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

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

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

Theorem 2.82. (Algebraic dot product on $\mathbb{R}^2$ formula implies geometric dot product formula on $\mathbb{R}^2$).
We've used the bilinearity of the geometric dot product to prove the algebraic dot product formula. Now we show that we can derive the geometric dot product formula from the algebraic dot product formula. More specifically if $\hat{\mathfrak{E}}$ is the standard basis of $\mathbb{R}^2$, defining $\cdot : \mathbb{R}^2 \to \mathbb{R}^2$ by $\mathbf{v}_1 \cdot \mathbf{v}_2 = ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1 ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_1 + ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_2 ([\mathbf{v}_2]_{\hat{\mathfrak{E}}})_2$ implies $\mathbf{v}_1 \cdot \mathbf{v}_2 = ||\mathbf{v}_2||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2) = ||\mathbf{v}_1||||\mathbf{v}_2|| \cos(\theta)$, where θ is the unsigned angle between $\mathbf{v}_1$ and $\mathbf{v}_2$.

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

Orthogonal linear functions on $\mathbb{R}^2$ preserve the algebraic dot product on $\mathbb{R}^2$ (see the previous lemma), so
$\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{f}(\mathbf{v}_1) \cdot \mathbf{f}(\mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) \cdot ||\mathbf{v}_2||\hat{\mathfrak{e}}_1 = \begin{pmatrix} ([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathfrak{E}}})_1 \\ ([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathfrak{E}}})_2 \end{pmatrix} \cdot \begin{pmatrix} ||\mathbf{v}_2|| \\ 0 \end{pmatrix} = ||\mathbf{v}_2||([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathfrak{E}}})_1$. We have
$([\mathbf{f}(\mathbf{v}_1)]_{\hat{\mathfrak{E}}})_1 = \text{proj}(\mathbf{f}(\mathbf{v}_1) \to \hat{\mathfrak{e}}_1) = \text{proj}(\mathbf{f}(\mathbf{v}_1) \to \mathbf{f}(\hat{\mathbf{v}}_2)) = \text{proj}(\mathbf{f}(\mathbf{v}_1) \to \mathbf{f}(\mathbf{v}_2)) = \text{proj}(\mathbf{f}(\mathbf{v}_1) \to \widehat{\mathbf{f}(\mathbf{v}_2)})$
$= \text{proj}(\mathbf{f}(\mathbf{v}_1) \to \mathbf{f}(\mathbf{v}_2)) = \text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$, where the last equality is by Lemma 2.71.
Therefore $\mathbf{v}_1 \cdot \mathbf{v}_2 = ||\mathbf{v}_2||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$, which is the definition of the geometric dot product on $\mathbb{R}^2$. □

Remark 2.83. Most proofs of the above theorem use the law of cosines. I personally do not find the law of cosines intuitive, and believe it is best seen as a consequence of the equivalence between the geometric and algebraic dot product formulas. We prove the law of cosines in this way in Theorem 2.90.

Definition 2.84. (Dot product on $\mathbb{R}^n$). Now that we have motivated the algebraic dot product on $\mathbb{R}^2$ by proving the previous theorem, we have a sensible way to define a dot product on $\mathbb{R}^n$. We can't do this by generalizing the geometric dot product on $\mathbb{R}^2$ because it is not clear how to define the concept of "angle" in $\mathbb{R}^n$.

Let $\hat{\mathfrak{E}}$ be the standard basis of $\mathbb{R}^n$. We define $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

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

Theorem 2.85. (Dot product on $\mathbb{R}^n$ as matrix-matrix product). $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1^\top \mathbf{v}_2$. (Proof left as exercise).

Theorem 2.86. (Length in $\mathbb{R}^n$). We can notice that in $\mathbb{R}^3$, the length of a vector $\mathbf{v} \in \mathbb{R}^3$ expressed relative to the standard basis $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}\}_{i=1}^3$ is $\sqrt{\sum_{i=1}^3 ([\mathbf{v}]_{\hat{\mathfrak{E}}})_i^2} = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

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

Definition 2.87. (Angle in $\mathbb{R}^n$). The dot product on $\mathbb{R}^2$ satisfies $\mathbf{v}_1 \cdot \mathbf{v}_2 = ||\mathbf{v}_1|| ||\mathbf{v}_2|| \cos(\theta)$, so $\theta = \cos^{-1}\left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_1|| ||\mathbf{v}_2||}\right)$. We define angle in $\mathbb{R}^n$ in analogy to the dot product on $\mathbb{R}^2$. The angle between vectors in $\mathbb{R}^n$ is $\theta := \cos^{-1}\left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_1|| ||\mathbf{v}_2||}\right)$, where the dot product here is the dot product on $\mathbb{R}^n$.

Remark 2.88. (Geometric dot product on $\mathbb{R}^n$).

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

Definition 2.89. (Orthogonality of vectors in $\mathbb{R}^n$). We say that vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ are *orthogonal* iff the angle in $\mathbb{R}^n$ between $\mathbf{v}_1$ and $\mathbf{v}_2$ is $\frac{\pi}{2}$. So, orthogonality is a generalized notion of perpendicularity. Thus, $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ are orthogonal iff $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Theorem 2.90. (Law of cosines in $\mathbb{R}^n$).

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

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

The only reason this theorem was included was to demonstrate the point of Remark 2.83.

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

Remark 2.91. The above theorem reveals that the algebraic dot product on $\mathbb{R}^2$ can also be discovered as an orthogonality condition between vectors. When $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ are orthogonal, they form a right triangle, so Pythagorean theorem gives $||\mathbf{v}_1||^2 + ||\mathbf{v}_2||^2 = ||\mathbf{v}_1 - \mathbf{v}_2||^2$. Use $\mathbf{v}_1 = \sqrt{\sum_{i=1}^2 (([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_i)^2}$ to discover that we must have $([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_1 + ([\mathbf{v}_1]_{\hat{\mathfrak{E}}})_2 = 0$.

Theorem 2.92. (Vector projection in terms of algebraic dot product).

Theorem 2.93. (Vector projection is bilinear).

Proof. □

Cross product

See [...]

Part I

Multilinear algebra and tensors

3

Preview of differential forms

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

4

A motivated introduction to tensors

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

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

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

4.1 Multilinear functions and tensor product spaces

Definition 4.1. (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 a *k-linear function* iff for all $\mathbf{v}_1 \in V_1, ..., \cancel{\mathbf{v}_i \in V_i}, ..., \mathbf{v}_k \in V_k$, 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*. A 2-linear function is called a *bilinear function*.

Example 4.2. (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 4.3. (Vector space of multilinear functions).

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

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

Definition 4.4. (Tensor product space).

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

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

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

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

and

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

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

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

Remark 4.5. (Tensor terminology). Some authors use the word "tensor" to mean "$\binom{p}{q}$ tensor". (We have not defined $\binom{p}{q}$ tensors yet, but we will in Definition 4.29). We will use the word "tensor" to either mean an element of a tensor product space or a $\binom{p}{q}$ tensor, but we only do this when the meaning is clear from context.

Definition 4.6. (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 4.7. (Associativity of tensor product).

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

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

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

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

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

Theorem 4.8. (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

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

Proof. This set spans $V_1 \otimes ... \otimes V_k$ by definition of $V_1 \otimes ... \otimes V_k$. To show linear independence, assume that $\sum_{i=1}^{n} c_{i_1 ... i_n} \mathbf{e}_{i1} \otimes ... \otimes \mathbf{e}_{in_i}$ is the zero tensor. We must show that all of the $c_{i_1 ... i_n}$'s are 0.

come back later

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

Theorem 4.9. (Universal property of the tensor product).

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

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

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

Theorem 4.10. (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 $\mathbf{f} \in \mathcal{L}(V_1 \times V_2 \to W) \mapsto \mathbf{h} = \mathbf{f} \circ \mathbf{g}^{-1}$, where $\mathbf{g}$ and $\mathbf{h}$ were defined in the proof of Theorem 4.9. We already know this map is an injection because $\mathbf{h}$ is uniquely determined by $\mathbf{f}$ (see the proof of Theorem 4.9). It is a surjection because, given $\mathbf{h}$, we can choose $\mathbf{f}$ so that $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ (this is the condition that $\mathbf{h}$ uniquely satisfies); choose $\mathbf{f}$ so that $\mathbf{f}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{h}(\mathbf{v}_1 \otimes \mathbf{v}_2)$. It is linear because, given vector spaces Y, Z, W, the map $\circ$ which composes linear functions, $\circ : \mathcal{L}(Y \to Z) \times \mathcal{L}(Z \to W) \to \mathcal{L}(Y \to W)$, is a bilinear map. (Check this fact for yourself. The consequences of this are explored in [...]). Therefore the map $\mathcal{L}(V_1 \times V_2 \to W) \mapsto \mathcal{L}(V_1 \otimes V_2 \to W)$ described above is a linear isomorphism. □

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

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

The theorem we will discover is that when V and W are finite-dimensional vector spaces, there is a natural isomorphism $\mathcal{L}(V \to W) \cong W^* \otimes V$, 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 the dual space V^*, which (we will see) indicates that thinking of linear functions as vectors is involved, and it also includes the tensor product $\otimes$, which indicates that multilinear structure is involved.

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

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

Let $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ be the standard basis of $K^n = K^{\dim(V)}$ and let $\hat{\mathfrak{F}} = \{\hat{\mathfrak{f}}_1, ..., \hat{\mathfrak{f}}_m\}$ be the standard basis of $K^m = K^{\dim(W)}$. So, in the example, $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ and $\hat{\mathfrak{F}} = \{\hat{\mathfrak{f}}_1, \hat{\mathfrak{f}}_2, \hat{\mathfrak{f}}_3\}$. The first "big leap" is to notice that the $m \times n$ matrix with ij entry 1 and all other entries 0 is $\hat{\mathfrak{f}}_i \hat{\mathfrak{e}}_j^\top$, where $\hat{\mathfrak{f}}_i \hat{\mathfrak{e}}_j^\top$ is the product of a $m \times 1$ matrix with a $1 \times n$ matrix (see Theorem 2.55). This means that the above 3×2 matrix can be expressed as

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

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

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

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

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

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

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

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

Consider one of the matrices in the weighted sum, $\mathbf{g}_i \mathbf{h}_j^\top$. Initially, we may be tempted to directly investigate the linear function represented by $\mathbf{g}_i \mathbf{h}_j^\top$. This will work, but we can be even more specific; $\mathbf{g}_i \mathbf{h}_j^\top$ is a matrix-matrix product, so it corresponds to a composition of linear functions (see Theorem 2.51 and possibly Theorem 2.53). Asking "to what linear functions do $\mathbf{g}_i$ and $\mathbf{h}_j^\top$ correspond?" will prove fruitful.

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

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

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

Overall, we have shown that the matrix with respect to bases of a linear function $\mathbf{f} : V \to W$ can be expressed as a linear combination of the (primitive (see Derivation 2.44)) matrices that represent the linear maps $\mathbf{g}_i \circ \phi_{\mathbf{h}_j}$. Therefore, the linear function $\mathbf{f}$ is a linear combination of the linear functions $\mathbf{g}_i \circ \phi_{\mathbf{h}_j}$:

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

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

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

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

Theorem 4.11. Let V and W be vector spaces. Any rank-1 linear function $V \to W$ can be expressed as $\mathbf{w} \circ \phi$, for some $\mathbf{w} \in W$ and some linear function $\phi : V \to K$, where $\mathbf{w} : K \to W$ is the linear map defined by $\mathbf{w}(c) = c\mathbf{w}$.

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

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

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

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

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

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

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

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

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

Definition 4.12. (Dual space). Let V be a (not necessarily finite-dimensional) vector space over a field K. The *dual vector space* to V is the vector space over K, denoted V^*, consisting of the linear functions $V \to K$ under the operations of function addition and function scaling:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

As was alluded to before, the above is achieved due to the bilinearity of $\circ$ and the seeming-bilinearity[1] of $\otimes$:

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

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

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

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

So, we have proved the following theorem.

Theorem 4.13. ($\mathcal{L}(V \to W) \cong W \otimes V^*$ 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 2.38).

Remark 4.14. (Rank-1 linear transformations correspond to elementary tensors).

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

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

Remark 4.15. (Tensor product space as the structure behind composition).

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

Remark 4.16. (The two key ideas). Now that we have gone through the derivation, we can specifically see how the two key ideas of thinking of linear functions as vectors and "multilinear elements" have manifested.

We thought of the linear function $\mathbf{f} : V \to W$ 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 $\otimes$.

4.3 Introduction to dual spaces

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

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

Definition 4.17. (Dual space). Let V be a (not necessarily finite-dimensional) vector space over a field K. The *dual vector space* to V is the vector space over K, denoted V^*, consisting of the linear functions $V \to K$ under the operations of function addition and function scaling:

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

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

Derivation 4.18. (Induced dual basis, dimension of finite-dimensional V^*).

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

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

Since $V^* = \mathcal{L}(V \to K)$, any element $\phi \in V^*$ is represented relative to the basis E by the (primitive) $1 \times n$ matrix $\phi(E)$ (see Derivation 2.44 and Remark 2.48).

Recall from Definition 2.42 that the matrix $\phi(E)$ is

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

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

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

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

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

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

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

$\phi_{\mathbf{e}_i} :=$ the element of V^* that is represented relative to E by the (primitive) matrix $\hat{\mathfrak{e}}_i^\top$

$\phi_{\mathbf{e}_i}(\mathbf{v}) = \hat{\mathfrak{e}}_i^\top[\mathbf{v}]_E.$

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

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

So, in all, we have

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

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

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

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

Theorem 4.19. (Equivalent definition of dual basis).

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

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

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

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

Remark 4.20. (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 2.38) because it depends on how we choose the basis E for V. Additionally, the map $\mathbf{e}_i \mapsto \phi_{\mathbf{e}_i}$ is only a surjection when V is finite-dimensional, because when V is infinite-dimensional the cardinality of V^* is strictly greater than the cardinality of V.

Remark 4.21. (Not every basis for the dual space is an induced basis).

We don't have to pick a basis of V to pick a basis for V^*. Derivation 4.18 showed that when V is finite-dimensional, then V^* is finite-dimensional. Therefore, when V is finite-dimensional, we can pick an *arbitrary* basis for V^*.

Remark 4.22. (The misleading star notation for dual vectors).

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

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

Theorem 4.23. ($V \cong V^{**}$ naturally).

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

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

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

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

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

$\mathcal{L}_V$ is injective iff ($\phi(\mathbf{v}) = \mathbf{0}$ for all $\phi \in V^* \implies \mathbf{v} = \mathbf{0}$). Since V is finite-dimensional, we can show the contrapositive, ($\exists \phi \in V^*\ \phi(\mathbf{v}) = \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$) after choosing a basis for V and obtaining a dual basis for V^*. (In the infinite dimensional setting, we would need to assume the axiom of choice to get a basis for V). If $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is a basis for V, then each $\phi \in V^*$ is associated with the matrix $\phi(E)$, as shown in Derivation 4.18. We can choose ϕ to be such that $\phi(\mathbf{v}) \neq \mathbf{0}$ for $\mathbf{v} \neq \mathbf{0}$ by imposing that ϕ have a nontrivial kernel. □

Definition 4.24. (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 $\mathbf{f}^*(\chi) = \chi \circ \mathbf{f}$.

Theorem 4.25. (Dual transformation on finite-dimensional vector spaces is represented by transpose matrix).

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

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

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

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

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

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

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

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

In terms of functions, the above is expressed as

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

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

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

Remark 4.26. (Motivations for defining the dual transformation).

The previous theorem reveals a new way to motivate the definition of the dual transformation. The first motivated definition, which we have already seen in Definition 4.24, 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 4.27. (Dual of a composition).

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

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

(The proof is left as an exercise).

Theorem 4.28. probably will remove this

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

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\mathcal{L}_V"'] & W \arrow[d, "\mathcal{L}_W"] \\ V^{**} \arrow[r, "T^{**}"'] & W^{**} \end{tikzcd}.$$

4.4 $\binom{p}{q}$ tensors

Definition 4.29. ($\binom{p}{q}$ tensor).

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

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

Remark 4.30. (The four-fold nature of $\binom{p}{q}$ tensors).

- multilinear function $\leftrightarrow$ multilinear element (element of tensor product space)
- linear function $\leftrightarrow$ linear element (vector)

$\binom{p}{q}$ tensors are "multilinear elements" because of tensor product spaces but also "generalized linear transformations" (because of $\mathcal{L}(V \to V) \cong V \otimes V^*$).

Consider how this is true for vectors and for dual vectors. Vectors are 1-linear elements by definition, and they are less obviously "generalized linear transformations" because they are naturally identifiable with elements of V^{**}. Dual vectors are linear functions, and they are less obviously 1-linear elements because they form a vector space.

Generalized linear maps, Bonet and Wood?

Come back and make a good diagram

Definition 4.31. (Coordinates of a $\binom{p}{q}$ tensor).

Let V be a finite-dimensional vector space over a field K with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be the induced dual basis for V^*. The *coordinates of a* $\binom{p}{q}$ *tensor* $T \in T_q^p(V)$ *relative to* E *and* E^* are the scalars $T^{i_1 ... i_p}_{j_1 ... j_q} \in K$ for which

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

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

Remark 4.32. (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 5.39). This definition of tensor is clearly unmotivated, as it describes how tensors behave before explaining what they really are.

- (The more "concrete" but less insightful mathematical definition of a tensor). Mathematicians often define a $\binom{p}{q}$ tensor to be a multilinear map $(V^*)^{\times p} \times V^{\times q} \to K$. This definition is equivalent to the one we have used (we see why in Theorem 4.35), but it is less preferable because it obscures the concept of a "multilinear element" that tensor product spaces so nicely capture.

- (Another physicist's definition of a tensor). Physicists also occasionally define an "nth order tensor" to be[2] a linear map that sends a $(n-1)$ order tensor to a vector in V, where a tensor of order 2 is defined to be a linear map $V \to V$. This definition works because we have the natural isomorphism $T_1^1(V) = V \otimes V^* \cong \mathcal{L}(V \to V)$. Note also that, when a basis for V is fixed (which is often always done in physics, since often we have $V = \mathbb{R}^3$, so we can use the standard basis), there is no ambiguity when one says "second order tensor", as $T_0^2(V) \cong T_1^1(V) \cong T_2^0(V)$ due to the (unnatural) isomorphism $V \cong V^*$ that is obtained by choosing a basis (see Remark 4.20).

Definition 4.33. (Covariance and contravariance, coordinates of a $\binom{p}{q}$ tensor).

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

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

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

Definition 4.34. (Valence and order of a tensor). The *valence* of a $\binom{p}{q}$ tensor is the tuple $\binom{p}{q}$. The *order* of a $\binom{p}{q}$ tensor is $p + q$.

Theorem 4.35. (Four fundamental natural isomorphisms for $\binom{p}{q}$ tensors).

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

$$\boxed{\begin{gathered}\mathcal{L}(V_1 \times ... \times V_k \to W) \cong \mathcal{L}(V_1 \otimes ... \otimes V_k \to W) \\ \mathcal{L}(V \to W) \cong W \otimes V^* \\ (V \otimes W)^* \cong V^* \otimes W^*\end{gathered}}$$

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

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

so we have the natural isomorphism

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

Proof. The first line in the first box is Theorem 4.10, and the second line in the first box is Theorem 4.13. We need to prove the third line in the first box; we need to prove that *taking the dual distributes over the tensor product.*

We do so by defining an isomorphism in the "reverse" direction. We define this isomorphism on elementary tensors and extend linearly. Given $\phi \otimes \psi \in V^* \otimes W^*$, we produce the linear map $\mathbf{f}_{\phi\otimes\psi} \in (V \otimes W)^*$, where $\mathbf{f}_{\phi\otimes\psi} : V \otimes W \to K$ 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. □

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

Theorem 4.36. (Other useful natural isomorphisms for $\binom{p}{q}$ tensors).

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

$$V \otimes K \cong V$$
$$V \otimes W \cong W \otimes V.$$

The proof of this theorem is left as an exercise.

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

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

Theorem 4.37. ($\binom{1}{1}$ tensor corresponding to a matrix).

Let V and W be finite-dimensional vector spaces, and let E and F be bases for V and W, respectively. If (a_{ij}) is the matrix of $\mathbf{f}$ relative to E and F, then $\mathbf{f}$ coorresponds to the $\binom{1}{1}$ tensor $\sum_{ij} a_{ji}\epsilon_i \otimes \mathbf{f}_j$.

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

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

(You might first try sending $\mathbf{f}$ to $\sum_{ij} a_{ij}\epsilon_i \otimes \mathbf{f}_j$. This won't work, but you'll get the idea to try $\sum_{ij} a_{ji}\epsilon_i \otimes \mathbf{f}_j = \sum_i \left(\epsilon_i \otimes \sum_j a_{ji}\mathbf{f}_j\right) = \sum_i \epsilon_i \otimes \mathbf{f}(\mathbf{e}_i)$). □

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

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

5

Bilinear forms, metric tensors, and coordinates of tensors

5.1 Bilinear forms and metric tensors

Definition 5.1. (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 a k-linear function $V_1 \times ... \times V_k \to K$. (Unfortunately, the word "k-form" is reserved to mean *differential* k-form. We have not defined differential k-forms yet).

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

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

Remark 5.2. (Linear k-forms are naturally identified with $\binom{0}{k}$ tensors).

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

Definition 5.3. (Nondegenerate bilinear form, the natural musical isomorphisms).

Let V and W be finite-dimensional vector spaces. If we have a bilinear form B on V and W, then there are natural linear maps $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ defined 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 denote $\mathbf{v}^{\flat_1} := \flat_1(\mathbf{v})$ and $\mathbf{w}^{\flat_2} := \flat_2(\mathbf{w})$.

What would it take for $\flat_1$ and $\flat_2$ to be linear isomorphisms? Well, if we knew that $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ were linear injections, then we would have $\dim(V) \leq \dim(W)$ and $\dim(W) \leq \dim(V)$, so we would have $\dim(V) = \dim(W)$, that is, $\dim(V) = \dim(W^*) = \dim(W) = \dim(V^*)$. Then, since $\flat_1$ and $\flat_2$ would be linear injections between finite-dimensional vector spaces of the same dimension, surjectivity would then follow automatically and $\flat_1$ and $\flat_2$ would be linear isomorphisms (see Theorem 2.37).

Therefore, if $\flat_1$ and $\flat_2$ are injections, then they are linear isomorphisms. When are $\flat_1$ and $\flat_2$ injections? This is the case if and only if their kernels are $\{\mathbf{0}\}$. In other words, $\flat_1$ and $\flat_2$ are isomorphisms iff the bilinear form B satisfies

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

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

A bilinear form B that satisfies the above conditions is called *nondegenerate*. We have contrived nondegenerate bilinear forms to be those for which $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ are natural linear isomorphisms. Note that when $\flat_1$ and $\flat_2$ are isomorphisms, they are indeed natural because they do not depend on a choice of basis (see Definition 2.38). 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 $\sharp_2 = \flat_2^{-1} : V^* \to W$.

Definition 5.4. (The adjoint).

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 $\mathbf{g} : V \to W$ obtained by using the musical isomorphisms on the domain and codomain of the dual transformation $\mathbf{f}^* : W^* \to V^*$. We have $\mathbf{g} = \flat_1^{-1} \circ \mathbf{f}^* \circ \flat_2$, or, equivalently, $\mathbf{g} \circ \flat_2 = \mathbf{f}^* \circ \flat_1$. So, the following diagram commutes:

$$\begin{tikzcd}
V \arrow[r, "\mathbf{g}"] \arrow[d, "\flat_1"'] & W \arrow[d] \\
W^* \arrow[r, "\mathbf{f}^*"'] & V^*
\end{tikzcd}$$

The map $\mathbf{g} : V \to W$, which can intuitively be thought of as "the dual transformation after identifying $V \cong W^*$ and $W \cong V^*$", is called *the adjoint of* $\mathbf{f}$. Looking at the commutative diagram, we see that the adjoint satisfies $\mathbf{g}(\mathbf{v}_0)^{\flat_2} = \mathbf{f}^*(\mathbf{v}_0^{\flat_1})$. This condition further unravels after referring to the definitions of $\flat_1$ and $\flat_2$:

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

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

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

Theorem 5.5. (Induced bilinear form on the duals).

Let V and W be vector spaces. If B is a nondegenerate bilinear form on V and W, then there is an induced nondegenerate bilinear form $\widetilde{B}$ on W^* and V^*, $\widetilde{B} = B \circ R^{-1}$, where $R : V \times W \to W^* \times V^*$ is defined by $R(\mathbf{v}, \mathbf{w}) = (\mathbf{v}^{\flat_1}, \mathbf{w}^{\flat_2})$. The bilinear form $\widetilde{B}$ supplies natural isomorphisms $\widetilde{\flat}_1 : V^* \to W^{**}$, $\widetilde{\flat}_2 : W^* \to V^{**}$ defined by $\phi^{\widetilde{\flat}_1}(\psi) = \widetilde{B}(\psi, \phi)$ and $\psi^{\widetilde{\flat}_2}(\phi) = \widetilde{B}(\psi, \phi)$. (Check that $\widetilde{B}$ is actually a nondegenerate bilinear form is an exercise).

The usefulness of this theorem is not apparent until we present Theorem 5.27.

Definition 5.6. (Metric tensor).

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

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

Definition 5.7. (The notation $\natural = \flat_1 = \flat_2$).

When V is a finite-dimensional 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. This leads us to define $\natural := \flat_1 = \flat_2$ and $\natural^{-1} := \flat_1^{-1} = \sharp_1 = \flat_2^{-1} = \sharp_2$.

Definition 5.8. (Inner product).

A metric tensor g on V is an *inner product* on V iff it is also *positive-definite*, that is, iff it is a metric tensor that satisfies $(g(\mathbf{v},\mathbf{v}) \neq \mathbf{0} \iff \mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in V)$. (We automatically have the reverse implication for any bilinear form g for the reasons discussed in Definition 5.3).

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

Definition 5.9. (Vector space with inner product). Let V be a vector space over K. Iff there is an inner product $\langle \cdot, \cdot \rangle$ on V, then V is called a *vector space with inner product*, or an *inner product space.*

Remark 5.10. (Positive-definite $\implies$ nondegenerate, but the converse does not hold).

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

Example 5.11. The dot product on $\mathbb{R}^n$ is an inner product on $\mathbb{R}^n$. (Proof left as exercise).

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

5.1.1 Length and orthogonality with respect to an inner product

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

Definition 5.12. (Length of a vector with respect to an inner product).

In analogy to the fact that the length of a vector in $\mathbb{R}^n$ can be expressed using the dot product on $\mathbb{R}^n$ (see Theorem 2.86), 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 5.13. (Angle between vectors with respect to an inner product).

In analogy to the definition of the angle between vectors in $\mathbb{R}^n$, (see Definition 2.87), we define the *angle* θ *between vectors* $\mathbf{v}_1, \mathbf{v}_2 \in V$ *with respect to the inner product on* V to be $\cos^{-1}\left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_1|| ||\mathbf{v}_2||}\right)$.

Definition 5.14. (Adjoint of a linear function $V \to W$).

Let V and W be vector spaces with inner products $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_W$, respectively. Consider a linear function $\mathbf{f} : V \to W$, and let $\mathbf{f}^* : W^* \to V^*$ be its adjoint (see Definition 5.4). The condition on $\langle \cdot, \cdot \rangle$ induced by using the musical isomorphism $\natural : V \to V^*$ to identify $V \cong V^*$ (see the end of Definition 5.4) is

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

Remark 5.15. (Geometric inner product).

Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$. Then $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = ||\mathbf{v}_1|| \, ||\mathbf{v}_2|| \cos(\theta)$, where θ is the angle between $\mathbf{v}_1$ and $\mathbf{v}_2$ with respect to the inner product on V. (This fact is the generalization of the geometric dot product on $\mathbb{R}^n$, which was discussed in Remark 2.88).

Theorem 5.16. (Cauchy-Schwarz inequality for vector spaces over $\mathbb{R}$).

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

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

Proof. Define $f : \mathbb{R} \to [0, \infty) \subseteq \mathbb{R}$ by $f(c) = \langle c\mathbf{v}_1 + \mathbf{v}_2, c\mathbf{v}_1 + \mathbf{v}_2 \rangle = c^2 \langle \mathbf{v}_1, \mathbf{v}_1 \rangle + 2c \langle \mathbf{v}_1, \mathbf{v}_2 \rangle + \langle \mathbf{v}_2, \mathbf{v}_2 \rangle$. Since $\langle \cdot, \cdot \rangle$ is positive-definite, $f(c) \geq 0$, with $f(c) = 0$ only when $c\mathbf{v}_1 + \mathbf{v}_2 = 0$. Since f is nonnegative, then it must have either one or zero real roots, meaning $b^2 - 4ac = (2\langle \mathbf{v}_1, \mathbf{v}_2 \rangle)^2 - 4(\langle \mathbf{v}_1, \mathbf{v}_2 \rangle)(\langle \mathbf{v}_2, \mathbf{v}_2 \rangle) \leq 0$. 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 5.17. (Orthogonality of vectors with respect to an inner product).

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

Definition 5.18. (Orthonormal basis).

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

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

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

Theorem 5.19. (Gram-Schmidt algorithm).

5.1.2 Symmetric and orthogonal linear functions

This subsection is presented only completeness, so reading the entirety of this subsection is not necessary. The only results that are necessary to know are the conditions (3) and (4) of Definition 5.22 satisfied by an orthogonal linear function.

Let V be a vector space with inner products $\langle \cdot, \cdot \rangle$. Consider a linear function $\mathbf{f} : V \to V$, and let $\mathbf{f}^* : V \to V$ be its adjoint (see Definition 5.4). The defining condition of the adjoint (see the end of Definition 5.4) is

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

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

Definition 5.20. (Transpose of a matrix).

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

Definition 5.21. (Symmetric linear function).

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

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

Proof.
($1 \iff 2$).

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

($\impliedby$). We have $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{f}(\mathbf{v}_2) \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $\langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle$. Therefore $\langle \mathbf{v}_1, \mathbf{f}(\mathbf{v}_2) \rangle = \langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle$. Due to the cancelability of inner products (see Theorem **??**), we have $\mathbf{f}(\mathbf{v}_2) = \mathbf{f}^*(\mathbf{v}_2)$ for all $\mathbf{v}_2 \in V$. So $\mathbf{f} = \mathbf{f}^*$.
($2 \iff 3a$).

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

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

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

Definition 5.22. (Orthogonal linear function).

Let V be a vector space over K (where[1] we have $K \neq \mathbb{Z}/2\mathbb{Z}$), consider a linear function $\mathbf{f} : V \to V$, and let $\mathbf{f}^*$ be its adjoint. We define $\mathbf{f}$ to be *orthogonal* iff the following equivalent conditions hold:

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

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

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

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

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

(5 $\implies$ 3). Replace the dot product $\cdot$ on $\mathbb{R}^n$ with the inner product $\langle \cdot, \cdot \rangle$ on V in the equation stated in the proof of Lemma 2.81 to show that $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \frac{1}{2}(||\mathbf{v}_1 + \mathbf{v}_2||^2 - (||\mathbf{v}_1||^2 + ||\mathbf{v}_2||^2))$. That is, the inner product $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ is a function of $||\mathbf{v}_1||$ and $||\mathbf{v}_2||$. Applying the previous formula, the inner product $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_2) \rangle$ is a function of $||\mathbf{f}(\mathbf{v}_1)||$ and $||\mathbf{f}(\mathbf{v}_2)||$: $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_2) \rangle = \frac{1}{2}(||\mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2)||^2 - (||\mathbf{f}(\mathbf{v}_1)||^2 + ||\mathbf{f}(\mathbf{v}_2)||^2))$. Since $\mathbf{f}$ is linear, this becomes $\frac{1}{2}(||\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2)||^2 - (||\mathbf{f}(\mathbf{v}_1)||^2 + ||\mathbf{f}(\mathbf{v}_2)||^2))$. If length is preserved, then this is the same as $\frac{1}{2}(||\mathbf{v}_1 + \mathbf{v}_2||^2 - (||\mathbf{v}_1||^2 + ||\mathbf{v}_2||^2) = \langle \mathbf{v}_1, \mathbf{v}_2 \rangle$.

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

(1 $\iff$ 2).

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

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

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

(3 $\iff$ 6).

($\implies$). 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 $\hat{U}$ 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 $[\mathbf{f}(\hat{\mathbf{u}}_i)]_{\hat{U}}$ of the matrix of $\mathbf{f}$ relative to $\hat{U}$ are orthonormal.

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

(6 $\iff$ 7).

($\implies$). The ij entry of $[\mathbf{f}(\hat{U})]_{\hat{U}} [\mathbf{f}(\hat{U})]_{\hat{U}}^\top$ is (ith row of $[\mathbf{f}(\hat{U})]_{\hat{U}}$) $\cdot$ (jth column of $[\mathbf{f}(\hat{U})]_{\hat{U}})\top$ =

[1] 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.

(ith row of $[\mathbf{f}(\hat{U})]_{\hat{U}}$) $\cdot$ (jth row of $[\mathbf{f}(\hat{U})]_{\hat{U}}$) $= \langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \delta^i_j$.

($\Longleftarrow$). 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) $\Longrightarrow$ (6) $\Longrightarrow$ (7).

(5 $\Longrightarrow$ 1). Use the fact that $[\mathbf{f}(\hat{U})]^\top_{\hat{U}}$ is the matrix of the adjoint $\mathbf{f}^* : V \to V$. $\square$

5.1.3 Self-duality and choice of basis

Definition 5.23. (Self-dual basis).

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

Theorem 5.24. (For vector spaces over $\mathbb{R}$, a musical isomorphism is the unique isomorphism enabling self-duality of a basis).

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

This theorem sheds light on the "unnatural" isomorphism $V \to V^*$ which sends $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ that was discussed in Remark 4.20). We said that the isomorphism which sends $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ is not natural because it depends on a choice of basis. (This is still true). Recently, we have seen that a nondegenerate bilinear form, such as an inner product, on V induces the natural musical isomorphism $\natural : V \to V^*$. This theorem shows that the unnatural isomorphism $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ can be viewed as arising from the usually natural but now, in this case, "unnatural" musical isomorphism $\natural : V \to V^*$ that arises from choosing a basis E for V.

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

Theorem 5.25. (Self-dual $\Longleftrightarrow$ orthonormal).

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

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

We have: self-dual $\Longleftrightarrow$ $\mathbf{e}_i^\natural = \phi^{\mathbf{e}_i}$ $\Longleftrightarrow$ $\mathbf{e}_i^\natural(\mathbf{e}_j) = \phi^{\mathbf{e}_i}(\mathbf{e}_j)$ $\Longleftrightarrow$ $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \phi^{\mathbf{e}_i}(\mathbf{e}_j)$ $\Longleftrightarrow$ $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta^i_j$ $\Longleftrightarrow$ orthonormal. $\square$

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

Key ideas:

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

Theorem 5.26. (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

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

Recall from Theorem 4.23 that $\mathbf{v}$ is identified with $\Phi_{\mathbf{v}} \in V^{**}$, where $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$.

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

$$\phi^{\mathbf{e}_i}(\mathbf{v}) = \phi^{\mathbf{e}_i}\Big(\sum_{j=1}^{n}([\mathbf{v}]_E)^j \mathbf{e}_i\Big) = \sum_{j=1}^{n}\Big(([\mathbf{v}]_E)^j \phi^{\mathbf{e}_i}(\mathbf{e}_j)\Big) = \sum_{j=1}^{n}([\mathbf{v}]_E)^j \delta^i_j = ([\mathbf{v}]_E)^i.$$

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

$$\phi(\mathbf{e}_i) = \Big(\sum_{j=1}^{n}([\phi]_{E^*})_j \phi^{\mathbf{e}_j}\Big)(\mathbf{e}_i) = \sum_{j=1}^{n}\Big(([\phi]_{E^*})_j \phi^{\mathbf{e}_j}(\mathbf{e}_i)\Big) = \sum_{j=1}^{n}([\phi]_{E^*})_j \delta^j_i = ([\phi]_{E^*})_i.$$

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

1. Recall from Theorem 4.23 that the definition of $\Phi_{\mathbf{v}}$ is $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$. We immediately get the result from this.
2. Since $\mathbf{v} \mapsto \Phi_{\mathbf{v}}$ is an isomorphism (see Theorem 4.23), then $([\mathbf{v}]_E)^i = ([\Phi_{\mathbf{v}}]_{E^{**}})^i$. By the second line, we have $([\Phi_{\mathbf{v}}]_{E^{**}})^i = \Phi_{\mathbf{v}}(\phi^{\mathbf{e}_i})$. Thus $([\mathbf{v}]_E)^i = \Phi_{\mathbf{v}}(\phi^{\mathbf{e}_i})$.

□

Theorem 5.27. (Relationship between coordinates of vectors and dual vectors for vector spaces).

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

Let V and W be finite-dimensional vector spaces with bases E and F, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_m}\}$ be corresponding induced dual bases of V^* and W^*. Let g be a metric tensor on V and W, and let $\widetilde{g}$ be the induced metric tensor on W^* and V^* (see Theorem 5.5). Let $\flat_1 : V \to W^*$ and $\flat_2 : W \to V^*$ be the musical isomorphisms induced by g.

Define the $n \times m$ matrices

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

Then

$$\begin{aligned} ([\mathbf{v}]_E)^i &= \sum_{j=1}^{n} g^{ij}([\mathbf{v}^{\flat_1}]_{F^*})_j = (i\text{th row of } \widetilde{\mathbf{g}}) \cdot [\mathbf{v}^{\flat_1}]_{F^*} \\ [\mathbf{v}]_E &= \widetilde{\mathbf{g}}[\mathbf{v}^{\flat_1}]_{F^*} \\ ([\mathbf{v}^{\flat_1}]_{F^*})_i &= \sum_{j=1}^{n} g_{ij}([\mathbf{v}]_E)^j = (i\text{th row of } \mathbf{g}) \cdot [\mathbf{v}]_E \\ [\mathbf{v}^{\flat_1}]_{F^*} &= \mathbf{g}[\mathbf{v}]_E. \end{aligned}$$

We will explain the significance of the lower indices in g_{ij} and the upper indices in g^{ij} in Remark 5.32.

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

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

$$([\mathbf{v}^{\flat_1}]_{F^*})_i = \mathbf{v}^{\flat_1}(\mathbf{f}_i) = g(\mathbf{v}, \mathbf{f}_i) = g\Big(\sum_{j=1}^n ([\mathbf{v}]_E)^j \mathbf{e}_j, \mathbf{f}_i\Big) = \sum_{j=1}^n ([\mathbf{v}]_E)^j g(\mathbf{e}_j, \mathbf{f}_i) = \sum_{j=1}^n ([\mathbf{v}]_E)^j g_{ji} = \sum_{j=1}^n g_{ij}([\mathbf{v}]_E)^j.$$

We prove the first equation by using the induced metric tensor $\widetilde{g}$ on V^*. Since $\mathbf{v} \overset{\flat_1}{\mapsto} \mathbf{v}^{\flat_1} \in W^*$ and $\mathbf{v}^{\flat_1} \overset{\widetilde{\flat}_1}{\mapsto} (\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1} \in V^{**}$, where these maps are isomorphisms, we have $([\mathbf{v}]_E)^i = ([(\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1}]_{E^{**}})^i$. Use the fact $[(\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1}]_{E^{**}})^i = (\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1}(\phi^{\mathbf{e}_i})$, which is implied by the second equation on the first line inside the box of the previous theorem, to get the first equation:

$$\begin{aligned}
([\mathbf{v}]_E)^i &= ([(\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1}]_{E^{**}})^i = (\mathbf{v}^{\flat_1})^{\widetilde{\flat}_1}(\phi^{\mathbf{e}_i}) = \widetilde{\flat}_1(\mathbf{v}^{\flat_1})(\phi^{\mathbf{e}_i}) = \widetilde{g}(\mathbf{v}^{\flat_1}, \phi^{\mathbf{e}_i}) = \widetilde{g}\Big(\sum_{j=1}^n ([\mathbf{v}^{\flat_1}]_{F^*})_j \psi^{\mathbf{f}_j}, \phi^{\mathbf{e}_i}\Big) \\
&= \sum_{j=1}^n ([\mathbf{v}^{\flat_1}]_{F^*})_j \widetilde{g}(\psi^{\mathbf{f}_j}, \phi^{\mathbf{e}_i}) = \sum_{j=1}^n ([\mathbf{v}^{\flat_1}]_{F^*})_j g^{ji} = \sum_{j=1}^n g^{ij}([\mathbf{v}^{\flat_1}]_{F^*})_j.
\end{aligned}$$

□

Lemma 5.28. (Primitive matrix of element of V^* as transposed coordinates).

Let V be a finite-dimensional vector space, let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be a basis for V, and consider an element $\phi \in V^*$. Then the (primitive) matrix $\phi(E)$ of ϕ relative to E (see Remark 2.48) is $\phi(E) = [\phi]_{E^*}^\top$.

Proof.

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

. □

Lemma 5.29. Let V and W be finite-dimensional vector spaces with bases E and F, respectively, let g be a metric tensor on V, and let $\widetilde{g}$ be the induced metric tensor on W^* and V^*. Then

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

Most importantly, since we earlier defined $\widetilde{\mathbf{g}} := (\widetilde{g}(\psi^{\mathbf{f}_i}, \phi^{\mathbf{e}_j}))$, we have that $\widetilde{\mathbf{g}} = \mathbf{g}^{-1}$ as a consequence. (Use $\psi = \psi^{\mathbf{f}_i}$ and $\phi = \phi^{\mathbf{e}_i}$ in the above to see this). That is,

$$\boxed{(g_{ij})^{-1} = (g^{ij})}$$

Proof. We show the first equation of each line. The second equation on each line follows by transposing the first equation on each line and using the fact that $\mathbf{g}$ and $\widetilde{\mathbf{g}}$ are symmetric matrices.

First, we show the first equation of the first line. We have $g(\mathbf{v}, \mathbf{w}) = \mathbf{v}^{\flat_1}(\mathbf{w})$. Notice that the previous lemma implies that the $1 \times m$ primitive matrix $\mathbf{v}^{\flat_1}(F)$ of $\mathbf{v}^{\flat_1} : W \to K$ relative to F is $\mathbf{v}^{\flat_1}(F) = [\mathbf{v}^{\flat_1}]_{F^*}^\top$. Applying the characterizing property of primitive matrices (see Derivation 2.44), we have $g(\mathbf{v}, \mathbf{w}) = \mathbf{v}^{\flat_1}(\mathbf{w}) = \mathbf{v}^{\flat_1}(F)[\mathbf{w}]_F = [\mathbf{v}^{\flat_1}]_{F^*}^\top [\mathbf{w}]_F$. So $g(\mathbf{v}, \mathbf{w}) = [\mathbf{v}^{\flat_1}]_{F^*}^\top [\mathbf{w}]_F$. The fourth equation of Theorem 5.27 states that $[\mathbf{v}^{\flat_1}]_{F^*} = \mathbf{g}[\mathbf{v}]_E$. Thus $g(\mathbf{v}, \mathbf{w}) = (\mathbf{g}[\mathbf{v}]_E)^\top [\mathbf{w}]_F = [\mathbf{v}]_E^\top \mathbf{g} [\mathbf{w}]_F$, which is the first equation of the first line.

Now we show the first equation of the second line. We have $\widetilde{g}(\psi, \phi) = g(\flat_1^{-1}(\psi), \flat_2^{-1}(\phi))$ by definition of $\widetilde{g}$ (see Theorem 5.5). We need to find the $\mathbf{v} \in V$ and $\mathbf{w} \in W$ for which $\mathbf{v} = \flat_1^{-1}(\psi)$ and $\mathbf{w} = \flat_2^{-1}(\phi)$, and then compute $\widetilde{g}(\psi, \phi) = g(\mathbf{v}, \mathbf{w}) = [\mathbf{v}]_E^\top \mathbf{g}[\mathbf{w}]_F$.

As was noted in the first part of this proof, $[\psi]_{F^*} = \mathbf{g}[\mathbf{v}]_E$ due to Theorem 5.27. Therefore $\flat_1^{-1}(\psi) = \mathbf{v}$, where $[\mathbf{v}]_E = \mathbf{g}^{-1}[\psi]_{F^*}$. Similarly, $\flat_2^{-1}(\phi) = \mathbf{w}$, where $[\mathbf{w}]_F = \mathbf{g}^{-1}[\phi]_{E^*}$.

Therefore $\widetilde{g}(\psi, \phi) = g(\mathbf{v}, \mathbf{w}) = [\mathbf{v}]_E^\top \mathbf{g}[\mathbf{w}]_F = (\mathbf{g}^{-1}[\psi]_{F^*})^\top \mathbf{g}(\mathbf{g}^{-1}[\phi]_{E^*}) = [\psi]_{F^*}^\top \mathbf{g}^{-1}\mathbf{g}\mathbf{g}^{-1}[\phi]_{E^*} = [\psi]_{F^*}\mathbf{g}[\phi]_{E^*}$. In this last step, we used that $\mathbf{g}^{-1}$ is symmetric. □

Theorem 5.30. (Relationship between coordinates of vectors and dual vectors).

Consider the hypotheses of Theorem 5.27. The fact that $(g_{ij})^{-1} = (g^{ij})$, proved in the previous lemma, immediately implies that

$$\boxed{\begin{aligned}
([\mathbf{v}]_E)^i &= \sum_{j=1}^n g^{ij}([\mathbf{v}^{\flat_1}]_{F^*})_j = (i\text{th row of } \mathbf{g}^{-1}) \cdot [\mathbf{v}^{\flat_1}]_{F^*} \\
[\mathbf{v}]_E &= \mathbf{g}^{-1}[\mathbf{v}^{\flat_1}]_{F^*} \\
([\mathbf{v}^{\flat_1}]_{F^*})_i &= \sum_{j=1}^n g_{ij}([\mathbf{v}]_E)^j = (i\text{th row of } \mathbf{g}) \cdot [\mathbf{v}]_E \\
[\mathbf{v}^{\flat_1}]_{F^*} &= \mathbf{g}[\mathbf{v}]_E
\end{aligned}}$$

Remark 5.31. suppose $V = W$

Special case of orthonormal bases: $g_{ij} = \delta_{ij}$. Then have $[\mathbf{v}^\natural]_{E^*} = \mathbf{g}[\mathbf{v}]_E = [\mathbf{v}]_E$, so $(\mathbf{v}^\natural)(\mathbf{v}_0) = [\mathbf{v}]_E^*(\mathbf{v}_0)$.

Remark 5.32. (Lower indices in g_{ij} and upper indices in g^{ij}).

We use lower indices in the notation $g_{ij} := g(\mathbf{e}_i, \mathbf{e}_j)$ because a metric tensor on V such as g is an element of $\mathcal{L}(V \times V \to K)$, so it can be identified (see Remark 5.2) with a $\binom{0}{2}$ tensor. We use lower indices for $g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j)$ to follow the index conventions of Definition 4.33.

Also... while $\mathbf{g}$ is covariant when $V = W$, $\mathbf{g}^{-1}$ is contravariant when $V = W$: when $V = W$, a bilinear form $(W^* \times V^*) \cong (V^* \times V^*) \to F$ is identifiable with a linear function $V^* \otimes V^* \to F$, which is an element of $(V^* \otimes V^*)^* \cong W \otimes V \cong T_0^2(V)$).

Remark 5.33. δ^i_j vs. δ^{ij} vs. δ^i_j

5.2.1 Change of basis for $\binom{p}{q}$ tensors

Definition 5.34. (Coordinates of a $\binom{p}{q}$ tensor).

For convenience, we now restate Definition 4.31.

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be the induced dual basis for V^*. The *coordinates of a $\binom{p}{q}$ tensor $T \in T_q^p(V)$ relative to E and E^** are the scalars $T^{i_1...i_p}_{j_1...j_q}$ for which

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

Lemma 5.35. Let V be a finite-dimensional vector space with bases E and F, and let E^* and F^* be the corresponding induced dual bases for V^*. Then

$$([\phi]_{F^*})_i = [\phi]_{E^*}^\top [\mathbf{f}_i]_E.$$

Proof. Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$. Then, starting with the fact $([\phi]_{F^*})_i = \phi(\mathbf{f}_i)$ (see Theorem 5.26), we have

$$([\phi]_{F^*})_i = \phi(\mathbf{f}_i) = \phi\Big(\sum_{j=1}^n ([\mathbf{f}_i]_E)_j \mathbf{e}_j\Big) = \sum_{j=1}^n ([\mathbf{f}_i]_E)_j \phi(\mathbf{e}_j) = \sum_{j=1}^n ([\mathbf{f}_i]_E)_j ([\phi]_{E^*})_i$$
$$= \Big(([\phi]_{E^*})_1 \quad \ldots \quad ([\phi]_{E^*})_n\Big) [\mathbf{f}_i]_E = [\phi]_{E^*}^\top [\mathbf{f}_i]_E.$$

□

Theorem 5.36. (Change of basis for vectors and dual vectors).

Let V be a finite-dimensional vector space with bases E and F, and let E^* and F^* be the corresponding induced dual bases for V^*. Then

$$\boxed{\begin{aligned} [\mathbf{v}]_F &= [\mathbf{E}]_F [\mathbf{v}]_E = [\mathbf{F}]_E^{-1} [\mathbf{v}]_E \\ [\phi]_{F^*} &= [\mathbf{E}]_F^{-\top} [\phi]_{E^*} = [\mathbf{F}]_E^\top [\phi]_{E^*} \\ [\mathbf{E}]_F^{-1} &= [\mathbf{F}]_E \\ [\mathbf{E}^*]_{F^*} &= [\mathbf{E}]_F^{-\top} \end{aligned}}$$

where $\mathbf{v} \in V$ and $\phi \in V^*$.

Proof. The first line of the boxed equation is Theorem 2.56, and the third line is Theorem 2.57. We will prove $[\mathbf{E}^*]_{F^*} = [\mathbf{E}]_F^{-\top}$, which is the fourth line. The second line then follows by applying the fourth line to the equation $[\phi]_{F^*} = [\mathbf{E}^*]_{F^*}[\mathbf{v}]_{E^*}$ (this last equation is implied by the first line).

For the proof, let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$, and $E^* = \{\epsilon^1, ..., \epsilon^n\}$. We show $[\mathbf{E}^*]_{F^*} = [\mathbf{E}]_F^{-\top}$ by examining the ij entry of $[\mathbf{E}^*]_{F^*}$ and making use of the previous lemma. We have

$$([\mathbf{E}^*]_{F^*})^i_j = [\epsilon^j]_{F^*} \underset{\text{lemma}}{=} [\epsilon^j]_{E^*}^\top [\mathbf{f}_i]_E = \hat{\mathfrak{e}}_j^\top [\mathbf{f}_i]_E = ([\mathbf{f}_i]_E)^j = ([\mathbf{F}]_E)^j_i = ([\mathbf{F}]_E^\top)^i_j \underset{\text{third line}}{=} ([\mathbf{E}]_F^{-\top})^i_j.$$

□

Remark 5.37. (What covariance and contravariance *really* mean).

The first two equations of the previous theorem can be restated as

$$[\mathbf{v}]_F = [\mathbf{F}]_E^{-1}[\mathbf{v}]_E$$
$$[\phi]_{F^*}^\top = [\phi]_{E^*}^\top [\mathbf{F}]_E.$$

(We have simply copied the first equation from the previous theorem. The second equation has been obtained by transposing its counterpart from the previous theorem).

Paying close attention to the second above equation, we see that when we treat the coordinates of dual vectors taken relative to the E^* basis as row vectors (i.e. as transposed column vectors), then these row vectors transform over to the F^* basis with use of $[\mathbf{F}]_E$. On the other hand, the first above equation states that the coordinates of vectors relative to E (when treated as column vectors, as usual) transform over to the F basis with use of $[\mathbf{F}]_E^{-1}$. For this reason, dual vectors are considered to be *covariant*, since they "co-vary" *with* $[\mathbf{F}]_E$, and vectors are considered to be *contravariant*, since they "contra-vary" *against* $[\mathbf{F}]_E$.

Theorem 5.38. (Change of basis for vectors and dual vectors in terms of basis vectors and basis dual vectors).

Let V be a finite-dimensional vector space with bases $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ and $F^* = \{\delta^1, ..., \delta^n\}$ be the corresponding induced dual bases for V^*. We have

$$\boxed{\begin{aligned} \mathbf{f}_i &= \sum_{j=1}^n ([\mathbf{f}_i]_E)_j \mathbf{e}_j = \sum_{j=1}^n ([\mathbf{F}]_E)^j_i \mathbf{e}_j \\ \delta^i &= \sum_{j=1}^n ([\delta^i]_{E^*})_j \epsilon^j = \sum_{j=1}^n ([\mathbf{F}]_E^{-\top})^j_i \epsilon^j \end{aligned}}$$

Proof. The first line in the boxed equation follows from the definition of $[\cdot]_F$ (and was stated earlier in Theorem 2.60). The second line in the boxed equation follows by applying the first line to the bases F^* and E^* for V^*. Specifically, the second equation in the second line follows because $\delta^i = \sum_{j=1}^n ([\mathbf{F}^*]_{E^*})^j_i \epsilon^j$, where we have $[\mathbf{F}^*]_{E^*} = [\mathbf{F}]_E^{-\top}$ due to the previous theorem. □

Theorem 5.39. (Change of basis for a $\binom{p}{q}$ 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 $E^* = \{\epsilon^1, ..., \epsilon^n\}$ and $F^* = \{\delta^1, ..., \delta^n\}$ be[2] the corresponding induced dual bases for V^*.

We now derive how to change the coordinates of a $\binom{p}{q}$ tensor in $T^p_q(V)$. To do so, it is enough to relate the coordinates of the elementary $\binom{p}{q}$ tensor

$$T = \mathbf{f}_{i_1} \otimes ... \otimes \mathbf{f}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q}$$

taken relative to F and F^* to the coordinates of T relative to E and E^*.

To obtain this relation, we apply the previous theorem to each basis vector in T:

$$\begin{aligned} T &= \Big(\sum_{j_1=1}^n ([\mathbf{F}]_E)^{j_1}_{i_1} \mathbf{e}_{j_1} \Big) \otimes ... \otimes \Big(\sum_{j_p=1}^n ([\mathbf{F}]_E)^{j_p}_{i_p} \mathbf{e}_{j_p} \Big) \otimes \Big(\sum_{i_1=1}^n ([\mathbf{F}_E]^{-1})^{j_1}_{i_1} \epsilon^{i_1} \Big) \otimes ... \otimes \Big(\sum_{i_q=1}^n ([\mathbf{F}_E]^{-1})^{j_q}_{i_q} \epsilon^{i_q} \Big) \\ &= \sum_{j_1=1}^n \cdots \sum_{j_p}^n \sum_{i_1=1}^n \cdots \sum_{i_q=1}^n \Big(([\mathbf{F}]_E)^{j_1}_{i_1} ... ([\mathbf{F}]_E)^{j_p}_{i_p} ([\mathbf{F}_E]^{-1})^{j_1}_{i_1} ... ([\mathbf{F}_E]^{-1})^{j_q}_{i_q} \mathbf{e}_{j_1} \otimes ... \otimes \mathbf{e}_{j_p} \otimes \epsilon^{i_1} \otimes ... \otimes \epsilon^{i_q} \Big). \end{aligned}$$

Though we have argued using an elementary tensor, this shows that an arbitrary $\binom{p}{q}$ 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 an $\binom{i_1 ... i_p}{j_1 ... j_q}$ component relative to E and E^* of

$$\sum_{k_1=1}^n \cdots \sum_{k_p}^n \sum_{\ell_1=1}^n \cdots \sum_{\ell_q=1}^n \Big(([\mathbf{F}]_E)^{k_1}_{\ell_1} ... ([\mathbf{F}]_E)^{k_p}_{\ell_p} ([\mathbf{F}_E]^{-1})^{k_1}_{\ell_1} ... ([\mathbf{F}_E]^{-1})^{k_q}_{\ell_q} T^{k_1 ... k_p}_{\ell_1 ... \ell_q} \Big).$$

(It is possible to this expression using the fact that $([\mathbf{F}]_E)^i_j ([\mathbf{F}]_E)^{-1})^i_j = \delta^i_j$. Let's not do that, because that would require introducing the max function to account for whether $p \geq q$ or $q < p$).

This change of basis formula is sometimes called the *Ricci transformation law*, or the *tensor transformation law*.

At this stage, it would be remiss not to mention what is called *Einstein summation notation.* In Einstein summation notation, we assume that there is an "implied summation" over any index that appears in both a lower and upper index. We can use Einstein notation to write the $\binom{i_1 ... i_p}{j_1 ... j_q}$ component of the previously mentioned $\binom{p}{q}$ tensor relative to E and E^* as

$$([\mathbf{F}]_E)^{k_1}_{\ell_1} ... ([\mathbf{F}]_E)^{k_p}_{\ell_p} ([\mathbf{F}_E]^{-1})^{k_1}_{\ell_1} ... ([\mathbf{F}_E]^{-1})^{k_q}_{\ell_q} T^{k_1 ... k_p}_{\ell_1 ... \ell_q} \quad \text{(Einstein notation)}.$$

Remark 5.40. (Tensors as "multidimensional matrices" that "transform like tensors").

As was mentioned in Remark 4.32, physicists often define tensors to be "multidimensional matrices" that follow the change of basis formula of the previous theorem.

5.2.2 Tensor contraction

Derivation 5.41. (Composition of linear functions with contraction).

Let V, W and Z be vector spaces over a field K. Notice that the map $\circ$ which composes linear functions $V \to W$ and $W \to Z$ is itself a bilinear map $\mathcal{L}(V \to W) \times \mathcal{L}(W \to Z) \overset{\circ}{\to} \mathcal{L}(V, Z)$. (Check

[2]We have used the notation ϵ^i and δ^i for the induced dual bases only for convenience. As usual, we have $\epsilon^i = \phi^{\mathbf{e}_i}$ and $\delta^i = \psi^{\mathbf{f}_i}$, where $\phi^{\mathbf{e}_i}(\mathbf{v}) = \hat{\mathfrak{e}}_i^\top [\mathbf{v}]_E$ and $\psi^{\mathbf{f}_i}(\mathbf{v}) = \hat{\mathfrak{e}}_i^\top [\mathbf{v}]_F$ as usual.

this as an exercise!). Also recall from Section 4.2 that every element of $\mathcal{L}(V \to W)$ and $\mathcal{L}(W \to Z)$ are linear combinations of rank-1 compositions of linear functions, i.e., of "elementary compositions". We can understand the composition map $\circ$ more deeply by looking at how it acts on such elementary compositions.

Lastly, recall the convention of Section 4.2 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

$$(\mathbf{w} \circ \phi, \mathbf{z} \circ \psi) \overset{\circ}{\longrightarrow} (\mathbf{z} \circ \psi) \circ (\mathbf{w} \circ \phi) = \mathbf{z} \circ (\psi \circ \mathbf{w}) \circ \phi.$$

Notice that $\phi \circ \mathbf{w}$ is the linear map $K \to K$ sending $c \mapsto c\psi(\mathbf{w})$. If we extend the convention of Section 4.2 mentioned above 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 two equalities, we were able to commute $\psi(\mathbf{w})$ because it is a linear map $K \to K$).

The action of $\phi \in W^*$ on $\mathbf{w} \in W$ is said to be the result of evaluating the *natural pairing map on W and W^**, or, equivalently, the result of *contracting W against W^**. Therefore, we see that the composition of linear maps, when we restrict the linear maps to be elementary compositions, involves *contraction*. These notions are formalized in the next definition.

Definition 5.42. (Tensor contraction).

Let V be a vector space, and consider also its dual space V^*. There is a natural bilinear form C on V and V^*, often called the *natural pairing (of V and V^*)*, that is defined by $C(\mathbf{v}, \phi) = \phi(\mathbf{v})$.

Now suppose that V is finite-dimensional. We define the $\binom{k}{\ell}$ *contraction* on elementary $\binom{p}{q}$ tensors, and extend with multilinearity. The $\binom{k}{\ell}$ contraction of an elementary tensor is defined as follows:

$$\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_p \otimes \phi^1 \otimes ... \otimes \phi^q$$
$$\binom{k}{\ell} \underset{\longmapsto}{\text{contraction}}$$
$$C(\mathbf{v}_k, \phi^\ell)(\mathbf{v}_1 \otimes ... \otimes \cancel{\mathbf{v}_k} \otimes ... \mathbf{v}_p \otimes \phi^1 \otimes ... \otimes \cancel{\phi^\ell} \otimes ... \otimes \phi^q)$$
$$=$$
$$\phi^\ell(\mathbf{v}_k)(\mathbf{v}_1 \otimes ... \otimes \cancel{\mathbf{v}_k} \otimes ... \mathbf{v}_p \otimes \phi^1 \otimes ... \otimes \cancel{\phi^\ell} \otimes ... \otimes \phi^q).$$

Remark 5.43. (Contraction with upper and lower indices).

The convention we laid out in 4.33 requires that lower indices (e.g. those which appear in $\mathbf{v}_k$) are always contracted against upper indices (e.g. chose which appear in ϕ^ℓ), and vice versa. Lower indices are never contracted against other lower indices are never contracted against other lower indices, and upper indices are never contracted against other upper indices.

Remark 5.44. (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 4.35, 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 5.41, 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 5.45. (Induced contraction between a $\binom{p}{q}$ tensor and an $\binom{k}{\ell}$ tensor when either $p, \ell \neq 0$ or $q, k \neq 0$).

Example: double contraction (double dot product) is the induced contraction between order 2 tensors. (The valence doesn't matter because the result is a scalar).

Definition 5.46. (Double dot product).

The double dot product : is a map $T_2^2(V) \otimes T_2^2(V)^* \to V^{\otimes 4}$, and is a "double contraction" on $T_2^2(V) \otimes T_2^2(V)^*$; since $T_2^2(V) = V \otimes V \otimes V^* \otimes V^*$, the V^*'s of the first $T_2^2(V)$'s evaluate the V's of the second two $T_2^2(V)$'s. Note that $T_2^2(V)$ is naturally isomorphic to its dual due to the commutativity of tensor products, so : can also be thought of as a "double contraction" on $T_2^2(V) \otimes T_2^2(V) \cong T_4^4(V)$.

Theorem 5.47. (Trace is a special case of tensor contraction).

An element of $\mathcal{L}(V, V)$ with matrix $\mathbf{A} = (a_{ij})$ coorresponds to the $\binom{1}{1}$ tensor $\sum_{ij} a_i^j \epsilon_j \otimes \mathbf{e}^i$, which contracts to $\sum_{ij} a_i^j \epsilon_j(\mathbf{e}^i) = \sum_{ij} a_i^j \delta_i^j = \sum_i a_i^i = \text{tr}(\mathbf{A})$.

Theorem 5.48. (Tensor contraction is basis-independent).

5.2.3 Slanted indices

Theorem 5.49. Correspondence between multidimensional matrices and $\binom{p}{q}$ tensors

Theorem 5.50. (Converting vectors to dual vectors and dual vectors to vectors in a $\binom{p}{q}$ tensor).

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be the induced dual basis for V^*.

Given a $\binom{p}{q}$ tensor $\mathbf{T} \in T_q^p(V)$, we can convert it to a $\binom{p-1}{q+1}$ tensor. First, we send

$$\begin{gathered}\mathbf{T} = \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_k} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \\ \longmapsto \\ \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_k}^{\flat_1} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q}.\end{gathered}$$

Using Theorem [...], we compute $\mathbf{e}_{i_k}^{\flat_1}$ to be

$$\mathbf{e}_{i_k}^{\flat_1} = \sum_{r=1}^n ([\mathbf{e}_{i_k}^{\flat_1}]_{E^*})_r \epsilon^r = \sum_{r=1}^n \Big(\sum_{j=1}^n g_{rj}([\mathbf{e}_{i_k}]_E)^j \Big) \epsilon^r = \sum_{r=1}^n \Big(\sum_{j=1}^n g_{rj} \delta_{i_k}^j \Big) \epsilon^r = \sum_{r=1}^n g_{i_k r} \epsilon^r,$$

so $\mathbf{T}$ is ultimately sent to

$$\begin{aligned}&\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_{k-1}} \otimes \sum_{r=1}^n \Big(g_{i_k r} \epsilon^r \Big) \otimes \mathbf{e}_{i_{k+1}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \\ &= \sum_{r=1}^n g_{i_k r} \Big(\mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_{k-1}} \otimes \epsilon^r \otimes \mathbf{e}_{i_{k+1}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \Big).\end{aligned}$$

Rearranging the position of ϵ^r within the sum, we see that this $\binom{p}{q}$ tensor is identifiable with the tensor

$$\sum_{r=1}^n g_{i_k r} \Big(\mathbf{e}_{i_1} \otimes ... \otimes \cancel{\mathbf{e}_{i_k}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^r \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \Big).$$

Overall, we have converted a $\binom{p}{q}$ tensor to a $\binom{p-1}{q+1}$ tensor as follows:

$$\begin{gathered}\mathbf{T} = \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_k} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \\ \longmapsto \\ \sum_{r=1}^n g_{i_k r} \Big(\mathbf{e}_{i_1} \otimes ... \otimes \cancel{\mathbf{e}_{i_k}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^r \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \Big).\end{gathered}$$

If the coordinates of $\mathbf{T}$ were originally $T^{i_1...i_p}_{j_1...j_q}$, then they get sent to $\sum_{r=1}^{n} g_{i_k r} T^{i_1...\cancel{i_k}...i_p}_{r j_1.......j_q}$.

Following a similar process to above, we can convert a $\binom{p}{q}$ tensor to a $\binom{p+1}{q-1}$ tensor like this:

$$\mathbf{T} = \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_\ell} \otimes ... \otimes \epsilon^{j_q}$$
$$\longmapsto$$
$$\sum_{r=1}^{n} g^{j_\ell r} \Big(\epsilon^r \otimes \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \cancel{\epsilon^{j_\ell}} \otimes ... \otimes \epsilon^{j_q} \Big).$$

(Note that we have used the contravariant metric tensor, with upper indices, here). Here, if the coordinates of $\mathbf{T}$ were originally $T^{i_1...i_p}_{j_1...j_q}$, then they get sent to $\sum_{r=1}^{n} g^{i_k r} T^{r i_1.......i_p}_{j_1...\cancel{j_\ell}...j_q}$.

To make sure the above two mappings are invertible (i.e. reversable), we adopt the convention that if we have two identifications $V \to V^*$, then the result of the second identification (which will be $\mathbf{e}^\flat_{i_u}$ for some positive integer u) goes immediately to the left of the previous identification (which will be $\mathbf{e}_{i_v}$ for some positive integer $v < u$), where the first identification is immediately to the left of the leftmost "original" basis dual vector (which is ϵ^{j_1} in both of the above mappings). Similarly, if we have two identifications $V^* \to V$, then the result of the second identification goes immediately to the left of the previous identification, where the first identification is immediately to the right of the leftmost "original" basis vector (which is $\mathbf{e}_{i_1}$ in both of the above mappings).

Definition 5.51. (Index notation and slanted index notation).

[We now present conventions that we will not use]

Let V be a finite-dimensional vector space, let E be a basis for V, and let E^* be the induced dual basis for V^*. We define $v^i := ([\mathbf{v}]_E)^i$ and $\phi_i := ([\phi]_{E^*})_i$.

Let V and W be finite-dimensional vector spaces with bases E and F, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_m}\}$ be corresponding induced dual bases of V^* and W^*. Suppose that we also have a metric tensor g on V and W, so $V \cong W^*$ naturally via the musical isomorphism $\flat_1 : V \to W^*$ that is induced by g. In this situation, we define $v_i := (\mathbf{v}^{\flat_1})_i$.

Due to Theorem [...], we have

$$v^i = \sum_{j=1}^{n} g^{ij} v_j$$
$$v_i = \sum_{j=1}^{n} g_{ij} v^j.$$

Due to the above two equations, physicists will say that one converts between the coordinates of a vector and the coordinates of its corresponding the dual vector by "multiplying by the metric". (Physicists also use Einstein summation notation, so this operation looks more like straightforward multiplication because, in Einstein summation notation, the sum is "implied", and not written).

As was shown out in the previous theorem, we can also "multiply by the metric" to convert the coordinates of a $\binom{p}{q}$ tensor to the coordinates of a $\binom{p-1}{q+1}$ or $\binom{p+1}{q-1}$ tensor. For this reason, there is also notation for converting coordinates of tensors with use of "multiplying by the metric". We present that notation now.

The conventions established at the end of the previous theorem correspond to the notation of "slanted indices".

Sticking to $\binom{1}{1}$ tensors for now, we define

$$T^i{}_j := \sum_k g^{ik} T_{kj}$$
$$T_j{}^i := \sum_k g^{ik} T_{jk}$$

Staggered index notation for $\binom{p}{q}$ tensors. Raise index i of T_{ij}:. (Remember, the index of T that gets summed over is the one that gets lowered or raised; it is raised, in this instance). The i is written to the left of the j because it was originally left of j in T_{ij}.

For slanted indices to have meaning, we must be given the context of some "original" slanting scheme. So, to keep things simple, it is easiest to use a straightforward and easy-to-read slanting scheme, such as $T^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}$ for an arbitrary $\binom{p}{q}$ tensor. When indices are raised or lowered,

then we can raise/lower indices in whatever order necessary by multiplying by g^{ij} or g_{ij} and summing as necessary.

To write an arbitrary index for a $\binom{p}{q}$ tensor, we *could* come up with notation for arbitrary staggering of indices

Instead, to keep things neater, we use the convention of using

6

Exterior powers, the determinant, and orientation

6.1 Exterior powers

6.1.1 Alternating tensors

Definition 6.1. (Permuting a $\binom{p}{q}$ tensor).

Let V be a vector space over a field K. Given a permutation $\sigma \in S_k$ and a tensor $\mathbf{T} \in T^p_q(V)$, we define the map $(\cdot)^\sigma : \mathbf{T}^p_q(V) \to T^p_q(V)$ 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 6.2. (Alternating $\binom{p}{q}$ tensor).

Let V be a vector space. We say a $\binom{p}{q}$ tensor $\mathbf{T} \in T^p_q(V)$ is *alternating* iff $\mathbf{T}^\sigma = \operatorname{sgn}(\sigma)\mathbf{T}$.

Remark 6.3. (We want to construct an algebra of alternating tensors).

Consider $(A, K, \cdot)$, where A is a set, K is a field, and $\cdot : A \times A \to A$ is a binary function on A. We say that A is an *algebra over* K iff

- A is a vector space over K
- $\cdot : A \times A \to A$ is bilinear

Note that $(\cup_{p,q\in\mathbb{N}}\{\text{alternating } \binom{p}{q} \text{ tensors}\}, K, \otimes)$ is *not* an algebra.

To see why, we show the contrapositive of the statement "$\mathbf{T}$ and $\mathbf{S}$ are alernating implies $\mathbf{T} \otimes \mathbf{S}$ is alternating"; given a tensor $\mathbf{T} \otimes \mathbf{S}$ that is not alternating, we show that it is possible for both $\mathbf{T}$ and $\mathbf{S}$ to be alternating. To this end, consider $\mathbf{T} = \mathbf{v}_1 \otimes \mathbf{v}_2$ and $\mathbf{S} = \mathbf{v}_3 \otimes \mathbf{v}_4$. We assume that $\mathbf{T} \otimes \mathbf{S} = \mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \mathbf{v}_3 \otimes \mathbf{v}_4$ is alternating, so, in particular, we do *not* have $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \mathbf{v}_3 \otimes \mathbf{v}_4 = -\mathbf{v}_3 \otimes \mathbf{v}_4 \otimes \mathbf{v}_1 \otimes \mathbf{v}_2$. (We have obtained the tensor on the right hand side by swapping the vector $\mathbf{v}_i$ from $\mathbf{T}$ with its counterpart $\mathbf{v}_{i+2}$ from $\mathbf{S}$). It is still possible to have $\mathbf{v}_1 \otimes \mathbf{v}_2 = -\mathbf{v}_2 \otimes \mathbf{v}_1$ and $\mathbf{v}_3 \otimes \mathbf{v}_3 = -\mathbf{v}_4 \otimes \mathbf{v}_3$

???

We will construct a function $\wedge : T^p_q(V) \times T^p_q(V) \to T^p_q(V)$, called the *wedge product*, such that $(\cup_{p,q\in\mathbb{N}}\{\text{alternating } \binom{p}{q} \text{ tensors}\}, K, \wedge)$ *is* an algebra.

6.1.2 Constructing the wedge product

Definition 6.4. (Alternatization of a $\binom{p}{q}$ tensor).

Let V be a vector space. Before defining the wedge product, we first define a function alt : $T^p_q(V) \to T^p_q(V)$ which converts any tensor $\mathbf{T} \in T^p_q(V)$ into an alternating tensor $\operatorname{alt}(T)$ by defining alt on elementary tensors and extending linearly: for an elementary tensor $\mathbf{T} \in T_q^p(V)$ of order k, we define $\text{alt}(\mathbf{T}) := \frac{1}{k!} \sum_{\sigma \in S_k} \text{sgn}(\sigma) \mathbf{T}^\sigma$. Note, the division by $k!$ is used so that $\text{alt}(\mathbf{T}) = \mathbf{T}$ when $\mathbf{T}$ is alternating.

How might you come up with this formula? Well, you first might start by noticing the case of $k = 2$, without the division by $2!$. That is, notice that we can send $\mathbf{v} \otimes \mathbf{w}$ to the alternating tensor $\mathbf{v} \otimes \mathbf{w} - \mathbf{w} \otimes \mathbf{v}$.

Proof. We need to show that $\text{alt}(\mathbf{T})$ is alternating, i.e., that $\text{alt}(\mathbf{T})^\pi = \text{sgn}(\pi)\text{alt}(\mathbf{T})$. We have

$$\text{alt}(\mathbf{T})^\pi = \Big(\sum_{\sigma \in S_k} \text{sgn}(\sigma)(\mathbf{T}^\sigma) \Big)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma)(\mathbf{T}^\sigma)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma) \mathbf{T}^{\pi \circ \sigma}.$$

Since S_k is closed under taking inverses, then for every $\tau \in S_k$ there is a $\sigma \in S_k$ such that $\tau = \pi \circ \sigma$; namely, $\sigma = \pi^{-1} \circ \tau$. So the sum becomes

$$\sum_{\tau \in S_k} \text{sgn}(\pi^{-1} \circ \tau) \mathbf{T}^\tau = \text{sgn}(\pi^{-1}) \sum_{\tau \in S_k} \text{sgn}(\tau) \mathbf{T}^\tau = -\text{sgn}(\pi)\text{alt}(\mathbf{T}).$$

We have shown $\text{alt}(\mathbf{T})^\pi = \text{sgn}(\pi)\text{alt}(\mathbf{T})$. □

Definition 6.5. (Wedge product).

Let V be a vector space. We define the *wedge product* $\wedge : T_q^p(V) \times T_q^p(V) \to \text{alt}(T_q^p(V))$ by $\mathbf{T} \wedge \mathbf{S} := \text{alt}(\mathbf{T} \otimes \mathbf{S})$.

Lemma 6.6. (Lemma for associativity of wedge product).

Let V be a vector space, and consider $\mathbf{T}, \mathbf{S} \in T_q^p(V)$. If $\text{alt}(\mathbf{T}) = \mathbf{0}$, then $\mathbf{T} \wedge \mathbf{S} = \mathbf{0} = \mathbf{S} \wedge \mathbf{T}$.

Proof. Assume $\text{alt}(\mathbf{T}) = \mathbf{0}$. Let $T = \mathbf{v}_{i_1} \otimes ... \otimes \mathbf{v}_{i_k}$ and $\mathbf{S} = \mathbf{v}_{j_1} \otimes ... \otimes \mathbf{v}_{j_\ell}$. We must show $\text{alt}(\mathbf{T} \otimes \mathbf{S}) = \mathbf{0}$.

To do so, let be H is the subgroup of $S_{k+\ell}$ whose elements fix all of $j_1, ..., j_\ell$, and consider the right cosets $\{H\sigma \mid \sigma \in S_{k+\ell}\}$ of H in $S_{k+\ell}$. Since these right cosets partition $S_{k+\ell}$, then

$$\begin{aligned}
\text{alt}(\mathbf{T} \otimes \mathbf{S}) &= \sum_{[\pi]_\tau \sigma \in \{\text{right cosets}\}} (-1)^{[\pi]_\tau \sigma} (\mathbf{T} \otimes \mathbf{S})^{\pi_\tau \sigma} = \sum_{\sigma \in S_{k+m}} \sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau \sigma} (\mathbf{T} \otimes \mathbf{S})^{\pi_\tau \sigma} \\
&= \sum_{\sigma \in S_{k+m}} \Big(\sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} (\mathbf{T} \otimes \mathbf{S})^{\pi_\tau} \Big)^\sigma.
\end{aligned}$$

Since $[\pi]_\tau \in H$, where H is the subgroup of $S_{k+\ell}$ 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

$$\sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} (\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} = \sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} \mathbf{T}^{[\pi]_\tau} \otimes \mathbf{S} = \Big(\sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} \mathbf{T}^{[\pi]_\tau} \Big) \otimes \mathbf{S}.$$

Now define $\pi \in S_k$ by $\pi = \tau^{-1} [\pi]_\tau \tau$, where $\tau = (i_1, ..., i_k)$, using one-line notation (so $\tau(i) = j_i$). Then the above is

$$\Big(\sum_{\pi S_k} (-1)^\pi \mathbf{T}^\pi \Big) \otimes \mathbf{S} = \text{alt}(\mathbf{T}) \otimes \mathbf{S} = \mathbf{0} \otimes \mathbf{S} = \mathbf{0}.$$

The last equality follows by the seeming-multilinearity of $\otimes$. □

Theorem 6.7. (Wedge product is associative).

Let V be a vector space. Then for all $\mathbf{T}, \mathbf{S}, \mathbf{R} \in T_q^p(V)$, 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. Let $\mathbf{T}, \mathbf{S}, \mathbf{R}$ be $\binom{p}{q}$ tensors. We will show $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$; a similar argument shows $\mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R}) = \text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$.

First, we have by definition of $\wedge$ that

$$(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \text{alt}((\mathbf{T} \wedge \mathbf{S}) \otimes \mathbf{R})$$

Subtracting $\text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$ from both sides and using linearity of alt, we get that

$$(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} - \text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R}) = \text{alt}((\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \otimes \mathbf{R}) = (\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R}.$$

If we show $(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R} = \mathbf{0}$, then our claim is true. The previous lemma says that if $\text{alt}(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) = \mathbf{0}$, then $(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R} = \mathbf{0}$. And this is true, since $\text{alt}(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) = \text{alt}(\mathbf{T} \wedge \mathbf{S}) - \text{alt}(\mathbf{T} \otimes \mathbf{S}) = \mathbf{T} \wedge \mathbf{S} - \mathbf{T} \wedge \mathbf{S} = \mathbf{0}$, by linearity of alt and with use of the fact that $\mathbf{T} \wedge \mathbf{S}$ is alternating. □

6.1.3 Exterior powers

Theorem 6.8. (Properties of the wedge product).

Let V be vector space over a field K, and consider the set of $\binom{p}{q}$ tensors on V. Then the wedge product $\wedge$ satisfies the following properties...

1. $\wedge$ looks as if it is bilinear, just as was the case with $\otimes$. That is, ...
   1.1. $(\mathbf{T}_1 + \mathbf{T}_2) \wedge \mathbf{T}_3 = \mathbf{T}_1 \wedge \mathbf{T}_3 + \mathbf{T}_2 \wedge \mathbf{T}_3$ for all $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3 \in T^p_q(V)$.
   1.2. $\mathbf{T}_1 \wedge (\mathbf{T}_2 + \mathbf{T}_3) = \mathbf{T}_1 \wedge \mathbf{T}_2 + \mathbf{T}_1 \wedge \mathbf{T}_3$ for all $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3 \in T^p_q(V)$.
   1.3. $(c\mathbf{T}_1) \wedge \mathbf{T}_2 = c(\mathbf{T}_1 \wedge \mathbf{T}_2) = \mathbf{T}_1 \wedge (c\mathbf{T}_2)$ for all $\mathbf{T}_1, \mathbf{T}_2 \in T^p_q(V)$ and $c \in K$.
2. $\wedge$ looks as it it is an alternating map on vectors: $\mathbf{v}_1 \wedge \mathbf{v}_2 = -\mathbf{v}_2 \wedge \mathbf{v}_1$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.
3. $\wedge$ is *skew-commutative*: if $\mathbf{T}_1 \in T^p_q(V)$ and $\mathbf{T}_2 \in T^k_\ell(V)$, then $\mathbf{T}_1 \wedge \mathbf{T}_2 = (-1)^{k+\ell}(\mathbf{T}_2 \wedge \mathbf{T}_1)$
4. $\mathbf{T} \wedge \mathbf{T} = \mathbf{0}$ for all $\mathbf{T} \in T^p_q(V)$ when $K \neq \mathbb{Z}/2\mathbb{Z}$.

Conditions (2) and (3) are logically equivalent, and conditions (2) and (4) are logically equivalent when $K \neq \mathbb{Z}/2\mathbb{Z}$. Again, we need to require $K \neq \mathbb{Z}/2\mathbb{Z}$ here so that $2 \neq 0$, which allows division by 2.

Theorem 6.9. (Exterior powers).

Let V be a vector space over a field K. In Remark 6.3, we said that we wanted to find a function $\wedge : T^p_q(V) \times T^p_q(V) \to T^p_q(V)$ for which $(\cup_{p,q\in\mathbb{N}}\{\text{alternating } \binom{p}{q} \text{ tensors}\}, K, \wedge)$ is an algebra. We have just done so by constructing the wedge product $\wedge$. Notice also that this algebra can be expressed using the alternization map alt that we constructed, as $(\cup_{p,q\in\mathbb{N}}\text{alt}(T^p_q(V)), K, \wedge)$.

There is no standard notation for the algebra $(\cup_{p,q\in\mathbb{N}}\text{alt}(T^p_q(V)), K, \wedge)$. Instead, we define

$$\Lambda^k(V) := \bigcup_{k\in\mathbb{N}} \text{alt}(T^k_0(V)), K, \wedge).$$

The algebra $(\cup_{p,q\in\mathbb{N}}\text{alt}(T^p_q(V)), K, \wedge)$ can then be expressed as

$$\bigcup_{p,q\in\mathbb{N}} \text{alt}(T^p_q(V)), K, \wedge) = \Lambda^p(V) \cup \Lambda^q(V^*)$$

Definition 6.10. (Wedge product space).

Let V and W be vector spaces. We define the *wedge product space* $V \wedge W$ to be the image of the alternization map alt, so $V \wedge W = \text{alt}(V \otimes W)$.

define alt to operate on V and W

Remark 6.11. Intuition of k-blades and volume

Theorem 6.12. (Basis and dimension for exterior powers).
don't prove until have proven thm about basis for tensor product spaces

Definition 6.13. (Alternating 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 a *k-alternating function* iff for all $\sigma \in S_n$, we have $\mathbf{f}(\mathbf{v}_{\sigma(1)}, ..., \mathbf{v}_{\sigma(k)}) = \text{sgn}(\sigma)\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_k)$. Equivalently, $\mathbf{f}$ is k-alternating iff $\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_i, \mathbf{v}_{i+1}, \mathbf{v}_k) = -\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_{i+1}, \mathbf{v}_i, ..., \mathbf{v}_k)$ for all $i \in \{1, ..., k\}$.

When k is clear from the context, k-linear functions are called *alternating functions.*

Definition 6.14. (Vector space of alternating functions).
If $V_1, ..., V_k, W$ are vector spaces over a field K, then we use $(\text{alt}\mathcal{L})(V_1 \times ... \times V_k \to W)$ to denote the vector space over K formed by the set of k-alternating functions $V_1 \times ... \times V_k \to W$ under the operations of function addition and function scaling.

Theorem 6.15. (Universal property for exterior powers).
Let V_1, V_2, W be vector spaces, 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 $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ that uniquely depends on $\mathbf{f}$, where $\mathbf{g} : V_1 \times V_2 \to V_1 \otimes V_2$.

Proof. The proof is similar to the proof of the universal property of tensor product spaces (Theorem 4.9). The only difference is that the maps we define in this proof are extended using alternatingness and bilinearity, rather than just bilinearity. □

Theorem 6.16. (Fundamental natural isomorphisms for exterior powers).
Theorem 4.35 stated that there are natural isomorphisms

$$
\begin{aligned}
\mathcal{L}(V_1 \times ... \times V_k \to W) &\cong \mathcal{L}(V_1 \otimes ... \otimes V_k \to W) \\
\mathcal{L}(V \to W) &\cong W \otimes V^* \\
(V \otimes W)^* &\cong V^* \otimes W^* \\
T^p_q(V) &\cong T^q_p(V^*)
\end{aligned}
$$

Analogously, there are natural isomorphisms

$$
\boxed{
\begin{aligned}
(\text{alt}\mathcal{L})(V_1 \times ... \times V_k \to W) &\cong (\text{alt}\mathcal{L})(V_1 \otimes ... \otimes V_k \to W) \\
(\text{alt}\mathcal{L})(V \to W) &\cong W \wedge V^* \\
(V \wedge W)^* &\cong V^* \wedge W^* \\
\Lambda^k(V)^* &\cong \Lambda^k(V^*)
\end{aligned}
}
$$

Proof. To show the first equation in the box, show that the map sending an alternating bilinear function to its unique linear counterpart defined on wedge product spaces (which is guaranteed to exist by the universal property for exterior powers) is a linear isomorphism. (This is what we did when we proved the coorresponding fact for tensor product spaces; those steps can essentially be repeated for this proof. See Theorem 4.10). 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 4.2, when we derived the natural isomorphism $\mathcal{L}(V \to W) \cong W \otimes V^*$. That is, take an element $\mathbf{f} \in (\text{alt}\mathcal{L})(V \to W)$, decompose it into a linear combination of "alternating elementary compositions", and then send each alternating elementary composition $\mathbf{w} \circ \phi \mapsto \mathbf{w} \wedge \phi$. The formal check that this map is a linear isomorphism is essentially the same as the check described at the end of Section 4.2.

The third line in the box is proved similarly as was in Theorem 4.35; the only difference is that it is necessary to extend with alternatingness and bilinearity rather than just bilinearity. The fourth line follows from the third line.

read `https://math.stackexchange.com/questions/18595/exterior-power-of-dual-space/18628#18628`. even better, ask Owen □

6.1.4 Push-forward and pull-back

Definition 6.17. (Pushforward of vectors, pullback of dual vectors).

Dual transformation $\mathbf{f}^* : W^* \to V^*$ of a linear function $\mathbf{f} : V \to W$ is in some sense a pullback from W^* to V^*

Theorem 6.18. interp as actual fns

Pushforward of vectors, $\otimes^p f : T_0^p(V) \to T_0^p(W)$, and pullback of dual vectors, $\otimes_q f^* : T_q^0(W) \to T_q^0(V)$. Note that the pullback is often denoted as f^*T.

Definition 6.19. Push-forward $\Lambda^k f$ and pullback $\Lambda^k f^*$

Theorem 6.20. Theorem. Let $\mathbf{f} : V \to V$, so $\mathbf{f}^* : V^* \to V^*$. Consider $\Lambda^n(V^*) = \text{alt}(T_k^0(V))$. If $\phi^i = \mathbf{f}^*(\epsilon^i)$, then $\phi^1 \overline{\wedge} ... \overline{\wedge} \phi^n = \det(\mathbf{f}^*)\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n$. Apply both sides to $(\mathbf{e}_1, ..., \mathbf{e}_n)$ to see $(\phi^1 \overline{\wedge} ... \overline{\wedge} \phi^n)(\mathbf{e}_1, ..., \mathbf{e}_n) = \det(\phi^i(\mathbf{e}_j))$. Extend with multilinearity and alternatingness to see $(\phi^1 \overline{\wedge} ... \overline{\wedge} \phi^n)(\mathbf{v}_1, ..., \mathbf{v}_n) = \det(\phi^i(\mathbf{v}_j))$.

Above thm holds for any $k \leq n$. Argue using vector subspaces.

6.2 The determinant

Definition 6.21. (The determinant).

Let K, and let $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ be the standard basis of K^n. We want to define a function $(K^n)^{\times n} \to K$ which, given $\mathbf{c}_1, ..., \mathbf{c}_n \in K^n$, returns the n-dimensional volume of the parallelapiped spanned by $\mathbf{c}_1, ..., \mathbf{c}_n$. We will denote this function by $\det : (K^n)^{\times n} \to K$. We require that det satisfy the following axioms:

1. $\det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$, since we want the unit n-cube to have an n-dimensional volume of 1.
2. det is multilinear, because...
    - The volume of a parallelapiped that is the disjoint union of two smaller parallelapipeds should be the sum of the volumes of the smaller parallelapipeds.
    - Scaling one of the sides of a parallelapiped by $c \in K$ should increase that paralellapiped's volume by a factor of c.
3. $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n) = 0$ when $\mathbf{c}_i = \mathbf{c}_j$ for all $\mathbf{c}_k \in K$, $k \in \{1, ..., n\}$. This should hold because when two sides of a parallelapiped coincide, its *n-dimensional* volume is zero.

When $K \neq \mathbb{Z}/2\mathbb{Z}$ so that $2 \neq 0$, which enables division by 2, then, due to the multilinearity of det, the third axiom is logically equivalent to det being an alternating function. (Proof left as exercise). This is the case when $K = \mathbb{R}$, for example. The fact that det is alternating means that our intuitive assumptions about volume require that volume be *signed*, or *oriented*; the volume of the parallelapiped spanned by $\mathbf{c}_1, ..., \mathbf{c}_j, ..., \mathbf{c}_i, ..., \mathbf{c}_n$ is the negation of the volume of the parallelapiped spanned by $\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n$.

The fact that the third axiom is logically equivalent to alternatingness also gives us a concise characterization of the determinant: $\det : (K^n)^{\times n} \to K$, when $K \neq \mathbb{Z}/2\mathbb{Z}$, is the unique multilinear alternating function satisfying $\det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$.

Definition 6.22. (Determinant of a square matrix).

We define the *determinant of a square matrix* to be the result of applying det to the column vectors of that matrix.

Theorem 6.23. (Consequent properties of the determinant).

5. det is invariant under linearly combining input vectors into a different input vector. That is, $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_n) = \det(\mathbf{c}_1, ..., \mathbf{c}_i + \sum_{j=1, j \neq i}^n d_j \mathbf{c}_j, ..., \mathbf{c}_n)$ for all $i \in \{1, ..., n\}$.

6. $\det(\mathbf{c}_1, ..., \mathbf{c}_n) = 0$ iff $\{\mathbf{c}_1, ..., \mathbf{c}_n\}$ is a linearly dependent set.

Proof.

5. Using the axiom $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n) = 0$ when $\mathbf{c}_i = \mathbf{c}_j$ together with the multilinearity of the determinant, we have

$$\begin{aligned}\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_n) &= \det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_n) + \sum_{j=1, j\neq i}^{n} \Big(d_j \det(\mathbf{c}_1, ..., \mathbf{c}_j, ..., \mathbf{c}_j, ..., \mathbf{c}_n)\Big) \\ &= \det\Big(\mathbf{c}_1, ..., \mathbf{c}_i + \sum_{j=1, j\neq i}^{n} (d_j \mathbf{c}_j), ..., \mathbf{c}_n\Big).\end{aligned}$$

6.
($\det(\mathbf{c}_1, ..., \mathbf{c}_n) = 0 \implies \{\mathbf{c}_1, ..., \mathbf{c}_n\}$ is a linearly dependent set). If the input vectors are linearly dependent, we can use the invariance of det under linearly combining some columns into others (which we just proved) to produce an equal determinant in which two columns are the same. By the third axiom, this determinant is zero.
($\det(\mathbf{c}_1, ..., \mathbf{c}_n) = 0 \impliedby \{\mathbf{c}_1, ..., \mathbf{c}_n\}$ is a linearly dependent set). Suppose for contradiction that the determinant of a set of n linearly independent vectors is zero. These n linearly independent vectors form a basis for K^n, so we have shown that the determinant of a basis set is zero. But then, using multilinearity together with the invariance of det under linearly combining some vectors into a different vector, we can show that $\det(\mathbf{c}_1, ..., \mathbf{c}_n) = 0$ for *all* $\mathbf{c}_1, ..., \mathbf{c}_n \in K^n$. This contradicts the first axiom; we must have $\det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$.

□

Derivation 6.24. (Permutation formula for the determinant).

We now derive *permutation formula* for the det. The formula we obtain shows that the function det specified in Definition [...] exists; since this formula is derived from the axioms of the determinant, it also shows that det is unique.

Consider vectors $\mathbf{c}_1, ..., \mathbf{c}_n \in K^n$, and[1] set $(a^i_j) := ([\mathbf{c}_i]_{\hat{\mathfrak{E}}})^j$. Then...

[1]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 $\binom{1}{1}$ tensor rather those of a $\binom{2}{0}$ or $\binom{0}{2}$ tensor.

$$
\begin{aligned}
\det((a^i_j)) = \det(\mathbf{c}_1, ..., \mathbf{c}_n) &= \det\Big(\sum_{i_1=1}^n a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., \sum_{i_n=1}^n a^n_{i_n}\hat{\mathfrak{e}}_{i_n}\Big) \\
&= \sum_{i_1=1}^n \det\Big(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., \sum_{i_n=1}^n a^n_{i_n}\hat{\mathfrak{e}}_{i_n}\Big) \\
&= \sum_{i_1=1}^n \det(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., \sum_{i_n=1}^n a^n_{i_n}\hat{\mathfrak{e}}_{i_n}\Big) \\
&\vdots \\
&= \sum_{i_1=1}^n ... \sum_{i_n=1}^n \det(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., a^n_{i_n}\hat{\mathfrak{e}}_{i_n}) \\
&= \sum_{i_1=1}^n ... \sum_{i_n=1}^n \det(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., a^n_{i_n}\hat{\mathfrak{e}}_{i_n}), \text{ where } i_1, ..., i_n \text{ are distinct from each other} \\
&= \sum_{\sigma \in S_n} \det(a^1_{\sigma(1)}\hat{\mathfrak{e}}_{\sigma(1)}, ..., a^n_{\sigma(n)}\hat{\mathfrak{e}}_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)} \det(\hat{\mathfrak{e}}_{\sigma(1)}, ..., \hat{\mathfrak{e}}_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)}\mathrm{sgn}(\sigma) \det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) \\
&= \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)}\mathrm{sgn}(\sigma)
\end{aligned}
$$

Therefore, we have

$$
\boxed{\det(\mathbf{c}_1, ..., \mathbf{c}_n) = \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)}\mathrm{sgn}(\sigma)}
$$

In this derivation, we have mostly used the multilinearity of the determinant. Though, the expression labeled with "where $i_1, ..., i_n$ are distinct from each other" results from the previous line due to the third axiom of the determinant, $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n) = 0$ when $\mathbf{c}_i = \mathbf{c}_j$.

There are four major steps in the derivation. The first step is to use the multilinearity of the determinant to turn the determinant of (a^i_j) into a sum of the determinants of matrices that only have one nonzero entry in each column (see the line directly above the line labeled with "where $i_1, ..., i_n$ are distinct from each other"). The second step is to disregard all determinants in this previous sum whose matrix arguments have two or more columns that have their nonzero entries in the same row (i.e. whose matrix arguments are matrices of linearly dependent columns). This leaves us with a sum of determinants of diagonal matrices whose columns have been shuffled (this corresponds to the line labeled with "where $i_1, ..., i_n$ are distinct from each other" and the line directly below it). The third step, which corresponds to the third to last line, is to use multilinearity to pull out all the constants. The fourth step is to use the alternatingness of the determinant so that every determinant argument in the sum is the identity matrix; this results in multiplying each term in the sum by $\mathrm{sgn}(\sigma)$.

Remark 6.25. Mention that $\mathrm{sgn}(\sigma)$ is well-defined.

Theorem 6.26. (Determinant of a matrix is transpose-invariant).

Let $\mathbf{A}$ be a square matrix with entries in K. Recall from the discussion after the statement of the permutation formula for the determinant that the determinant of a matrix is a sum of determinants of diagonal matrices whose columns have been shuffled. Each shuffled diagonal matrix in this sum can be momentarily converted to a diagonal matrix, transposed, and then re-shuffled (so that the columns of the shuffled-transposed-reshuffled matrix are in the order of the columns of the original shuffled diagonal matrix). Reversing the expansion that was accomplished with the multilinearity of det in the derivation of the permutation formula for the determinant, we see that the sum of the determinants of these shuffled-transposed-reshuffled matrices is equal to the determinant of $\mathbf{A}^\top$. Therefore

$$\det(\mathbf{A}) = \det(\mathbf{A}^\top).$$

Theorem 6.27. (Laplace expansion for the determinant).

Consider an $n \times n$ matrix $\mathbf{A} = (a^i_j)$, and let $\mathbf{A}^i_j$ denote the so-called *ij minor matrix* obtained by erasing the ith row and jth column of $\mathbf{A}$. We have

$$\det(\mathbf{A}) = \sum_{i=1}^{n} a^i_j \det(\mathbf{A}^i_j) \text{ for all } i \in \{1, ..., n\}$$

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a^i_j \det(\mathbf{A}^i_j) \text{ for all } j \in \{1, ..., n\}$$

The first equation is called the *Laplace expansion for the determinant along the ith row*, and the second equation is called the *Laplace expansion for the determinant along the jth column*. Note that each equation implies the other because $\det(\mathbf{A}) = \det(\mathbf{A}^\top)$.

Proof. We prove the second equation of the theorem.

Consider all terms in the permutation formula's sum for $\det(\mathbf{A})$ that have the factor a^i_j. Let $\mathbf{B}$ denote the shuffled diagonal matrix that corresponds to one of these terms. We can view $\det(\mathbf{B})$ as $\det(\mathbf{B}) = \pm a^i_j \det(\mathbf{B}^i_j)$, where $\mathbf{B}^i_j$ is the determinant of the matrix obtained by removing the ith column and jth row from $\mathbf{B}$. The $\pm$ sign is a result of the fact that the matrices $\mathbf{B}$ and $\mathbf{B}^i_j$ may have different inversion counts.

The main effort of this proof is to determine the $\pm$ sign and specify how the inversion counts of $\mathbf{B}$ and $\mathbf{B}^i_j$ differ.

As a first step, note that the difference in the inversion count between $\mathbf{B}$ and $\mathbf{B}^i_j$ is the number of inversions that involve a^i_j. Thus, our problem reduces to determining an expression for the number of inversions that involve a^i_j. So, divide the matrix $\mathbf{B}$ into quadrants that are centered on a^i_j. Let k_1, k_2, k_3, k_4 be the number of inversions in the upper left, upper right, lower left, and bottom right corners of $\mathbf{A}$, respectively. The number of inversions involving a^i_j is k_2+k_3. Since we know $k_1+k_2+1 = i$ and $k_1 + k_3 + 1 = j$, we have $k_2 + k_3 = i + j - 2 - 2k_1 = i + j - 2(k_1 + 1)$. (We also know $k_1 + k_2 + k_3 + k_4 = n$, but this is not that helpful). Thus, if σ is the permutation corresponding to $\mathbf{B}$ and π is the permutation corresponding to $\mathbf{B}^i_j$, then $\mathrm{sgn}(\sigma) = \mathrm{sgn}(\pi)(-1)^{i+j-2(k_1+1)} = \mathrm{sgn}(\pi)(-1)^{i+j}$. Thus $\mathrm{sgn}(\sigma) = (-1)^{i+j}\mathrm{sgn}(\pi) \iff \mathrm{sgn}(\pi) = (-1)^{i+j}\mathrm{sgn}(\sigma)$.

So,

$$\begin{aligned}
a^i_j \det(\mathbf{B}^i_j) &= a^i_j \sum_{\pi \in S_n} a_1^{\pi(1)} ... \cancel{a_j^{\pi(i)}} ..., a_n^{\pi(n)} \mathrm{sgn}(\pi) \\
&= a^i_j \sum_{\sigma \in S_n} a_1^{\pi(1)} ... \cancel{a_j^{\pi(i)}} ..., a_n^{\pi(n)} (-1)^{i+j} \mathrm{sgn}(\sigma) \\
&= (-1)^{i+j} a^i_j \det(\mathbf{B})
\end{aligned}$$

Thus $a^i_j \det(\mathbf{B}^i_j) = (-1)^{i+j} a^i_j \det(\mathbf{B}) \iff \det(\mathbf{B}) = (-1)^{i+j} a^i_j \det(\mathbf{B}^i_j)$. Now sum all of the $\mathbf{B}$'s (the diagonal shuffled matrices) to get $\det(\mathbf{A}) = \sum_{j=1}^n a^i_j \det(\mathbf{A}^i_j)$. □

Theorem 6.28. (Determinant of an upper triangular matrix).

Theorem 6.29. (Adjoint and Cramer's rule).

Definition 6.30. (Determinant of a linear function $V \to V$).

Let V be a finite-dimensional vector space, and let E be any basis for V. We define the *determinant of a linear function* $\mathbf{f} : V \to V$ to be the determinant of the matrix of $\mathbf{f}$ relative to E and E, $\det(\mathbf{f}) := \det([\mathbf{f}(E)]_E)$.

what about when $\mathbf{f}$ maps into W? pushforward makes more sense for this

Remark 6.31. We have not yet shown that the determinant of a linear function $V \to V$ does not depend on the basis chosen for V. We will see that this is the case soon.

Theorem 6.32. (The pushforward on the top exterior power is multiplication by the determinant).

Let V be an n-dimensional vector space over a field K, let $\mathbf{f} : V \to V$ 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 $\Lambda^n(\mathbf{f}) : \Lambda^n(V) \to \Lambda^n(V)$ which 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 is a multilinear alternating map. The pushforward $\Lambda^n(\mathbf{f})$ is also a map of 1-dimensional vector spaces, so it must be multiplication by a constant. We will determine what this constant is.

Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be a basis of V, and consider the action of the pushforward on the basis vectors of E,

$$\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n).$$

Because $\Lambda^n(\mathbf{f})$ is multilinear and alternating, the wedge product on the left hand side closely resembles the determinant $\det(\mathbf{f}(\mathbf{e}_1), ..., \mathbf{f}(\mathbf{e}_n)) = \det([\mathbf{f}(E)]_E) = \det(\mathbf{f})$. So, set $(a^i_j) = [\mathbf{f}(E)]_E$, and then use essentially the same argument as was made to derive the permutation formula on the left hand side of the above. We obtain

$$\begin{aligned}
\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n) &= \sum_{\sigma \in S_n} \Big(a^1_{\sigma(1)} ... a^n_{\sigma(n)} \mathrm{sgn}(\sigma) \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n \Big) \\
&= \Big(\sum_{\sigma \in S_n} a^1_{\sigma(1)} ... a^n_{\sigma(n)} \mathrm{sgn}(\sigma) \Big) \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n = \det(\mathbf{f})(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n)
\end{aligned}$$

So, we have the statement on the basis E

$$\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n) = \det(\mathbf{f})(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n).$$

Using the seeming-multilinearity of $\wedge$, we can extend this fact to apply to any list of vectors:

$$\boxed{\mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_n) = \det(\mathbf{f}) \mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n \text{ for all } \mathbf{v}_1, ..., \mathbf{v}_n \in V}$$

In other words,

$$\Lambda^n(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n) = \det(\mathbf{f}) \mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n \text{ for all } \mathbf{v}_1, ..., \mathbf{v}_n \in V.$$

Thus, $\Lambda^n(\mathbf{f})$ is multiplication by $\det(\mathbf{f})$.

Theorem 6.33. (Product rule for determinants).

Theorem 6.34. (Determinant of an inverse function).

$\det(\mathbf{f}^{-1}) = \frac{1}{\det(\mathbf{f})}$. Follows because determinant is a group homomorphism (the product rule).

Theorem 6.35. (Determinant is basis-invariant).

Consider a linear function $\mathbf{f} : V \to V$, and let E and F be bases for E. The matrix $[\mathbf{f}(E)]_E$ of $\mathbf{f}$ relative to E and E is related to the matrix $[\mathbf{f}(F)]_F$ of $\mathbf{f}$ relative to F and F by $[\mathbf{f}(E)]_E = [\mathbf{F}]_F [\mathbf{f}(E)]_E [\mathbf{F}]_F^{-1}$ (see Theorem 2.59).

Theorem 6.36. $\det(\mathbf{f}^*) = \det(\mathbf{f})$. Follows because of two facts. (1) if $\mathbf{A}$ is matrix of $\mathbf{f}$ wrt orthonormal bases $\hat{U}_1$, $\hat{U}_2$, then $\mathbf{A}^\top$ is matrix of $\mathbf{f}^*$ wrt $\hat{U}_2^*$, $\hat{U}_1^*$. (2) $\det(\mathbf{A}^\top) = \det(\mathbf{A})$, which is true because a determinant is a sum of determinants of diagonal matrices $\mathbf{D}$, which have the property $\det(\mathbf{D}^\top) = \det(\mathbf{D})$.

Lemma 6.37. (Pushforward on the dual).

use k, not n

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be the induced dual basis for V^*. For all $\phi^1, ..., \phi^n \in V^*$, we have

$$\phi^1 \wedge ... \wedge \phi^n = \det(\phi^i(\mathbf{e}_j))\epsilon^1 \wedge ... \wedge \epsilon^n.$$

Proof. Let $\mathbf{f} : V \to V$ be the linear function whose dual $\mathbf{f}^* : V^* \to V^*$ sends $\epsilon^i \overset{\mathbf{f}^*}{\mapsto} \phi^i$. Then $\det(\mathbf{f}^*)$ is the determinant of the matrix of $\mathbf{f}^*$ relative to E^* and E^*, which is

$$\Big([\mathbf{f}^*(\epsilon^1)]_{E^*} \quad \dots \quad [\mathbf{f}^*(\epsilon^n)]_{E^*}\Big) = \Big([\phi^1]_{E^*} \quad \dots \quad [\phi^n]_{E^*}\Big).$$

[Take a moment to talk about geometric interpretation of this].

We have $([\phi_i]_{E^*})_j = \phi^i(\mathbf{e}_j)$ by Theorem 5.26. Therefore $\det(\mathbf{f}^*) = \det(\phi^i(\mathbf{e}_j))$, where $(\phi^i(\mathbf{e}_j))$ is the matrix with ij entry $\phi^i(\mathbf{e}_j)$.

(Sidenote: notice $\det(\mathbf{f}) = \det(\mathbf{f}^*) = \det(\phi^i(\mathbf{e}_j))$). □

Lemma 6.38. (Action of dual k-wedge on dual basis).

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be the induced dual basis for V^*. Then

$$(\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k)(\mathbf{e}_1, ..., \mathbf{e}_k) = 1.$$

Proof. Using a similar argument to the one that showed the permutation formula for the determinant on the ϵ^i, we have

$$\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k = \overline{\mathrm{alt}}(\epsilon^1 \overline{\otimes} ... \overline{\otimes} \epsilon^k) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \epsilon^{\sigma(1)} \overline{\otimes} ... \overline{\otimes} \epsilon^{\sigma(k)}.$$

Here, $\overline{\mathrm{alt}}$ denotes the map [...] induced by alt.

Therefore

$$(\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k)(\mathbf{e}_1, ..., \mathbf{e}_k) = \sum_{\sigma \in S_n} \Big(\mathrm{sgn}(\sigma) \epsilon^{\sigma(1)} \overline{\otimes} ... \overline{\otimes} \epsilon^{\sigma(k)}\Big)(\mathbf{e}_1, ..., \mathbf{e}_k) = \sum_{\sigma \in S_n} \Big(\mathrm{sgn}(\sigma)(\epsilon^{\sigma(1)} \overline{\otimes} ... \overline{\otimes} \epsilon^{\sigma(k)})(\mathbf{e}_1, ..., \mathbf{e}_k)\Big).$$

Now we focus on the inner term, $(\epsilon^{\sigma(1)} \overline{\otimes} ... \overline{\otimes} \epsilon^{\sigma(k)})(\mathbf{e}_1, ..., \mathbf{e}_k)$. By definition of $\overline{\otimes}$, we have

$$(\epsilon^{\sigma(1)} \overline{\otimes} ... \overline{\otimes} \epsilon^{\sigma(k)})(\mathbf{e}_1, ..., \mathbf{e}_k) = \epsilon^{\sigma(1)}(\mathbf{e}_1) ... \epsilon^{\sigma(k)}(\mathbf{e}_k)$$

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

$$(\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k)(\mathbf{e}_1, ..., \mathbf{e}_k) = \mathrm{sgn}(i) \cdot 1 = 1 \cdot 1 = 1.$$

□

Theorem 6.39. (Action of dual k-wedge on vectors).

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be the induced dual basis for V^*. Then

$$(\phi^1 \overline{\wedge} ... \overline{\wedge} \phi^k)(\mathbf{e}_1, ..., \mathbf{e}_k) = \det(\phi^i(\mathbf{e}_j)).$$

Proof. By Lemma [...], we have $\phi^1 \overline{\wedge} ... \overline{\wedge} \phi^k = \det(\phi^i(\mathbf{e}_j))(\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k)(\mathbf{e}_1, ..., \mathbf{e}_k)$. By the previous lemma, $(\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k)(\mathbf{e}_1, ..., \mathbf{e}_k) = 1$, so the result follows. □

6.3 Orientation of finite-dimensional vector spaces

Orientation is the mathematical formalization of the notions of "clockwise" and "counterclockwise"; it is the notion which distinguishes different "rotational configurations" from each other.

Our discussion of orientation will be as follows. First, we treat inner product spaces, so that we can speak of orthonormality; we will define an orientation on an inner product space to be a choice of an *ordered* orthonormal basis. Then, in order to complete the definition of orientation for inner product spaces, we introduce rotations in n-dimensions. After we finish the definition of orientation for inner product spaces, we end the subsection on oriented inner product spaces by presenting the fact that the determinant "tracks" orientation. This fact allows us to generalize the notion of orientation to finite dimensional vector spaces that may or may not have an inner product. Lastly, we show how the top exterior power of a finite-dimensional vector space can be used for the purposes of orientation.

6.3.1 First notions of orientation for inner product spaces

Definition 6.40. (Ordered basis).

We will formalize the notion of orientation relying on the concept of an ordered basis. An *ordered basis*, which is a basis for a finite-dimensional vector space in which the order that vectors are specified matters.

For example, if V is a 2-dimensional vector space, then the ordered bases $E_1 = \{\mathbf{e}_1, \mathbf{e}_2\}$ and $E_2 = \{\mathbf{e}_2, \mathbf{e}_1\}$ for V are not equal, $E_1 \neq E_2$.

Definition 6.41. (Permutation acting on an ordered basis).

Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be an ordered basis for some finite-dimensional vector space. Given a permutation $\sigma \in S_n$, we define $E^\sigma := \{\mathbf{e}_{\sigma(1)}, ..., \mathbf{e}_{\sigma(n)}\}$.

We now discover a consequence of imposing that the bases under consideration be ordered.

Derivation 6.42. (Intuition for the alternatingness 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 *alternatingness of ordered bases.* The notion of "equivalent under rotation" is formalized in the following definition.

We now work to formalize the alternatingness of ordered bases by stating it in the context of a 2-dimensional inner product space. The notion of *rotational equivalence* is what facilitate this formalization. Before we define rotational equivalence, however, we must define what a 2-rotation is.

Definition 6.43. (2-rotation).

Let V be a 2-dimensional inner product space, and let $\hat{U}$ be an orthonormal ordered basis for V. A 2-*rotation on the* 2-*dimensional inner product space* V is a linear function $\mathbf{R}_\theta : V \to V$ whose matrix relative to $\hat{U}$ and $\hat{U}$ is

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

for some $\theta \in [0, 2\pi)$.

Now consider the case when V is an n-dimensional inner product space, $n > 2$. Suppose $\hat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ is an orthonormal ordered basis for V. An *extension of a 2-rotation on the oriented subspace*[2] *span*$(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ *to* V is a linear function $V \to V$ for which

- The map $(\mathbf{v}_i, \mathbf{v}_j) \mapsto \mathbf{R}_\theta(\mathbf{v}_1, ..., \mathbf{v}_i, ..., \mathbf{v}_j, ..., \mathbf{v}_n)$ is a 2-rotation on $\text{span}(\mathbf{v}_i, \mathbf{v}_j)$ for some i, j
- The map $(\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \cancel{\mathbf{v}_j}, ..., \mathbf{v}_n) \mapsto \mathbf{R}_\theta(\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \cancel{\mathbf{v}_j}, ..., \mathbf{v}_n)$ is the identity on $\text{span}(\mathbf{v}_1, ..., \cancel{\mathbf{v}_i}, ..., \cancel{\mathbf{v}_j}, ..., \mathbf{v}_n)$.

Note that the extension of a 2-rotation restricts to a 2-rotation on the subspace $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$. It's also worth noting that the matrix of such an extension relative to $\hat{U}$ and $\hat{U}$ is, for some $i, j \in \{1, ..., n\}$,

$$\begin{array}{c} \begin{array}{cccccc} & & i\text{th column} & & j\text{th column} & \end{array} \\ \begin{pmatrix} 1 & \dots & \cos(\theta) & 0 & -\sin(\theta) & 0 \\ 0 & & 0 & \vdots & 0 & \vdots \\ 0 & & \vdots & 1 & \vdots & \vdots \\ \vdots & & 0 & \vdots & 0 & \vdots \\ 0 & \dots & \sin(\theta) & 0 & \cos(\theta) & 1 \end{pmatrix} \end{array}.$$

(The columns other than the ith and jth columns are the columns of the $n \times n$ identity matrix).

If $\mathbf{R}_\theta$ is a 2-rotation on a 2-dimensional inner product space or is an extension of a 2-dimensional rotation on an n-dimensional inner product space, it is simply called a 2-*dimensional rotation*. In this looser terminology, the phrase "a 2-rotation defined on the oriented subspace $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$" really means "an extension of a 2-rotation, defined on the oriented subspace $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$".

Definition 6.44. (Equivalence under rotation for 2-dimensional inner product spaces).

We define orthonormal ordered bases $\hat{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ and $\widetilde{U} = \{\widetilde{\mathbf{u}}_1, \widetilde{\mathbf{u}}_2\}$ of a 2-dimensional inner product space V to be *equivalent under rotation* iff there exists a 2-dimensional rotation $\mathbf{R}_\theta$ for which $\widetilde{U} = \mathbf{R}_\theta(\hat{U})$. (Recall Definition 2.42 for the meaning of $\mathbf{R}_\theta \hat{U}$)).

Theorem 6.45. (Alternatingness of ordered bases for a 2-dimensional inner product space).

Now we see how the notion of rotational equivalence for 2-dimensional inner product spaces formalizes the alternatingness of ordered bases. Let $\hat{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ be an orthonormal ordered basis of a 2-dimensional inner product space V. When $\theta = -\frac{\pi}{2}$ or $\theta = \frac{\pi}{2}$, the matrix of $\mathbf{R}_\theta$ relative to $\hat{U}$ and $\hat{U}$ is

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

Thus, as we noticed in the informal discussion (Derivation [...]), the following ordered bases are rotationally equivalent:

$$\{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\} \sim \{-\hat{\mathbf{u}}_2, \hat{\mathbf{u}}_1\} \sim \{\hat{\mathbf{u}}_2, -\hat{\mathbf{u}}_1\}$$
$$\{\hat{\mathbf{u}}_2, \hat{\mathbf{u}}_1\} \sim \{-\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\} \sim \{\hat{\mathbf{u}}_1, -\hat{\mathbf{u}}_2\}.$$

We now generalize the equivalence under rotation and the "swap negate principle" to n-dimensional inner product spaces.

Definition 6.46. (Equivalence under rotation for n-dimensional inner product spaces).

We define orthonormal ordered bases $\hat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ and $\widetilde{U} = \{\widetilde{\mathbf{u}}_1, ..., \widetilde{\mathbf{u}}_n\}$ of an n-dimensional inner product space V to be *equivalent under rotation*, and thus write $\hat{U} \sim \widetilde{U}$, iff there is a composition of 2-dimensional rotations $\mathbf{R} = \mathbf{R}_k \circ ... \circ \mathbf{R}_1$ defined on some oriented subspaces $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ for which $\mathbf{R}(\hat{U}) = \widetilde{U}$. This notion of "equivalence under rotation" is indeed an equivalence relation on orthonormal ordered bases of V.

[2]When we say "oriented subspace", we mean that the orientation of $\text{span}(\mathbf{u}_i, \mathbf{u}_j)$ is given by the ordered basis $\{\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j\}$.

Theorem 6.47. (Alternatingness of ordered bases for an n-dimensional inner product space).

Similarly to what was done in the previous derivation, we use $\theta = -\frac{\pi}{2}$ and $\theta = \frac{\pi}{2}$ in the matrix relative to bases of a 2-dimensional rotation defined on the oriented subspace $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ to obtain the formal statement of the alternatingness of ordered bases: for any orthonormal ordered basis $\hat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ of V, we have

$$\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}, ..., \hat{\mathbf{u}}_n\} \sim -\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_{i+1}, \hat{\mathbf{u}}_i, ..., \hat{\mathbf{u}}_n\}.$$

Equivalently, for any orthonormal ordered basis $\hat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ of V and any permutation $\sigma \in S_n$,

$$\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}, ..., \hat{\mathbf{u}}_n\} \sim \text{sgn}(\sigma)\{\hat{\mathbf{u}}_{\sigma(1)}, ..., \hat{\mathbf{u}}_{\sigma(n)}\}.$$

Definition 6.48. (Orientation of permuted ordered bases).

Let V be an n-dimensional inner product space, and fix an orthonormal ordered basis $\hat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ for V. Suppose $\hat{U}^\sigma = \{\hat{\mathbf{u}}_{\sigma(1)}, ..., \hat{\mathbf{u}}_{\sigma(n)}\}$, where $\sigma \in S_n$, is a permutation of $\hat{U}$ that is not rotationally equivalent to $\hat{U}$ (so choose any σ with $\text{sgn}(\sigma) = -1$). By the alternatingness of ordered bases, any other permutation $\hat{U}^\pi = \{\hat{\mathbf{u}}_{\pi(1)}, ..., \hat{\mathbf{u}}_{\pi(n)}\}$, $\pi \in S_n$, of $\hat{U}$ is rotationally equivalent either to $\hat{U}$ or to $\hat{U}^\sigma$. In other words, there are only two equivalence classes[3] of "equivalence under rotation".

We can now begin to set up the notion of orientation. An *orientation for the n-dimensional inner product space* V is a choice of an orthonormal ordered basis $\hat{U}$ for V. Iff V is given the orientation $\hat{U}$, then the *orientation of a permutation of $\hat{U}$ (relative to $\hat{U}$)* is then defined to be *positive* iff that permutation of $\hat{U}$ is rotationally equivalent to $\hat{U}$, and is defined to be *negative* otherwise. Per the previous paragraph, every permutation of $\hat{U}$ is either positively oriented or negatively oriented relative to $\hat{U}$.

Remark 6.49. (The formalization of "counterclockwise" and "clockwise").

At the beginning of this section, we said that orientation would formalize the notions of "clockwise" and "counterclockwise".

At this point, we need some definitions and facts about n-rotations before we complete our development of orientation.

6.3.2 Rotations in n-dimensions

Definition 6.50. (n-rotation).

Let V be an n-dimensional inner product space, and consider $k \leq n$. An *extension of a k-rotation on the oriented subspace span*$(\mathbf{v}_{i_1}, ..., \mathbf{v}_{i_k})$ *to* V is a linear function $\mathbf{R} : V \to V$ such that

- The map $(\mathbf{v}_{i_1}, ..., \mathbf{v}_{i_k}) \mapsto \mathbf{R}(\mathbf{v}_1, ..., \mathbf{v}_{i_1}, ..., \mathbf{v}_{i_k}, ..., \mathbf{v}_n)$ is an n-rotation on $\text{span}(\mathbf{v}_{i_1}, ..., \mathbf{v}_{i_k})$ for some $i_1, ..., i_k$.
- The map $(\mathbf{v}_1, ..., \cancel{\mathbf{v}_{i_1}}, ..., \cancel{\mathbf{v}_{i_k}}, ..., \mathbf{v}_n) \mapsto \mathbf{R}(\mathbf{v}_1, ..., \cancel{\mathbf{v}_{i_1}}, ..., \cancel{\mathbf{v}_{i_k}}, ..., \mathbf{v}_n)$ is the identity on $\text{span}(\mathbf{v}_1, ..., \cancel{\mathbf{v}_{i_1}}, ..., \cancel{\mathbf{v}_{i_k}}, ..., \mathbf{v}_n)$.

We define a *n-rotation* to be a composition of extensions of $(n-1)$-rotations.

Theorem 6.51. Every n-rotation has determinant 1. [...]. Therefore, every n-rotation is an orthogonal linear function.

Proof. Every n-rotation is a composition of 2-rotations. Since 2-rotations have determinant 1, it follows that any n-rotation has determinant 1. $\square$

[3] 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.

Lemma 6.52. (3-rotations acting on 2-dimensional orthonormal oriented subspaces).

Let V be a inner product space with $\dim(V) \geq 2$, and consider 3-dimensional subspaces $\widetilde{W}$ and W of V. If $\widetilde{U} = \{\widetilde{\mathbf{u}}_1, \widetilde{\mathbf{u}}_2\}$ and $\hat{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ are orthonormal ordered bases of $\widetilde{W}$ and W, respectively, then there exists a 3-rotation which takes $\widetilde{U}$ to $\hat{U}^\sigma$ for some $\sigma \in S_2$.

Proof. By definition, a 3-rotation is of the form $\mathbf{R} = \mathbf{R}_\gamma \circ \mathbf{R}_\beta \circ \mathbf{R}_\alpha$, where $\mathbf{R}_\alpha, \mathbf{R}_\beta$, and $\mathbf{R}_\gamma$ are 2-rotations. (Sidenote: $\alpha, \beta, \gamma \in [0, 2\pi)$ are called *Euler angles*).

We must show that there exist α, β, γ for which $\mathbf{R}(\widetilde{U}) = \hat{U}$. We choose α, β, γ such that $\mathbf{R}(\widetilde{\mathbf{u}}_1) = \hat{\mathbf{u}}_1$. **proof that such α, β, γ exist?** 2-rotations are orthogonal linear functions, so they preserve the orthonormality of bases. Thus, $\mathbf{R}(\widetilde{\mathbf{u}}_2)$ must have length 1 and be orthogonal to $\mathbf{R}(\widetilde{\mathbf{u}}_1) = \hat{\mathbf{u}}_1$, so $\mathbf{R}(\widetilde{\mathbf{u}}_2)$ is either $\hat{\mathbf{u}}_2$ or $-\hat{\mathbf{u}}_2$. Therefore, $\mathbf{R}(\widetilde{U}) = \{\hat{\mathbf{u}}_1, \pm\hat{\mathbf{u}}_2\}$. More formally, $\mathbf{R}(\widetilde{U}) = \hat{U}^\sigma$ for some $\sigma \in S_2$. □

Remark 6.53. The contribution of the previous lemma to the theorem we prove next is that the previous lemma captures the notion of rotating a lesser dimensional subspace within a higher dimensional ambient vector space. This machinery is required in the next theorem in the case when n is odd.

Theorem 6.54. (n-rotations acting on orthonormal ordered bases).

Let V be an n-dimensional vector space, let $k \leq n$, and consider k-dimensional subspaces $\widetilde{W}$ and W of V. If $\widetilde{U} = \{\widetilde{\mathbf{u}}_1, ..., \widetilde{\mathbf{u}}_n\}$ and $\hat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ are orthonormal ordered bases of $\widetilde{W}$ and W, respectively, then there is an n-rotation taking $\widetilde{U}$ to some permutation $\hat{U}^\sigma$ of $\hat{U}$.

Proof. When $n \in \{2, 3\}$, the previous lemma yields the desired n-rotation as a composition of 3-rotations (for $n = 2$, just take the restriction of the 3-rotations). We need to show that the theorem holds when $n > 3$. We consider the cases when n is even and n is odd.

If n is even, then, for $i \in \{1, ..., n\}$, let $\mathbf{R}_i$ be the 3-rotation taking $\{\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}\}$ to $\{\widetilde{\mathbf{u}}_{\sigma_i(i)}, \widetilde{\mathbf{u}}_{\sigma_i(i+1)}\}$, where $\sigma_i \in S(\{i, i+1\})$ is some permutation. Then $\mathbf{R}_{\frac{n}{2}} \circ ... \circ \mathbf{R}_1$ is a composition of 3-rotations taking $\widetilde{U}$ to $\hat{U}^\sigma$, where $\sigma \in S_n$ is the permutation defined by $\sigma(i) = \sigma_{j(i)}(i)$, where $j(i) = i$ when i is odd and $j(i) = i - 1$ when i is even.

If n is odd, then let $\mathbf{R}$ be the 3-rotation taking $\{\hat{\mathbf{u}}_{n-2}, \hat{\mathbf{u}}_{n-1}, \hat{\mathbf{u}}_n\}$ to $\{\widetilde{\mathbf{u}}_{\sigma_{n-2}(n-2)}, \widetilde{\mathbf{u}}_{\sigma_{n-2}(n-1)}, \widetilde{\mathbf{u}}_{\sigma_{n-2}(n)}\}$, where $\sigma_{n-2} \in S(\{n-2, n-1, n\})$ is some permutation.

By the previous case (when n was even), there is a composition $\mathbf{R}_{\frac{n-3}{2}} \circ ... \circ \mathbf{R}_1$ of 3-rotations taking $\widetilde{U} - \{\widetilde{\mathbf{u}}_{n-2}, \widetilde{\mathbf{u}}_{n-1}, \widetilde{\mathbf{u}}_n\}$ to $\hat{U} - \{\hat{\mathbf{u}}_{n-2}, \hat{\mathbf{u}}_{n-1}, \hat{\mathbf{u}}_n\})^\sigma$, for some permutation $\sigma \in S_n$. Thus $\mathbf{R} \circ \mathbf{R}_{\frac{n-3}{2}} \circ ... \circ \mathbf{R}_1$ is a composition of 3-rotations taking $\widetilde{U}$ to $\hat{U}^\pi$, where $\pi(i) = \sigma(i)$ when $i \in \{1, ..., n-3\}$ and $\pi(i) = \sigma_{n-2}(i)$ when $i \in \{n-2, n-1, n\}$. □

6.3.3 Completing the definition of orientation for inner product spaces

Definition 6.55. (Orientation of arbitrary ordered bases).

Let V be an n-dimensional inner product space, and fix an orthonormal basis $\hat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ for V. We know how to "orient" ordered bases for V that happen to be permutations of $\hat{U}$. Now, we generalize the notion of orientation so that it applies to any ordered orthonormal basis of V.

We define the *orientation of an orthonormal ordered basis E of V* that is not necessarily a permutation of $\hat{U}$ to be the orientation of the unique permutation $\hat{U}^\sigma$, $\sigma \in S^n$, of $\hat{U}$ for which there exists an n-rotation taking E to $\hat{U}^\sigma$.

Then, we define the *orientation of an arbitrary orthonormal ordered basis E of V* to be the orientation of the unique orthonormal basis $\hat{U}_E$ obtained from performing the Gram-Schmidt process on E (see Theorem [...]).

Theorem 6.56. (The determinant tracks orientation).

Let V be an n-dimensional inner product space with an orientation given by an orthonormal ordered basis $\hat{U}$. Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be any ordered basis (not necessarily orthonormal) of V. We have $\det([\mathbf{E}]\hat{U}) > 0$ iff E is positively oriented relative to $\hat{U}$, and $\det([\mathbf{E}]\hat{U}) < 0$ iff E is negatively oriented relative to $\hat{U}$.

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 $\hat{U}_E$ from E. Then we pass the definition of orientation for orthonormal ordered bases of V that are not permutations of $\hat{U}$ to the definition of orientation for orthonormal ordered bases of V that are permutations of $\hat{U}$.

To begin the first step, consider $\det([\mathbf{E}]_{\hat{U}}) = \det([\mathbf{e}_1]_{\hat{U}}, ..., [\mathbf{e}_n]_{\hat{U}})$, and perform Gram-Schmidt on $\{[\mathbf{e}_1]_{\hat{U}}, ..., [\mathbf{e}_n]_{\hat{U}}\}$. In the ith step of Gram-Schmidt, a linear combination of the vectors $[\mathbf{e}_1]_{\hat{U}}, ..., \cancel{[\mathbf{e}_i]_{\hat{U}}}, ..., [\mathbf{e}_n]_{\hat{U}}$ is added to $[\mathbf{e}_i]_{\hat{U}}$. Recall from Theorem 6.23 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 $\hat{U}_E = \{\tilde{\mathbf{u}}_1, ..., \tilde{\mathbf{u}}_n\}$ is the orthonormal basis obtained by performing Gram-Schmidt on E, then

$$\det([\mathbf{E}]_{\hat{U}}) = \det([\mathbf{e}_1]_{\hat{U}}, ..., [\mathbf{e}_n]_{\hat{U}}) = \det([\tilde{\mathbf{u}}_1]_{\hat{U}}, ..., [\tilde{\mathbf{u}}_n]_{\hat{U}}) = \det([\hat{U}_E]_{\hat{U}}).$$

In performing this first step of the proof, the determinant has stayed the same as we've passed from E to $\hat{U}_E$. We now show that the determinant continues to stay the same as we pass from $\hat{U}_E$ to some permutation $\hat{U}^\sigma$ of $\hat{U}$.

Theorem 6.54 says that there is a n-rotation $\mathbf{R}$ taking $\hat{U}_E$ to $\hat{U}^\sigma$, for some $\sigma \in S_n$, and Theorem 6.51 guarantees that $\det(\mathbf{R}) = 1$. Thus, since $\hat{U}^\sigma = \mathbf{R}(\hat{U}_E)$, we have

$$\det([\hat{U}_E]_{\hat{U}}) = \det([\mathbf{R}(\hat{U}_E)]_{\hat{U}}) \det([\hat{U}_E]_{\hat{U}}) = \det([(\mathbf{R} \circ \mathbf{I})(\hat{U}_E)]_U) = \det([\mathbf{R}(\hat{U}_E)]_U) = \det([\hat{U}^\sigma]_U).$$

To conclude the proof, we will show that $\det([\hat{U}^\sigma]_{\hat{U}}) = \operatorname{sgn}(\sigma) \det(\hat{U}]_{\hat{U}}) = \operatorname{sgn}(\sigma)$. Since any permutation is a composition of "swaps" (a "swap" is a permutation defined on a two-element set), then $\hat{U}^\sigma$ can be obtained from $\hat{U}$ by repeatedly swapping vectors in $\hat{U}$. Whenever vectors are swapped in the determinant, the sign of the determinant is multiplied by -1. This accounts for the $\operatorname{sgn}(\sigma)$ factor in the equation $\det([\hat{U}^\sigma]_{\hat{U}}) = \operatorname{sgn}(\sigma) \det(\hat{U}]_{\hat{U}}) = \operatorname{sgn}(\sigma)$. □

We state the next theorem only for completeness. (We will not use the forward implication of the next theorem, and our last use of the reverse implication, which was already shown, was in the proof of Theorem 6.56).

Theorem 6.57. A linear function has determinant 1 if and only if it is an n-rotation.

Proof.
$(\Longleftarrow)$. We repeat the proof of Theorem 6.51. Every n-rotation is a composition of 2-rotations. Since 2-rotations have determinant 1, it follows that any n-rotation has determinant 1.
$(\Longrightarrow)$. We will not make use of the forward implication in later proofs, so we present a proof sketch of $(\Longrightarrow)$. If $\mathbf{f}$ is a linear function with $\det(\mathbf{f}) = 1$, then $\mathbf{f}$ must be orthogonal [why?], so, by the Cartan-Dieudonné theorem (see p. [...] of [...]), $\mathbf{f}$ is a composition of n-reflections. (n-reflections are defined similarly to n-rotations. Start with the 2-dimensional case, and define an n-reflection to be a composition of an *odd* number of extensions of $(n-1)$-dimensional reflections. The "odd" stiuplation is necessary because a composition of an even number of $(n-1)$-rotation can be shown to be an n-rotation). $\mathbf{f}$ can't decompose into an odd number of reflections because an n-reflection reverses orientation; if it decomposed into an odd number of reflections, then $\mathbf{f}$ would reverse orientation, which would contradict $\det(\mathbf{f}) = 1$. (Prove that an n-reflection reverses orientation by decomposing an n-reflection into an odd number of 2-reflections). Therefore, $\mathbf{f}$ is a composition of an even number of n-reflections, so it is an n-rotation. □

6.3.4 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 6.58. (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 6.59. (Alternatingness of ordered bases).

Notice that we still have the previous alternatingness of ordered bases due to the alternatingness of the determinant.

We now show how the top exterior power of a vector space can be used to describe orientation.

Theorem 6.60. (Orientation with n-wedges).

Let V be a finite-dimensional vector space with an orientation $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. All positively oriented bases of V are scalar multiples of E, and all negatively oriented bases of V are scalar multiples of $-E$, where $-E = E^\sigma$ for some σ with $\mathrm{sgn}(\sigma) < 0$.

Notice, we can identify $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ with $\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n \in \Lambda^n(V)$, because the alternatingness of ordered bases is manifested in elements of $\Lambda^n(V)$ due to the seeming-alternatingness 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 bases, when considered under the operations of "basis addition" and multiplication by a scalar, is a vector space that is isomorphic to $\Lambda^n(V)$.

Therefore, another way to give an orientation to a finite-dimensional vector space is to choose an element of $\Lambda^n(V)$. Notice also that the fact that the pushforward on the top exterior power is multiplication by the determinant (recall Theorem 6.32) plays nicely into this interpretation: once an orientation $\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n$ of V has been chosen, then we have $\mathbf{f}_1 \wedge ... \wedge \mathbf{f}_n = \det(\mathbf{f})\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n$, where $\mathbf{f}$ is the linear function $V \to V$ sending $\mathbf{e}_i \mapsto \mathbf{f}_i$.

6.3.5 Volume forms and hodge duality?

Might put this section here, it might also come up after diff forms introduced. Have to figure it out more anyways

- Volume form μ
  - Define volume form as n-form μ for which $\mu(\mathbf{X}_1, ..., \mathbf{X}_n) = \det(\mathbf{X}_1, ..., \mathbf{X}_n)$, where the det on the RHS is the function on $\mathfrak{X}(M)^{\times n} \times M$ sending $(\mathbf{X}_1, ..., \mathbf{X}_n, \mathbf{x}) \mapsto \det(\mathbf{X}_1(\mathbf{x}), ..., \mathbf{X}_n(\mathbf{x}))$. Use "8.15. Proposition" of Sjmaar.
  - Sjmaar p. 108 examples about how μ relates to ds, dA, dV
  - Use Lee Prop. 15.29 and 15.31 give formulas about volume forms
- Hodge duality.
  - Let V be a finite-dimensional vector space and consider an ordered basis $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ for V. Let $\sigma_W \in S_k$ and $\sigma_{W^\perp} \in S_{n-k}$. Set W to be the oriented subspace of V spanned by the ordered basis $\{\mathbf{e}_{\sigma(1)}, ..., \mathbf{e}_{\sigma(k)}\}$, and give $W^\perp \subseteq V$ the orientation specified by $\{\mathbf{e}_{\sigma(k+1)}, ..., \mathbf{e}_{\sigma(n)}\}$.

    Equivalently, the orientations chosen on W and $W^\perp$ can be represented as elements of $\Lambda^k(W) \subseteq \Lambda^k(V)$ and $\Lambda^{n-k}(W^\perp) \subseteq \Lambda^{n-k}(V)$, respectively. The orientation on W is specified by $\mathbf{e}_{\sigma(1)} \wedge ... \wedge \mathbf{e}_{\sigma(k)}$ and the orientation of $W^\perp$ is specified by $\mathbf{e}_{\sigma(k+1)} \wedge ... \wedge \mathbf{e}_{\sigma(n)}$.

    The *Hodge dual*, or *Hodge star*, is the map $* : \Lambda^k(W) \to \Lambda^{n-k}(W^\perp)$ which acts on the orientation chosen for a subspace $W \subseteq V$ to the orientation chosen for $W^\perp$.

It is enough to say that $*$ "acts" on $\mathbf{e}_{\sigma(1)} \wedge ... \wedge \mathbf{e}_{\sigma(k)}$ as the action on this element of $\Lambda^k(W)$ is implied: $*$ is a map of one-dimensional vector spaces, so we know

$$*(\mathbf{e}_{\sigma(1)} \wedge ... \wedge \mathbf{e}_{\sigma(k)}) =$$

6.4 The cross product

This section gives an intuitive and formal understanding of the cross product.

Lemma 6.61. (The algebraic dot product on $\mathbb{R}^n$ is positive definite, and therefore cancelable).

The algebraic dot product on $\mathbb{R}^n$ is *positive definite*; that is, it satisfies the property $(\mathbf{v} \cdot \mathbf{v} = 0 \iff \mathbf{v} = \mathbf{0})$.

As a consequence, we have the fact that, when $\mathbf{v} \in \mathbb{R}^n$ is nonzero and $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, then $((\mathbf{v}_1 \cdot \mathbf{v} = \mathbf{v}_2 \cdot \mathbf{v}) \implies \mathbf{v}_1 = \mathbf{v}_2)$.

Proof. We first show that the dot product on $\mathbb{R}^n$ satisfies the property $(\mathbf{v} \cdot \mathbf{v} = 0 \iff \mathbf{v} = \mathbf{0})$. The reverse implication follows immediately. For the forward implication, let $\mathbf{v} \in \mathbb{R}^n$, and suppose $\mathbf{v} \cdot \mathbf{v} = 0$; we must show $\mathbf{v} = \mathbf{0}$. We have $\mathbf{v} \cdot \mathbf{v} = \sum_{i=1}^n ([\mathbf{v}]_{\hat{\mathfrak{E}}})_i^2$. Each term in the sum is a nonnegative number. Therefore, the sum is only zero if all terms in the sum are zero, so $([\mathbf{v}]_{\hat{\mathfrak{E}}})_i = 0$ for each i, which means $\mathbf{v} = \mathbf{0}$.

Now we show that when $\mathbf{v} \in \mathbb{R}^n$ is nonzero and $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, then $((\mathbf{v}_1 \cdot \mathbf{v} = \mathbf{v}_2 \cdot \mathbf{v}) \implies \mathbf{v}_1 = \mathbf{v}_2)$. If $\mathbf{v}_1 \cdot \mathbf{v} = \mathbf{v}_2 \cdot \mathbf{v}$, then $(\mathbf{v}_1 - \mathbf{v}_2) \cdot \mathbf{v} = \mathbf{0}$ by the bilinearity of $\cdot$. Since $\mathbf{v} \neq \mathbf{0}$, then we must have $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$ due to the positive definiteness of $\cdot$. That is, $\mathbf{v}_1 = \mathbf{v}_2$. $\square$

Derivation 6.62. Prove that $(\mathbf{v}_1 \overline{\wedge} ... \overline{\wedge} \mathbf{v}_{n-1}) \cdot \mathbf{x} = \det(\mathbf{v}_1, ..., \mathbf{v}_{n-1}, \mathbf{x})$.

Then derive induced operation $\times$ on $V \times V$ coorresponding to $\overline{\wedge}$ when $\dim(V) = 3$

Definition 6.63. (Cross product). The *cross product* of vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ is the unique vector $(\mathbf{v}_1 \times \mathbf{v}_2) \in \mathbb{R}^3$ for which

$$(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{x} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^3.$$

Proof. Let $\mathbf{x} \in \mathbb{R}^3$. We need to show that the vector $\mathbf{c} \in \mathbb{R}^3$ for which $\mathbf{c} \cdot \mathbf{x} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{c})$ is unique. So, let $\mathbf{d} \in R^3$ satisfy this relation as well; suppose $\mathbf{d} \cdot \mathbf{x} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{c})$. Then $\mathbf{c} \cdot \mathbf{x} = \mathbf{d} \cdot \mathbf{x}$. Since $\cdot$ is cancelable (see the previous lemma), then $\mathbf{c} = \mathbf{d}$. $\square$

Theorem 6.64. (Magnitude, direction of cross product).

right hand rule

sin(theta) formula

7

Different definitions of tensor

7.0.1 Exterior powers

Theorem 7.1. Translate the above into a statement about $\Lambda^n(V)^* \cong \text{alt}(\mathcal{L}(V^{\times n} \to K))$. See Lee.

From here, probably switch over to using actual function interpretation of exterior powers.

Pushforward and pullback

this has been postponed until after exterior powers because there are two interps of pushforward and pullback

8

Tensors in physics and engineering

- In physics/engineering, we are usually in the situation in which a basis has been chosen for V. (Since V is usually $\mathbb{R}^n$, for which we usually choose the standard basis). So there is always an isomorphism $V \cong V^*$. Two ways to see this (that are really the same): (1) choice of basis allows for sending basis vectors to dual vectors in the way you'd expect, (2) choice of basis induces the bilinear form that is the dot product.
- Einstein notation
  - Levi-Civita symbol. (It is not a tensor!)
- Example: Cauchy stress tensor
- Summarize things that have been mentioned before
  - Outer product. It is denoted by $\mathbf{v}_1 \otimes \mathbf{v}_2$.
  - Double contraction
    - Example: Hooke's law
  - Metric tensor and slanted indices
- Abstract index notation? (`https://math.stackexchange.com/questions/455478/what-is-the-practic` `rq=1`)

Part II

Differential forms

9

Review of calculus

Notation for covariance and contravariance is not used in this chapter. The use of both upper and lower indices to distinguish between "covariant" and "contravariant" will not be used in the following chapter of multivariable calculus review, even though these concepts have already been introduced. Only lower indices will be used.

9.1 Single-variable calculus

Definition 9.1. (The derivative).

$$f'(t) := \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}$$

Definition 9.2. (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.

FIX ALIGNMENT

Leibniz notation

$$\frac{df}{dt} := f'$$
$$\frac{df}{dt}\Big|_{t=t_0} = f'(t_0)$$

Operator notation

$$\frac{d}{dt} f := f'$$
$$\frac{d}{dt} f(t) := \Big(\frac{d}{ds} f\Big)\Big|_{s=t} = f'(t)$$
$$\frac{df(t)}{dt} := \frac{d}{dt} f(t)$$

Definition 9.3. (Derivative with respect to a function).

This definition formalizes a convention that is often used but rarely explained.

Suppose $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ and $x : V \subseteq \mathbb{R} \to \mathbb{R}$ satisfy the differentiability conditions of the chain rule (see Theorem [...]), so that $f \circ x$ is differentiable. We define $\frac{df}{dx} :? \to ?$ to be the function defined by

$$\frac{df}{dx}\Big|_{t=t_0} := \frac{df}{dt}\Big|_{t=x(t_0)}.$$

That is, $\frac{df}{dx} := f' \circ x$.

With this notation, the chain rule is

$$\frac{d(f \circ x)}{dt} = \frac{df}{dx}\frac{dx}{dt}.$$

This is more elegant than the following statement of the chain rule employing a substitution, which is often presented in standard calculus textbooks:

$$\frac{d(f \circ x)}{dt}\Big|_{t=t_0} = \frac{df(u)}{du}\frac{du(t)}{dt}, \text{ where } u = x(t_0).$$

check the above standard chain rule statement

Remark 9.4. (Letters in the denominator).

The definition $\frac{df}{dt} := f'$ from Definition 9.2 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.

Definition 9.5. (Antiderivative). Let $f : I \to \mathbb{R}$ be a function defined on an interval I. An *antiderivative of f (on I)* is a function $F : I \to \mathbb{R}$ whose derivative is f, i.e., $F' = f$.

We denote a particular antiderivative of f by $\int f$. If f has an antiderivative $\int f$, then all other antiderivatives of f are of the form $\int f + c$, where $c \in \mathbb{R}$ is a constant.

Note, we have not shown yet that any function actually has an antiderivative. A corollary of the first part of the fundamental theorem of calculus is that every continuous function has an antiderivative.

Definition 9.6. (Difference notation). If $f : A \to \mathbb{R} \to \mathbb{R}$ is function defined on any nonempty set $A \subseteq \mathbb{R}$, then we define, for $a, b \in A$, the notation $f|_a^b := f(b) - f(a)$.

Theorem 9.7. (Fundamental theorem of calculus).

1. If $f : [a, b] \to \mathbb{R}$ is continuous, then the "accumulation function" $\int_a^x f$ is differentiable on $[a, b]$ and

$$\frac{d}{dx}\int_a^x f = f.$$

2. If $f : [a, b] \to \mathbb{R}$ is continuous, then

$$\int_a^b f = \Big(\int f\Big)\Big|_a^b.$$

Proof. See `http://www2.clarku.edu/~djoyce/ma121/FTCproof.pdf`. □

Remark 9.8. (Intuition for the fundamental theorem of calculus).

The first part of the theorem says that the rate of the accumulation f is f itself.

The second part of the fundamental theorem of calculus follows from the first part. It states that accumulating f over the region $[a, b]$ is "the same" as finding the difference in the antiderivative between b and a. More geometrically, if we define $F = \int f$, then it relates the definite integral of F' over the region $[a, b]$ to how F behaves on the boundary $\{\{a\}, \{b\}\}$. Stokes' theorem (Theorem [...]) generalizes this result.

(The second part of the fundamental theorem as stated above also holds if f is assumed to be Riemann-integrable. The focus of this book is geometry, not analysis, so we are not concerned with this).

Theorem 9.9. All continuous functions $[a, b] \to \mathbb{R}$ have antiderivatives.

This is a corollary of the first part of the fundamental theorem of calculus: an antiderivative of a continuous function f is its accumulation function.

9.2 Multivariable calculus

Lemma 9.10. (Multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}$).

Let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$.

Set $\mathbf{x}(t_0) = \mathbf{x}_0$.

It can be proved that

$$\frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = \frac{\partial f(\mathbf{x})}{\partial x_1}\Big|_{\mathbf{x}=\mathbf{x}_0} \frac{dx_1}{dt}\Big|_{t=t_0} + ... + \frac{\partial f(\mathbf{x})}{\partial x_n}\Big|_{\mathbf{x}=\mathbf{x}_0} \frac{dx_n}{dt}\Big|_{t=t_0}.$$

In other words,

$$\frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = (\underset{\mathbf{x}}{\nabla} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}(t)}{dt}\Big|_{t=t_0},$$

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 9.3. This is why we say the gradient $\nabla_{\mathbf{x}} f$ is "with respect to $\mathbf{x}$".

We can interpret the dot product to act on vector-valued functions (the dot product of vector-valued functions is equal to the dot product of the evaluated vector-valued functions at each point) so that the above is expressed as

$$\boxed{\frac{d(f \circ \mathbf{x})}{dt} = (\underset{\mathbf{x}}{\nabla} f) \cdot \frac{d\mathbf{x}(t)}{dt}}$$

Definition 9.11. (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 $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\frac{d\mathbf{x}}{dt}\Big|_{t=t_0} = \mathbf{v}$. We define the *directional derivative* $\frac{\partial f}{\partial \mathbf{v}}$ *of f in the direction of* $\mathbf{v}$ to be

$$\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{x}=\mathbf{x}_0} := \frac{d(f\circ \mathbf{x})}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = (\underset{\mathbf{x}}{\nabla} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}(t)}{dt}\Big|_{t=t_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \mathbf{v}$$

Therefore the directional derivative is expressed as

$$\boxed{\begin{aligned} &\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{x}=\mathbf{x}_0} = (\nabla_{\mathbf{x}} f)_{\mathbf{x}=\mathbf{x}_0} \cdot \mathbf{v} \\ &\frac{\partial f}{\partial \mathbf{v}} = \nabla f \cdot \mathbf{v} \end{aligned}}$$

In the second line, we interpret ∇ as the function sending $\mathbf{x} \mapsto \nabla_{\mathbf{x}}$.
Most authors denote $\frac{\partial f}{\partial \mathbf{v}}$ as $D_{\mathbf{x}} f(\mathbf{v})$ or as $Df[\mathbf{v}](\mathbf{x})$.

Theorem 9.12. (Gradient is direction of greatest increase).

Remark 9.13. (Directional derivative simplifies to partial derivative). We have $\frac{\partial}{\partial \hat{\mathfrak{e}}_i} = \frac{\partial}{\partial x_i}$.

Lemma 9.14. (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 $\mathbf{x}_0 = \mathbf{x}(t_0)$.

$$\frac{d(\mathbf{f}\circ\mathbf{x})(t)}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = \begin{pmatrix} \frac{d}{dt} f_1(\mathbf{x}(t)) \\ \vdots \\ \frac{d}{dt} f_m(\mathbf{x}(t)) \end{pmatrix} = \begin{pmatrix} (\nabla_{\mathbf{x}} f_1)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}}{dt}\Big|_{t=t_0} \\ \vdots \\ (\nabla_{\mathbf{x}} f_m)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}}{dt}\Big|_{t=t_0} \end{pmatrix} = \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix}\Bigg|_{\mathbf{x}=\mathbf{x}_0} \frac{d\mathbf{x}}{dt}\Big|_{t=t_0}.$$

In terms of functions, we have

$$\frac{d(\mathbf{f}\circ\mathbf{x})(t)}{dt} = \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix} \frac{d\mathbf{x}}{dt}$$

Recall from Derivation 2.44 and Theorem 2.54 that a matrix-vector product can be expressed as either a linear combination of column vectors or as a vector of dot products. We have already seen the second expression; here is the first:

$$\begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix} \frac{d\mathbf{x}}{dt} = \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \cdot \frac{d\mathbf{x}}{dt} \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \cdot \frac{d\mathbf{x}}{dt} \end{pmatrix} = \sum_{i=1}^{n} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_i} \frac{dx_i(t)}{dt}.$$

Definition 9.15. (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 9.3), we define the *Jacobian matrix* $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ to be

$$\boxed{\frac{\partial \mathbf{f}}{\partial \mathbf{x}} := \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix} = \Big(\frac{\partial f_i}{\partial x_j}\Big)}$$

Using the Jacobian, the multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}^m$ is now succinctly stated as

$$\boxed{\frac{d(\mathbf{f} \circ \mathbf{x})}{dt} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt}}$$

Definition 9.16. (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 $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\frac{d\mathbf{x}}{dt}\Big|_{t=t_0} = \mathbf{v}$. We define the *directional derivative* $\frac{\partial f}{\partial \mathbf{v}}$ *of* f *in the direction of* $\mathbf{v}$ to be

$$\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\Big|_{\mathbf{x}=\mathbf{x}_0} := \frac{d(\mathbf{f} \circ \mathbf{x})}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}_0} \frac{d\mathbf{x}}{dt}\Big|_{t=t_0}$$

So this most general definition of directional derivative is expressed as

$$\boxed{\begin{array}{l} \dfrac{\partial \mathbf{f}}{\partial \mathbf{v}}\Big|_{\mathbf{x}=\mathbf{x}_0} = \dfrac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}_0} \dfrac{d\mathbf{x}}{dt}\Big|_{t=t_0} \\ \dfrac{\partial \mathbf{f}}{\partial \mathbf{v}} = \dfrac{\partial \mathbf{f}}{\partial \mathbf{x}} \dfrac{d\mathbf{x}}{dt} \end{array}}$$

Remark 9.17. Linearity with respect to the $\mathbf{v}$ in the denominator.

10

Basic topology

Topology in $\mathbb{R}_{\text{std}}$

Open sets, closed sets and their characterizations

Define an *open set in* $\mathbb{R}_{\text{std}}$ to be an arbitrary union of open intervals. From this, the *interior point characterization of open sets* follows: a set $U \subseteq \mathbb{R}_{\text{std}}$ is open iff each $x \in U$ is an interior point, where we have defined $x \in A \subseteq \mathbb{R}_{\text{std}}$ to be an *interior point* of A iff $\forall x \in A\ \exists \text{open } U_x \ni x$ s.t. $U_x \subseteq A$. We define the *interior* of A to be the set of all interior points of A, $\text{int}(A) := \{x \in A \mid x \text{ is an interior point of } A\}$, so that the interior point characterization of open sets can be stated as: $U \subseteq \mathbb{R}_{\text{std}}$ is open iff $U = \text{int}(U)$. (Every set A satisfies $\text{int}(A) \subseteq A$. So $A = \text{int}(A)$ iff $A \subseteq \text{int}(A)$).

Next, see what the interior point characterization of open sets implies for complements of open sets.

$$
\begin{aligned}
&U \text{ is open} \\
U = \text{int}(U) \iff U \subseteq \text{int}(U) \iff &\forall x \in U\ x \in \text{int}(U) \\
&\forall x \in U\ \exists \text{open } U_x \ni x \text{ s.t. } U_x \subseteq U \\
&\forall x \in U\ \exists \text{open } U_x \ni x \text{ s.t. } U_x \cap U^c = \emptyset \\
&\forall x\ x \notin U^c \implies \sim (\forall \text{open } U_x \ni x\ \ U_x \cap U^c \neq \emptyset) \\
&\forall x\ x \notin U^c \implies \sim (\forall \text{open } U_x\ \ U_x \cap U^c - \{x\} \neq \emptyset) \\
&\forall x\ x \in U^c \implies \sim (x \in (U^c)') \\
&\forall x\ x \in (U^c)' \implies x \in U^c \\
&(U^c)' \subseteq U^c.
\end{aligned}
$$

Note that line 6 follows from line 5 because we can subtract x out of $U_x \cap U^c$ due to the hypothesis "$x \notin U^c$".

In line 7, we define the notion of a limit point. We say $x \in \mathbb{R}_{\text{std}}$ is a *limit point* of A iff $\forall \text{open } U_x\ \ U_x \cap U^c - \{x\} \neq \emptyset$. That is, $x \in \mathbb{R}_{\text{std}}$ is a *limit point* of A iff every open set containing x intersects A. The set of limit points of A is denoted A'.

Still looking at the above proof, we can notice that we accidentally proved $x \in \text{int}(U)$ iff $x \notin (U^c)'$. So, we proved $x \in \text{int}(U)$ iff $x \in ((U^c)')^c$, which means $\text{int}(U) = ((U^c)')^c \iff \text{int}(U)^c = (U^c)'$. The one to remember is $\text{int}(U)^c = (U^c)'$.

What is most important from the above is that we have seen 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* $\mathbb{R}_{\text{std}}$ to be any set which is the complement of an open set, or, equivalently, any set which contains all of its limit points. The fact that C is closed iff $C' \subseteq C$ is the *limit point characterization of closed sets.*

Unions and intersections of open and closed sets

It quickly follows from the definition of an open set as an arbitrary union of open intervals that an arbitrary union of open sets is an open set. DeMorgan's laws then imply that an arbitrary intersection of closed sets is a closed set.

What about intersections of open sets- or, equivalently, by DeMorgan's laws- unions of closed sets? In $\mathbb{R}_{\text{std}}$, it is apparent that any infinite sort of union of closed sets is not necessarily closed: consider $\cup_{i=1}^{\infty}\{\frac{1}{n}\} = \{\frac{1}{n} \mid n \in \mathbb{N}\}$, which does not consider its limit point of 0. Perhaps at finite unions of closed sets must be closed? This turns out to be true in $\mathbb{R}_{\text{std}}$.

When defining the general notion of a topological space, we *require* that finite unions of closed sets be closed. We will require general topological spaces to have the same fundamental property of $\mathbb{R}_{\text{std}}$ that causes finite unions of closed sets to be closed. (This fundamental property will be that "a basis refines with interior points").

Topologies

A *topology* τ on a set X is a collection of subsets of X such that...

1. There is a *cover* $\mathcal{B}$ of X that *generates* τ.
   - A collection $\mathbb{C}$ of subsets of X is a *cover* of a set X iff $\cup_{C \in \mathbb{C}} C = X$. A set $\mathbb{C}$ *generates* a topology τ iff each $U \in \tau$ is an arbitrary union of the elements of $\mathbb{C}$; each $U \in \tau$ is $U = \cup_{\alpha \in I} C_\alpha$, $\{C_\alpha\} \subseteq \mathbb{C}$. (In the example with $\mathbb{R}_{\text{std}}$, elements of $\mathbb{C}$ were open intervals).
   - For the same reasons as with $\mathbb{R}_{\text{std}}$ ($X = \mathbb{R}$ and $\tau = \{\text{open sets} \subseteq \mathbb{R}\}$), (1) is equivalent to the interior point characterization of open sets, which is in turn equivalent to the limit point characterization of closed sets. Interior points and limit points are defined in the same way as before. This is also equivalent to the fact: arbitrary unions of open sets are open $\iff$ arbitrary intersections of closed sets are closed.
2. Finite unions of closed sets are closed $\iff$ finite intersections of open sets are open.

We interpret the elements of τ as being open sets. Formally, we say that $U \subseteq X$ is an *open set* iff $U \in \tau$.

The above definition of a topology can quickly be seen to be equivalent to the most common definition. The most common definition requires τ to satisfy the following:

3. Arbitrary unions of open sets are open.
   - As noted in (1), we have that (1) and (3) are equivalent.
4. Finite intersections of closed sets are open.

Note that the most common definition makes no mention of any sort of basis.

From covers to bases

In order to derive the definition of a *basis* for a topology, we will show: if $\mathcal{B}$ is a cover that generates τ, then finite unions of closed sets are closed iff $\mathcal{B}$ "refines with interior points". We say $\mathcal{B}$ *refines with interior points* iff $\forall B_1, B_2 \in \mathcal{B} \;\; x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B}$ s.t. $x \in B_3$. This motivates defining a *basis* of a topology τ on a set X to be a cover of X that refines with interior points: since finite unions of closed sets are closed iff $\mathcal{B}$ refines with interior points, then *every topology is generated by some basis.* In fact, the reason that finite unions of closed sets are closed in $\mathbb{R}_{\text{std}}$ is that $\mathbb{R}_{\text{std}}$ refines with interior points.

Now we show that if there $\mathcal{B}$ is the cover that generates τ, then finite unions of closed sets are closed iff $\forall B_1, B_2 \in \mathcal{B} \;\; x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B}$ s.t. $x \in B_3$.

($\implies$). Assume a finite union of closed sets is closed. Then by DeMorgan's laws, any finite intersection of open sets is open. We must show that $\mathcal{B}$ refines with interior points. Basis elements are by definition open, so if $B_1, B_2 \in \mathcal{B}$, then $B_1 \cap B_2$ is a finite intersection of open sets, and is thus open. By the interior point characterization of open sets, $x \in B_1 \cap B_2$ implies there is some open set $U \ni x$ such that $U \subseteq B_1 \cap B_2$. Because U is open, then it is a union of basis elements, so $x \in B_3$ for some $B_3 \in \mathcal{B}$.

($\impliedby$). Assume $\mathcal{B}$ is a basis; that is, $\mathcal{B}$ is a cover that refines with interior points. We want to show that finite unions of closed sets are closed. By DeMorgan's laws, we can instead show that finite intersections of open sets are open.

So, set $V = \cap_{i=1}^{n} U_i$, where the U_i are open. If any U_i is empty, then their intersection is $\emptyset$, which is open, so assume no U_i is empty. We show that V is open by showing it satisfies the interior point characterization of open sets. Consider $x \in V$. Then $x \in U_i$ for all i. Each U_i is a union of basis elements, so, for each U_i, we have $x \in B_i$ for some $B_i \in \mathcal{B}$. Thus $x \in \cap_{i=1}^{n} B_i$. Using induction on the fact that $\mathcal{B}$ refines with interior points, there is a $B_x \in \mathcal{B}$ s.t. $x \in B_x \subseteq \cap_{i=1}^{n} B_i$. We have $\cap_{i=1}^{n} B_i = \cup_{x \in V} B_x$, so $\cap_{i=1}^{n} B_i$ is open.

Summary

We start by defining the elements of a *topology* τ on a set X to be generated by a cover $\mathcal{B}$ of X. (In the motivating example of $\mathbb{R}_{\text{std}}$, the elements of $\mathcal{B}$ are open intervals). In noticing the *interior point characterization of open sets*, we discover the definition of an *interior point*. The interior point characterization of open sets states that U is open iff $U = \text{int}(U)$.

Next, we rephrase the characterization of open sets in terms of complements of open sets, and discover the definitions *limit point*, *closed set*, and the theorem that is the *limit point characterization of closed sets*. The limit point characterization of closed sets states that C is closed iff $C' \subseteq C$, where C' is the set of limit points of C. A closer look at the proof of this reveals that the proof hinged upon the fact that $\text{int}(U)^c = (U^c)'$.

Lastly, we require any topology τ on a set X to be such that finite unions of closed sets are closed. (We are motivated to do so because $\mathbb{R}_{\text{std}}$ demonstrates this property). We show that if $\mathcal{B}$ is a cover of X that generates τ, then finite unions of closed sets are closed iff $\mathcal{B}$ refines with interior points. Therefore, we have the three following equivalent definitions of a topology:

τ is a *topology* on X iff any one of these equivalent statements is true:

- There is a cover $\mathcal{B}$ of X that generates τ, and finite unions of closed sets are closed.
- Arbitrary unions of open sets are open and finite intersections of open sets are open. (Open sets are elements of τ).
- There is a basis for τ.
  - $\mathcal{B}$ is a *basis* for τ iff $\mathcal{B}$ is a cover of X that refines with interior points; that is, iff $\mathcal{B}$ is a cover of X for which $\forall B_1, B_2 \in \mathcal{B} \;\; x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B}$ s.t. $x \in B_3$.

The definitions of and relations between open and closed sets all follow from the first condition in each definition of a topology (which is: there is a cover $\mathcal{B}$ of $X \iff$ arbitrary unions of open sets are open). The necessary and sufficient condition for finite intersections of closed sets (or finite unions of open sets) can be derived with only knowledge of open sets and the interior point characterization of open sets (though, it seems best to frame the proof of this necessary and sufficient condition as the last piece of the puzzle).

Interior and closure

Define interior as set of all interior points
Eqivalent def: largest open set contained in A
Then define closure as smallest closed set which contains A
Compare $\text{cl}(A)$ and A'; maybe investigate A' vs. $\text{cl}(A) - A'$; $A' - \text{cl}(A) \subseteq A' \cap A$?

$$
\begin{aligned}
x \in \text{cl}(A) &\iff \forall \text{closed } C \supseteq A \; x \in C && (10.1)\\
&\iff \forall \text{open } U \; U \cap A = \emptyset \implies x \notin U && (10.2)\\
&\iff \forall \text{open } U \ni x \; U \cap A \neq \emptyset. && (10.3)
\end{aligned}
$$

In line (2), use $U = C^C$. Line (3) follows from (2) because $(P \text{ and } \sim Q) \iff (Q \implies P)$.
More

- mention that open sets are also called “neighborhoods”
- define what it means for a point to “have a neighborhood”
- homeomorphisms are both open and closed maps
- product topology
- projection maps
- continuous functions
- closed and bounded is compact in $\mathbb{R}^n$

11

Manifolds

11.1 Introduction to manifolds

Definition 11.1. [Lee, p.2] (Manifold). A *n-manifold* is a topological space M that is...

- Hausdorff, or "point-separable"
- *second-countable*; that is, M has a countable basis
- *locally Euclidean of dimension n* in the sense that each point in M has a neighborhood that is homeomorphic to $\mathbb{R}^n$ (where $\mathbb{R}^n$ has the standard topology)

Theorem 11.2. [Lee, p.3] (Topological invariance of dimension).

Example 11.3. [Lee, p.5] (Graphs of continuous functions are manifolds).
Example 1.3

Definition 11.4. [Lee, p.25] (Closed n-dimensional upper half-space). Consider $\mathbb{R}^n$ with the standard topology. We define the *closed n-dimensional half space* to be the topology $\mathbb{H}^n \subseteq \mathbb{R}^n$,

$$\mathbb{H}^n := \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^n \geq 0\},$$

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 interior points of M and boundary points of M. To see how $\mathbb{H}^n$ facilitates this, note that the interior $\text{int}(\mathbb{H}^n)$ and boundary $\partial\mathbb{H}^n$ of $\mathbb{H}^n$, in the usual topological senses of "interior" and "boundary" (see Defn [...]), are

$$\text{int}(\mathbb{H}^n) = \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^n > 0\}$$
$$\partial\mathbb{H}^n = \{(x^1, ..., 0) \in \mathbb{R}^n\}.$$

Recall [from topology] $\mathbf{p} \in \mathbb{H}^n$ 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 an n-dimensional Hausdorff second-countable space, and suppose that for some open subset $U \subseteq M$ containing a point $\mathbf{p} \in U$, there is a homeomorphism $\mathbf{x} : U \subseteq M \to \mathbb{R}^n$, where $\mathbf{x}(\mathbf{p}) = (x^1(\mathbf{p}), ..., x^{n-1}(\mathbf{p}), 0)$ lies in $\partial\mathbb{H}^n$. The

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 11.5. [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
- 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 11.6. [Lee, p.26] (Topological interior and boundary vs. manifold interior and boundary).

We can obtain the *topological* interior and *topological* boundary of M by regarding M as a topological space and taking $\text{int}(M), \partial M$ in the usual topological senses of interior and boundary (see Definition [...] and Definition [...]).

In general, the topological interior and topological boundary are *not* the same as the manifold interior and manifold boundary.

To see this, we first need to remember that the topological notions of interior and boundary are only applicable when M is a subset of some other topological space X. We consider the case in which $X = M$, which is a relatively "natural" (but also trivial) choice, and the case $X = \mathbb{R}^m$ for some $m \geq n$, which is less natural, but still a good example.

When $X = M$, we have $\text{int}(M) = M$ and $\partial M =$

When $X = \mathbb{R}^m$, then because $M \subseteq X$, we have $M = \mathbb{R}^n$ where $n \leq m$. When $n = m$, then the topological and manifold notions of interior and boundary are equivalent, but when $n < m$, the topological interior is always empty.

"So for Hn as a subspace of Rn, there it is true that the topological boundary is the manifold boundary. And that's the model for all manifolds with boundary. But because of global topology, it is often not possible for those local Hns to live in a global ambient Rn, and without that we no longer have topological boundary = manifold boundary. (Consider a cylinder IxS1, it's not a subspace of R2)"

Definition 11.7. [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.

Definition 11.8. [Lee, p.415] (Manifold with corners).

$$\text{cl}(\mathbb{H}^n) = \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^i \geq 0 \text{ for all } i\}$$

$\mathbf{p} \in \text{cl}(\mathbb{H}^n)$ is a *corner point of cl($\mathbb{H}^n$)* iff more than one of its coordinates [more precise] vanish.

define manifold with corners

11.2 Coordinatizing manifolds

Definition 11.9. [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 map 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 11.10. [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 11.11. (Coordinate representations of functions on manifolds).

Let M be an n-manifold. The *coordinate representation of a function* $\mathbf{f} : M \to \mathbb{R}^k$ *relative to a chart* $(U, \mathbf{x})$ is the function $\widetilde{\mathbf{f}}_{(U,\mathbf{x})} = \mathbf{f} \circ \mathbf{x}^{-1} : \mathbb{R}^n \to \mathbb{R}^k$.

Now let N be an m-manifold. Let $(U, \mathbf{x})$ be a chart on M about $\mathbf{p}$ and let $(V, \mathbf{y})$ be a chart on N. Consider a function $\mathbf{F} : U \subseteq M \to V \subseteq N$, and additionally suppose $\mathbf{F}(\mathbf{p}) \in V$. 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 $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{y} \circ \mathbf{f} \circ \mathbf{x}^{-1} : \mathbb{R}^n \to \mathbb{R}^m$.

Theorem 11.12. delete?

If x^i is the ith coordinate function of a chart $(U, \mathbf{x})$, then the coordinate representation $\widetilde{x^i}_{(U,\mathbf{x})}$ of x^i relative to the chart $(U, \mathbf{x})$ is $\widetilde{x^i}_{(U,\mathbf{x})} = x^i \circ \mathbf{x}^{-1} = \epsilon^i \in (\mathbb{R}^n)^*$, where $\{\epsilon^1, ..., \epsilon^n\}$ is the dual basis of $(\mathbb{R}^n)^*$ induced by the standard basis $\{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ of $\mathbb{R}^n$. Recall, $\epsilon^i(\mathbf{p}) = [\mathbf{p}]^i_{\hat{\mathfrak{E}}}$.

Definition 11.13. [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 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 $\mathrm{cl}(\mathbb{H}^n)$, i.e., iff $\mathbf{x}(U) \subseteq \mathrm{cl}(\mathbb{H}^n)$

We classify a point $\mathbf{p} \in M$ according to the type of chart it lies in. $\mathbf{p} \in M$ is 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 $\mathbf{p}$

Theorem 11.14. [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.

Theorem 11.15. [Lee, p.415] (Manifolds with corners are topologically the same as manifolds with boundary).

The title of this theorem is true because $\mathrm{cl}(\mathbb{H}^n)$ is homeomorphic to $\mathbb{H}^n$.

We will see that *smooth* manifolds with corners will be different than *smooth* manifolds with boundary.

Definition 11.16. [Lee, p.11] (Differentiability classes, $C^k(\mathbb{R}^k \to \mathbb{R})$, smooth functions $\mathbb{R}^k \to \mathbb{R}$, diffeomorphisms).

Let M be a n-dimensional smooth manifold. A function $\mathbf{f} : \mathbb{R}^k \to \mathbb{R}$ is said to be of *differentiability class* C^k iff the partial derivatives of $\mathbf{f}$ of orders $0, 1, ..., k$ are continuous. In particular, C^0 functions are real-valued continuous functions on M, and C^1 functions are real-valued continuous functions on M whose first partial derivatives are also continuous. A function $\mathbf{f} : M \to \mathbb{R}$ has *differentiability class* C^∞ iff $\mathbf{f} \in C^k(M)$ for all $k \in \{0, 1, ...\}$.

We define $C^k(\mathbb{R}^k \to \mathbb{R})$ to be the set of functions of differentiability class C^k, and define $C^\infty(M \to \mathbb{R}) := \cup_{k=1}^{\infty} C^k(M)$. We will use the word "smooth" to mean C^∞. Following this convention, the set $C^\infty(\mathbb{R}^k \to \mathbb{R})$ is called the set of *smooth real-valued functions on* $\mathbb{R}^k$.

Lastly, we say that a smooth real-valued function on $\mathbb{R}^k$ is a *diffeomorphism* iff it is smooth, bijective, and has a smooth inverse.

11.3 Smooth manifolds

Derivation 11.17. (Smooth manifold).

Let M be an n-manifold. We say that a function $\mathbf{f} : M \to \mathbb{R}$ is *smooth relative to a chart* $(U, \mathbf{x})$ of M iff the coordinate representation $\widetilde{\mathbf{f}}_{(U,\mathbf{x})} : \mathbf{x}(U \cap V) \subseteq \mathbb{R}^n \to \mathbb{R}$ 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, so the domains of the coordinate representations $\widetilde{\mathbf{f}}_{(U,\mathbf{x})}$ and $\widetilde{\mathbf{f}}_{(V,\mathbf{y})}$ are homeomorphic. We would like to define a notion of *smooth manifold* to be such that, if M is a smooth manifold, then

$$\begin{gathered}\mathbf{f} : M \to \mathbb{R} \text{ is smooth relative to some chart } (U, \mathbf{x}) \\ \Longrightarrow \\ \mathbf{f} \text{ is smooth relative to all other charts } (V, \mathbf{y}) \text{ that overlap } (U, \mathbf{x}),\ U \cap V \neq \emptyset.\end{gathered}$$

In our present situation, this is not the case. If $(U, \mathbf{x})$ and $(V, \mathbf{y})$ intersect, $U \cap V \neq \emptyset$, so that the domains of $\widetilde{\mathbf{f}}_{(U,\mathbf{x})}$ and $\widetilde{\mathbf{f}}_{(V,\mathbf{y})}$ are homeomorphic, it is still possible for $\mathbf{f}$ to be smooth relative to $(U, \mathbf{x})$ but not relative to $(V, \mathbf{y})$. To see why, express $\widetilde{\mathbf{f}}_{(U,\mathbf{x})}$ 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 does not preserve smoothness. (In fact, there is always a choice of $\mathbf{x}$ for which $\widetilde{\mathbf{f}}_{(U,\mathbf{x})}$ is *not* smooth!).

Since smoothness *is* preserved by composing with a diffeomorphism, we define two charts to be *smoothly compatible* iff the transition map between them is a diffeomorphism, and define a *smooth atlas* for M to be one in which any two charts are smoothly compatible[1]. A chart that is an element of a smooth atlas is called a *smooth chart*.

As is noted in [Lee], there are in general "many possible atlases that give the 'same' smooth structure, in that they all determine the same collection of smooth functions on M." So that we can deal with a concrete choice of atlas, we define a *smooth structure* on M to be a maximal smooth atlas. No generality is lost when we take our atlas on M to be the maximal one, since all other atlases are contained in the maximal atlas.

We say that M is a *smooth manifold* iff there exists a smooth structure on M.

Definition 11.18. (Smooth manifolds with boundary and with corners).

every manifold is a manifold with boundary every manifold with boundary is a manifold with corners suffices to define smoothness for manifold with corners

[Lee, p.27]

Example 11.19. [Lee, p.20]

If $U \subseteq \mathbb{R}^n$ is an open subset and $\mathbf{f} : U \to \mathbb{R}^k$ is a smooth function, we have already observed above (Example 1.3) that the graph of $\mathbf{f}$ is a topological n-manifold in the subspace topology. Since graph($\mathbf{f}$) is covered by the single graph coordinate chart $\mathbf{x} : \text{graph}(\mathbf{f}) \to U$ (the restriction of $\boldsymbol{\pi}_1$), we can put a canonical smooth structure on graph($\mathbf{f}$) by declaring the graph coordinate chart $(\text{graph}(\mathbf{f}), \mathbf{x})$ to be a smooth chart.

Example 11.20. [Lee, p.20]

Level sets are also smooth manifolds.

11.4 Submersions, immersions, embeddings

- Let M and N be smooth manifolds with or without boundary and let $\mathbf{f} : M \to N$ be a smooth map. The *rank of* $\mathbf{f}$ *at* $\mathbf{p} \in M$ is the rank of the linear map $d\mathbf{f}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{\mathbf{f}(\mathbf{p})}(N)$. Iff $\mathbf{f}$ has the same rank at every point, it is said to have *constant* rank.

[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.

$\mathbf{f}$ is a *smooth submersion* iff its differential is surjective everywhere (equivalently, $\text{rank}(\mathbf{f}) = \dim(N)$), and is a *smooth immersion* iff its differential is injective everywhere (equivalently, $\text{rank}(\mathbf{f}) = \dim(M)$).

An *embedding* is an injective smooth immersion that is also a homeomorphism onto its image.

- (Proposition 5.2). (Images of embeddings as submanifolds). Suppose M is a smooth manifold with or without boundary, N is a smooth manifold, and $F : N \to M$ is a smooth embedding. Let $S = F(N)$. With the subspace topology, S is a topological manifold, and it has a unique smooth structure making it into an embedded submanifold of M with the property that F is a diffeomorphism onto its image.

11.5 Tangent vectors

No way to directly pull back geometric vectors from $\mathbb{R}^n_{\mathbf{F}(\mathbf{p})}$ to M. We can however pull back derivations at $\mathbf{F}(\mathbf{p}) \in \mathbb{R}^n$.

Definition 11.21. (Derivation at $\mathbf{p} \in \mathbb{R}^n$, tangent space to $\mathbb{R}^n$).
A *derivation at* $\mathbf{p} \in \mathbb{R}^n$ is a function $v_\mathbf{p} : C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}$ which satisfies

$$v_\mathbf{p}(f + g) = v_\mathbf{p}(f) + v_\mathbf{p}(g)$$
$$v_\mathbf{p}(fg) = v_\mathbf{p}(f)g(\mathbf{p}) + f(\mathbf{p})v_\mathbf{p}(g),$$

for all $f, g \in C^\infty(\mathbb{R}^n \to \mathbb{R})$. Due to these conditions, we will see in Theorem [...] that every derivation is a directional derivative.

The set of derivations at $\mathbf{p} \in \mathbb{R}^n$ is called the *tangent space to* $\mathbb{R}^n$ *at* $\mathbf{p}$, and is denoted $T_\mathbf{p}(\mathbb{R}^n)$.

Theorem 11.22. (Basis of $T_\mathbf{p}(\mathbb{R}^n)$).
When $M = \mathbb{R}^n$, the directional derivatives $\left\{ \frac{\partial}{\partial x^i}\Big|_\mathbf{p} \right\}_{i=1}^n$ form a basis for $T_\mathbf{p}(\mathbb{R}^n)$.

Proof. Take the standard basis $\{\hat{\mathfrak{e}}^i_\mathbf{p}\}_{i=1}^n$ for $\mathbb{R}^n_\mathbf{p}$. Since $\mathbf{v}_\mathbf{p} \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_\mathbf{p}$ is an isomorphism $\mathbb{R}^n_\mathbf{p} \to T_\mathbf{p}(\mathbb{R}^n)$, then the images of the $\hat{\mathfrak{e}}^i_\mathbf{p}$ under this isomorphism are a basis for $T_\mathbf{p}(\mathbb{R}^n)$. Thus $\left\{ \frac{\partial}{\partial \hat{\mathfrak{e}}^i}\Big|_\mathbf{p} \right\}_{i=1}^n = \left\{ \frac{\partial}{\partial x^i}\Big|_\mathbf{p} \right\}_{i=1}^n$ is a basis for $T_\mathbf{p}(\mathbb{R}^n)$. □

Lemma 11.23. (Properties of derivations at $\mathbf{p} \in \mathbb{R}^n$). Suppose $\mathbf{p} \in \mathbb{R}^n$, $v_\mathbf{p} \in T_\mathbf{p}(\mathbb{R}^n)$ is a derivation at $\mathbf{p}$, and $f, g \in C^\infty(\mathbb{R}^n \to \mathbb{R})$. 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.

- First set $f_1 \equiv 1$ and use the product rule with $f_1 \cdot f_1 = f_1$ to show $v_\mathbf{p}(f_1) = 0$. Any other constant function $f_c \equiv c$ is of the form $f_c = cf_1$, so $v_\mathbf{p}(f_c) = v_\mathbf{p}(cf_1) = cv_\mathbf{p}(f_1) = 0$ by linearity.
- Use the product rule.

□

Theorem 11.24. (Derivations and directional derivatives are in one-to-one correspondence).
The set of directional derivatives at $\mathbf{p} \in \mathbb{R}^n$ is equal to the set of derivations at $\mathbf{p}$:

$$\left\{ \frac{\partial}{\partial \mathbf{v}}\Big|_\mathbf{p} \mid \mathbf{v} \in \mathbb{R}^n \right\} = T_\mathbf{p}(\mathbb{R}^n).$$

Proof.

($\subseteq$). Showing that every directional derivative at $\mathbf{p} \in \mathbb{R}^n$ is a derivation at $\mathbf{p} \in \mathbb{R}^n$ follows straightforwardly from the definition of "derivation at $\mathbf{p} \in \mathbb{R}^n$".

($\supseteq$). Let $\mathbf{p} \in \mathbb{R}^n$ and let $v_\mathbf{p} : C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}$ be a derivation. We must show that $v_\mathbf{p}$ is a directinoal derivative at $\mathbf{p}$, and find a vector $\mathbf{v}_\mathbf{p} \in \mathbb{R}^n_\mathbf{p}$ for which $v_\mathbf{p} = \frac{\partial}{\partial \mathbf{v}}\Big|_\mathbf{p}$. So, we will find a $\mathbf{v}_\mathbf{p} \in \mathbb{R}^n_\mathbf{p}$ for which

$$v_\mathbf{p}(f) = \Big(\frac{\partial}{\partial \mathbf{v}}\Big|_\mathbf{p}\Big)(f) \text{ for all } f \in C^\infty(\mathbb{R}^n \to \mathbb{R}).$$

We use Taylor's theorem (cite) to write $f \in C^\infty(\mathbb{R}^n \to \mathbb{R})$ as

$$f(\mathbf{x}) = f(\mathbf{p}) + \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_\mathbf{p}(x^i - a^i) + \sum_{i,j=1}^n (x^i - p^i)(x^j - p^j)\int_0^1 (1-t)\frac{\partial^2 f}{\partial x^i \partial x^j}\Big|_{\mathbf{p}+t(\mathbf{x}-\mathbf{p})} dt,$$

where $x^i = [\mathbf{x}]^i_E$, $p^i = [\mathbf{p}]^i_E$.

Now we produce the vector $\mathbf{v}_\mathbf{p}$. Let $\{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ be the standard basis for $\mathbb{R}^n$ and let $\{\epsilon^1, ..., \epsilon^n\}$ be the induced dual basis for $(\mathbb{R}^n)^*$. We set $\mathbf{v}_\mathbf{p} = \sum_{i=1}^n v_\mathbf{p}(\epsilon^i)\mathbf{e}_i$. (Note that applying the derivation $v_\mathbf{p}$ to ϵ^i makes sense because ϵ^i, being the "ith coordinate function on $\mathbb{R}^n$", is a smooth function $\mathbb{R}^n \to \mathbb{R}$). This choice of $\mathbf{v}_\mathbf{p}$ makes more sense in hindsight after reading Lemma [coords]. (That lemma logically depends on this result, though, so we must make this relatively unmotivated choice of $\mathbf{v}_\mathbf{p}$!)

Apply $v_\mathbf{p}$ to f and use Lemma [...] to obtain

$$v_\mathbf{p}(f) = v_\mathbf{p}\Big(\sum_{i,j=1}^n \frac{\partial f}{\partial x^i}\Big|_\mathbf{p}(x^i - p^i)\Big) = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_\mathbf{p}(v_\mathbf{p}(x^i) - v_\mathbf{p}(p^i)) = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_\mathbf{p} v_\mathbf{p}(x^i).$$

When we apply $v_\mathbf{p}$ to x^i in the above, we are using a slight abuse of notation and interpreting x^i to be the coordinate function $\mathbf{x} \mapsto x^i$ evaluated at $\mathbf{x}$. Thinking of x^i in this way, we have $x^i = \epsilon^i$, so the above further simplifies to

$$\sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_\mathbf{p} v_\mathbf{p}(\epsilon^i) = (\nabla \mathbf{f})_\mathbf{p} \cdot \mathbf{v}_\mathbf{p} = \frac{\partial f}{\partial \mathbf{v}}\Big|_\mathbf{p}.$$

Thus, $v_\mathbf{p} = \frac{\partial}{\partial \mathbf{v}}\Big|_\mathbf{p}$. □

Theorem 11.25. (Geometric tangent vectors are naturally isomorphic to derivations).

The map $\mathbf{v}_\mathbf{p} \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_\mathbf{p}$ is a natural linear isomorphism $\mathbb{R}^n_\mathbf{p} \cong T_\mathbf{p}(\mathbb{R}^n)$.

Proof. We need to show $\mathbf{v} \mapsto \frac{\partial}{\partial \mathbf{v}}$ is linear, injective and surjective. The naturality of the isomorphism is immediate because it is a basis-independent definition.

Linearity follows immediately from the fact that the directional derivative $\frac{\partial}{\partial \mathbf{v}}\Big|_\mathbf{p}$ is linear with respect to $\mathbf{v}$ (see Remark 9.17). The map $\mathbf{v}_\mathbf{p} \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_\mathbf{p}$ is surjective because every derivation at $\mathbf{p} \in \mathbb{R}^n$ is a directional derivative at $\mathbf{p} \in \mathbb{R}^n$, so every derivation at $\mathbf{p} \in \mathbb{R}^n$ is of the form $\frac{\partial}{\partial \mathbf{v}}\Big|_\mathbf{p}$ for some $\mathbf{v} \in \mathbb{R}^n$.

It remains to show injectivity. We show the map $\mathbf{v} \mapsto \frac{\partial}{\partial \mathbf{v}}$ has a trivial kernel. So, assume $\mathbf{v}_\mathbf{p}$ is sent to the zero map; we need to show $\mathbf{v}_\mathbf{p} = \mathbf{0}$.

Let $E = \{\mathbf{e}^1, ..., \mathbf{e}^n\}$ be a basis for $\mathbb{R}^n$, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be its dual basis for $(\mathbb{R}^n)^*$. Note that since $\epsilon_i : \mathbb{R}^n \to \mathbb{R}$ is defined by $\epsilon^i(\mathbf{p}) = [\mathbf{p}]^i_E$, each ϵ_i is a C^∞ function on $\mathbb{R}^n$.

At $\mathbf{p} \in \mathbb{R}^n$, the ith coordinate of the zero map is $\frac{\partial}{\partial \mathbf{v_p}}\Big|_{\mathbf{q}}(\epsilon^i)$ (see Theorem 5.26). All coordinates of the zero map relative to any basis must be zero, so, using the linearity of the directional derivative with respect to $\mathbf{v_p}$ (again, see Remark 9.17), we have

$$0 = \frac{\partial}{\partial \mathbf{v_p}}\Big|_{\mathbf{p}}(\epsilon_i) = \frac{\partial}{\partial(\sum_j [\mathbf{v_p}]^j_E \mathbf{e}^j)}\Big|_{\mathbf{p}}(\epsilon_i) = \sum_j [\mathbf{v_p}]^j_E \Big(\frac{\partial}{\partial \mathbf{e}^j}\Big|_{\mathbf{p}}\Big)(\epsilon_i).$$

Using the fact that $\frac{\partial}{\partial \mathbf{e}^j}\Big|_{\mathbf{p}} = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$, the above becomes

$$0 = \sum_j [\mathbf{v_p}]^j_E \delta^i_j = [\mathbf{v_p}]^i_E.$$

We see $[\mathbf{v_p}]^i_E = 0$ for all i, so $\mathbf{v_p} = \mathbf{0}$. □

Definition 11.26. (Derivation at $\mathbf{p} \in M$, tangent space to a manifold).
Let M be a smooth n-manifold with or without boundary. A *derivation at* $\mathbf{p} \in M$ is a linear map $v_\mathbf{p} : C^\infty(M \to \mathbb{R}) \to \mathbb{R}$ that satisfies the product rule,

$$v_\mathbf{p}(fg) = v_\mathbf{p}(f)g(\mathbf{p}) + f(\mathbf{p})v_\mathbf{p}(g) \text{ for all } f, g \in C^\infty(M \to \mathbb{R}).$$

The set of derivations at $\mathbf{p} \in M$ is called the *tangent space to* M *at* $\mathbf{p}$, and is denoted $T_\mathbf{p}(M)$.

Theorem 11.27. (Properties of derivations at $\mathbf{p} \in M$).
Let M be a smooth n-manifold with or without boundary, let $\mathbf{p} \in \mathbb{R}^n$, $v_\mathbf{p} \in T_\mathbf{p}(\mathbb{R}^n)$ be a derivation at $\mathbf{p}$, and $f, g \in C^\infty(\mathbb{R}^n \to \mathbb{R})$. 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 [...]. □

Definition 11.28. [War, p.14, 15] (Basis of $T_\mathbf{p}(M)$ and its abuse of notation).
Let M be a smooth n-manifold with or without boundary, let $(U, \mathbf{x})$ be a smooth chart on M, and let $\mathbf{p} \in M$.

In an abuse of notation, we define $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$, $i \in \{1, ..., n\}$, to be the derivations at $\mathbf{p} \in U \subseteq M$ for which

$$\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}\Big)(f \circ \mathbf{x}^{-1}),$$

where the $\frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial}{\partial \hat{\mathfrak{e}}^i}\Big|_{\mathbf{x}(\mathbf{p})}$ on the right hand side are directional derivatives taking in smooth functions on $\mathbb{R}^n$ as their arguments. The $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ on the left hand side are what we are defining in our abuse of notation, and (we will see) are derivations $C^\infty(M \to \mathbb{R}) \to \mathbb{R}$; that is, they are elements of $T_\mathbf{p}(M)$.

Importantly, $\Big\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big\}_{i=1}^n$ is a basis of $T_\mathbf{p}(M)$. This follows because the $\frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$ on the right hand side are a basis for $T_{\mathbf{x}(\mathbf{p})}(\mathbb{R}^n)$ (see Theorem 11.22).

In the special case when $M = \mathbb{R}^n$, then $\mathbf{x}$ is the identity on $M = \mathbb{R}^n$, so $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ are the partial derivatives $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} = \frac{\partial}{\partial \hat{\mathfrak{e}}^i}\Big|_{\mathbf{p}}$. (This restates- and is true because of- Theorem 11.22).

Lastly, if we consider the coordinate representation $\widetilde{\mathbf{f}}_{(U,\mathbf{x})} = \mathbf{f} \circ \mathbf{x}^{-1}$ of $\mathbf{f}$ relative to the chart $(U, \mathbf{x})$, then the above definition becomes

$$\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \widetilde{\mathbf{f}}_{(U,\mathbf{x})}.$$

Proof. We need to show that the $\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}} : C^\infty(M \to \mathbb{R}) \to \mathbb{R}$ from the left side of the condition of the definition are derivations at $\mathbf{p} \in M$. In this proof, we put an "LHS" under $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ when we mean $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ to be from the left hand side of the condition in the above definition. All other occurrences of $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ in this proof are directional derivatives.

We need to show that $\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}$ are linear and follow the product rule. Linearity follows easily from the linearity of the directional derivative with respect to its argument from $C^\infty(\mathbb{R}^n \to \mathbb{R})$. We show the product rule holds:

$$\begin{aligned}
\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}(fg) &= \frac{\partial(fg \circ \mathbf{x}^{-1})}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\Big|_{\mathbf{x}^{-1}(\mathbf{x}(\mathbf{p}))} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \\
&= \Big(\frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial g}{\partial x^i}\Big|_{\mathbf{p}}\Big)\frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \\
&= \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial g}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \\
&= \frac{\partial(f \circ \mathbf{x})}{\partial x^i}\Big|_{\mathbf{p}} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial(g \circ \mathbf{x})}{\partial x^i}\Big|_{\mathbf{p}} \\
&= \Big(\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}\Big)(f)\; g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\; \Big(\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}\Big)(g).
\end{aligned}$$

Here, we used the chain and product rules and then reversed the product and chain rules. □

11.5.1 Misc. tangent vector stuff- reorganize

Tangent bundle, $T(M) := \sqcup_{\mathbf{p} \in M} T_{\mathbf{p}}(M)$.

(Proposition 3.18). For any smooth n-manifold M; the tangent bundle $T(M)$ has a natural topology and smooth structure that make it into a $2n$-dimensional smooth manifold.

Proof idea. Follow the special case of $M = \mathbb{R}^n$:

$$T(\mathbb{R}^n) = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} T_{\mathbf{p}}(\mathbb{R}^n) = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} \mathbb{R}^n_{\mathbf{p}} = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} (\mathbb{R}^n \times \{\mathbf{p}\}) = \mathbb{R}^n \times \mathbb{R}^n.$$

(Figure out why last equality works).

(Proposition 3.23). Every $\mathbf{v} \in T_{\mathbf{p}}(M)$ is the velocity of some smooth curve in M.

Vector bundles

- (Proposition A.23). Let $X_1, ..., X_n$ be topological spaces, and consider the product topology $X_1 \times ... \times X_n$. The *projection onto the ith factor* is the map $\boldsymbol{\pi}_i : X_1 \times ... \times X_n \to X_i$ defined by $\boldsymbol{\pi}(\mathbf{p}_1, ..., \mathbf{p}_n) = \mathbf{p}_i$.

- Let B be a topological space (thought of as the *base space*). A *vector bundle of rank k over B* is a tuple $(E, \boldsymbol{\sigma})$, where...
  - E is a topological space (thought of as the *entire space*, or *total space*).
  - $\boldsymbol{\sigma} : B \to \text{powerset}(E)$ is a map whose left-inverse restricted onto singletons, $\boldsymbol{\pi} = (\boldsymbol{\sigma}^{-1})|_{\cup_{\mathbf{p} \in E}\{\mathbf{p}\}}$ is a surjective continuous map satisfying...
    * For each $\mathbf{p} \in B$, the set $\boldsymbol{\sigma}(\mathbf{p})$ (called the *fiber over* $\mathbf{p}$) is a k-dimensional vector space.
    * For each $\mathbf{p} \in B$, there is a neighborhood $U \subseteq B$ with $\mathbf{p} \in U$ (called a *local trivialization*) for which there is a homeomorphism $\mathbf{F} : U \times \mathbb{R}^k \to \boldsymbol{\sigma}(U)$ satisfying two more conditions: for all $\mathbf{q} \in U$,
      · $(\boldsymbol{\pi} \circ \mathbf{F})(\mathbf{q}, \mathbf{v}) = \mathbf{q}$ for all $\mathbf{v} \in \mathbb{R}^k$
      · $\mathbf{v} \mapsto \mathbf{F}(\mathbf{q}, \mathbf{v})$ is a linear isomorphism $\mathbb{R}^k \cong \boldsymbol{\sigma}(\mathbf{q})$

  A vector bundle can be pictured as a hairbrush. The base space B is thought of as the surface of the hairbrush's handle, and the entire space E is thought of as the disjoint union of B with the protruding bristles of the brush (so the bristles are $E - B$). In this analogy, $\boldsymbol{\sigma}(\mathbf{p})$ is the set of points on a bristle above a particular point $\mathbf{p} \in B$ on the handle; it is therefore called the *fiber over* $\mathbf{p}$.

  The map $\boldsymbol{\pi}$ is called the *projection*.

  A *cross section*, or simply *section*, of E, is some continuous restriction $\boldsymbol{\sigma}|_A$ of $\boldsymbol{\sigma}$ onto a subset $A \subseteq B$, where A is chosen so that $\boldsymbol{\sigma}|_A$ is one-to-one.

  The common convention is to state the definition of a vector bundle in terms of the surjective continuous restriction onto singletons $\boldsymbol{\pi}$.

- Let E be a vector bundle over B. If B and E are smooth manifolds with or without boundary, $\boldsymbol{\pi}$ is a smooth map, and the local trivializations can be chosen to be diffeomorphisms, then E is called a *smooth vector bundle*. A *smooth (cross) section* of a smooth vector bundle E is a smooth section of E.

- Pedagogically, it's fine to say "smooth map from some subset of B to E" to mean "smooth section of E". So, since the tensor bundle is a vector bundle over M, a smooth tensor field is a smooth map from some subset $T^p_q(T(M))$ to M.

11.5.2 The differential of a smooth map $M \to N$

Definition 11.29. (Differential of a smooth map $\mathbb{R}^n \to \mathbb{R}^m$).

Consider a smooth map $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$. We define the *differential* $d\mathbf{F}_\mathbf{p} : T_\mathbf{p}(\mathbb{R}^n) \to T_{\mathbf{F}(\mathbf{p})}(\mathbb{R}^m)$ *of* $\mathbf{F}$ *at* $\mathbf{p}$ to be the linear function whose matrix relative to the bases $\{\frac{\partial}{\partial x^i}\}$ for $T_\mathbf{p}(\mathbb{R}^n)$ and $\{\frac{\partial}{\partial y^i}\}$ for $T_\mathbf{p}(\mathbb{R}^m)$ is the Jacobian matrix of $\mathbf{F}$ at $\mathbf{p}$:

$$d\mathbf{F}_\mathbf{p} := \frac{\partial \mathbf{F}}{\partial \mathbf{x}}\Big|_\mathbf{p} = \Big(\frac{\partial F^i}{\partial x^j}\Big|_\mathbf{p}\Big).$$

(Here, F^i is the ith component function of $\mathbf{F}$).
(We will show that if $v_\mathbf{p} \in T_\mathbf{p}(M)$, then $d\mathbf{F}_\mathbf{p}(v_\mathbf{p})$ is a derivation at $\mathbf{F}(\mathbf{p})$ in $T_{\mathbf{F}(\mathbf{p})}(N)$ later).
This means that $d\mathbf{F}_\mathbf{p}$ acts on a basis tangent vector $\frac{\partial}{\partial x^i}$ by

$$d\mathbf{F}_\mathbf{p}\Big(\frac{\partial}{\partial x^i}\Big) = \sum_{j=1}^m a^j_i \frac{\partial}{\partial y^j} = \sum_{j=1}^m \frac{\partial F^j}{\partial x^i}\Big|_\mathbf{p} \frac{\partial}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})},$$

where $(a^i_j) = \Big[d\mathbf{F}_\mathbf{p}\Big(\Big\{\frac{\partial}{\partial x^i}\Big\}\Big)\Big]_{\{\frac{\partial}{\partial y^j}\}}$ is the matrix of $d\mathbf{F}_\mathbf{p}$ relative to the bases $\{\frac{\partial}{\partial x^i}\}$ and $\{\frac{\partial}{\partial y^i}\}$.

Using linearity, we see $d\mathbf{F}_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big)$ acts on $f \in C^\infty(U \subseteq \mathbb{R}^n \to \mathbb{R})$ by

$$\begin{aligned} d\mathbf{F}_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big)(f) &= \Big(\sum_{j=1}^m \frac{\partial F^j}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})}\Big)(f) = \sum_{j=1}^m \Big[\Big(\frac{\partial F^j}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})}\Big)(f)\Big] \\ &= \sum_{j=1}^m \frac{\partial F^j}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial f}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})} = \sum_{j=1}^m \frac{\partial f}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})} \frac{\partial F^j}{\partial x^i}\Big|_{\mathbf{p}} = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}(f \circ \mathbf{F}). \end{aligned}$$

Note, this last $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ is a directional derivative, $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} = \frac{\partial}{\partial \hat{\mathfrak{e}}^i} : C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}^n$.
The above shows

$$d\mathbf{F}_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}(f \circ \mathbf{F}).$$

Since $d\mathbf{F}_{\mathbf{p}}$ is linear and $\Big\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big\}$ is a basis for $T_{\mathbf{p}}(\mathbb{R}^n)$, the above condition extends to any $v_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^n)$:

$$d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})(f) = v_{\mathbf{p}}(f \circ \mathbf{F}).$$

This characterization of the differential is coordinate free, and therefore provides an easy way to define the differential in a more general setting. We do this in the next definition.

Definition 11.30. (Differential of a smooth map $\mathbf{F} : M \to N$).
Let M and N be smooth n- and m- dimensional manifolds with or without boundary.
We define the *differential* $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{\mathbf{F}(\mathbf{p})}(N)$ *of* $\mathbf{F}$ *at* $\mathbf{p}$ by

$$d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})(f) = v_{\mathbf{p}}(f \circ \mathbf{F}),$$

where $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ and $f \in C^\infty(N)$.

Proof. We need to check that $d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})$ is a derivation at $\mathbf{F}(\mathbf{p}) \in N$. To do so, follow the proof of Theorem [..], which showed $\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}$ is a derivation at $\mathbf{x}(\mathbf{p}) \in \mathbb{R}^n$. □

Theorem 11.31. (Properties of differentials). Let M, N, and P be smooth manifolds with or without boundary, let $\mathbf{F} : M \to N$ and $\mathbf{G} : N \to P$ be smooth maps, and let $\mathbf{p} \in M$.

- (Chain rule). $d(\mathbf{G} \circ \mathbf{F})_{\mathbf{p}} = d\mathbf{G}_{\mathbf{F}(\mathbf{p})} \circ d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{(\mathbf{G}\circ\mathbf{F})(\mathbf{p})}(P)$.
- (Differential of the identity). $d(\mathbf{I}_M)_{\mathbf{p}} = \mathbf{I}_M$, where $\mathbf{I}_M : T_{\mathbf{p}}(M) \to T_{\mathbf{p}}(M)$ is the identity on M.
- If $\mathbf{F}$ is a diffeomorphism, then $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{\mathbf{F}(\mathbf{p})}(N)$ is a linear isomorphism, and $(d\mathbf{F}_{\mathbf{p}})^{-1} = d(\mathbf{F}^{-1})_{\mathbf{F}(\mathbf{p})}$.

Proof. See Exercise 3.7 of Lee. [cite more] □

11.6 Tangent covectors

Definition 11.32. (Cotangent space to a manifold).

$\phi_{\mathbf{p}} \in T^*_{\mathbf{p}}(M)$

Theorem 11.33. [Lee, p.281] (Differential of a smooth map $M \to \mathbb{R}$).

What happens when we take the differential of a smooth $f : M \to \mathbb{R}$? Well, by definition of the differential of a smooth map $M \to N$, we have

$$df_{\mathbf{p}}(v_{\mathbf{p}})(g) = v_{\mathbf{p}}(g \circ f).$$

Notice that $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{f(\mathbf{p})}(\mathbb{R})$, where $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ by Theorem [...]. There is therefore an induced map $\widetilde{df_{\mathbf{p}}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ for which the following diagram commutes:

$$\begin{tikzcd}
T_{\mathbf{p}}(M) \arrow[r, "v_{\mathbf{p}} \mapsto df_{\mathbf{p}}(v_{\mathbf{p}})"] \arrow[dr, "\widetilde{df_{\mathbf{p}}}"'] & T_{\mathbf{f(p)}}(\mathbb{R}) \arrow[d, "w \mapsto \sum_{i=1}^n w(x^i)\hat{\mathfrak{e}}_i"] \\
& \mathbb{R}
\end{tikzcd}$$

From the diagram, we see that map geometric "vector" $\widetilde{df_{\mathbf{p}}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ acts on a tangent vector $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by

$$\widetilde{df_{\mathbf{p}}(v_{\mathbf{p}})} = \sum_{i=1}^{1} df_{\mathbf{p}}(v_{\mathbf{p}})(x^i)\hat{\mathfrak{e}}_i = df_{\mathbf{p}}(v_{\mathbf{p}})(x^1)\hat{\mathfrak{e}}_1 = v_{\mathbf{p}}(x^1 \circ f)\hat{\mathfrak{e}}_1.$$

Since $\mathbb{R}^1$ is one-dimensional, then $\mathbb{R}^1 \cong \mathbb{R}$. As we switch from $\mathbb{R}^1$ to $\mathbb{R}$, the coordinate function $x^1 : \mathbb{R}^1 \to \mathbb{R}^1$ becomes the identity on $\mathbb{R}$ and $\hat{\mathfrak{e}}_1$ becomes the scalar $1 \in \mathbb{R}$. These final identifications[2] give

$$\widetilde{df_{\mathbf{p}}(v_{\mathbf{p}})} = v_{\mathbf{p}}(f).$$

In practice, we write $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ to mean $\widetilde{df_{\mathbf{p}}(v_{\mathbf{p}})}$. So, the above characterizing condition of the differential of $f : M \to \mathbb{R}$ is restated as

$$\boxed{df_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)}$$

Theorem 11.34. [Lee, p.281]

$$\boxed{\begin{aligned} df_{\mathbf{p}} &= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}} \\ df &= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx^i \end{aligned}}$$

Remark 11.35. [Lee, p.282, 283] Lee's Taylor expansion remark

[2] To be very formal, we could write the above as $\widetilde{\widetilde{df_{\mathbf{p}}(v_{\mathbf{p}})}}$ to indicate that it is the map obtained from $\widetilde{df_{\mathbf{p}}(v_{\mathbf{p}})}$ by identifying $\mathbb{R}^1 \cong \mathbb{R}$, but this identification is so trivial that we consider $\widetilde{df_{\mathbf{p}}(v_{\mathbf{p}})}$ and $\widetilde{\widetilde{df_{\mathbf{p}}(v_{\mathbf{p}})}}$ to be the same function.

Remark 11.36. (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 to compute

$$df_{\mathbf{p}}(\mathbf{v}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i(\mathbf{v}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} [\mathbf{v}]^i_{\hat{\mathfrak{E}}} = (\nabla_{\mathbf{x}} f) \cdot \mathbf{v} = \frac{\partial f}{\partial \mathbf{v}},$$

where $\frac{\partial f}{\partial \mathbf{v}}$ is the directional derivative of f in the direction of $\mathbf{v}$. This is to be expected because we defined the differential of a smooth map of smooth manifolds (see Definition [...]) so that its coordinate representation is represented by the Jacobian relative to the coordinate bases. Here, the row-matrix of partial derivatives is the Jacobian matrix of f; recall that the Jacobian is used to express the directional derivative of a smooth function $\mathbb{R}^n \to \mathbb{R}^m$; here we have $m = 1$ (see Theorem [jacobian]).

11.7 Tangent vectors and tangent covectors with coordinates

Theorem 11.37. (Induced bases in a chart).

Let M be a smooth n-manifold with or without boundary and let $(U, \mathbf{x})$ be a smooth chart on M about $\mathbf{p} \in M$. Consider the set $C^\infty(U \subseteq M \to \mathbb{R})$ of smooth real-valued functions defined on $U \subseteq M$ as a vector space over $\mathbb{R}$. Then

1. $\{x^i|_{\mathbf{p}}\}_{i=1}^n$ is a basis for $C^\infty(U \subseteq M \to \mathbb{R})$, where x^i is the ith coordinate function of $\mathbf{x}$, and where we denote $x^i|_{\mathbf{p}} := x^i(\mathbf{p})$.
2. The dual basis for $T_{\mathbf{p}}(M) = C^\infty(U \subseteq M \to \mathbb{R})^*$ induced by $\{x^i\}_{i=1}^n$ is $\left\{ \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \right\}$.
3. The dual basis for $T^*_{\mathbf{p}}(M) = T_{\mathbf{p}}(M)^* = C^\infty(U \subseteq M \to \mathbb{R})^{**}$ induced by $\left\{ \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \right\}$ is $dx^i|_{\mathbf{p}}$, where the d here is the differential of a function $M \to \mathbb{R}$ (see Theorem 11.33) and x^i is again the ith coordinate function of $\mathbf{x}$.

Proof.

1. easy
2. $\left(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \right)(x^j) = \delta^j_i$
3. proved as part of showing $df_{\mathbf{p}} = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}$ in Theorem 11.33

□

Theorem 11.38. (Coordinates of tangent vectors and dual vectors).

Theorem 5.26 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 with or without boundary and let $(U, \mathbf{x})$ be a smooth chart on M about

$\mathbf{p} \in M$. Then the ith coordinate of a tangent vector $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ relative to $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}_{i=1}^n$ and the ith coordinate of a tangent covector $\phi_{\mathbf{p}} \in T^*_{\mathbf{p}}(M)$ are

$$\boxed{\begin{aligned} &([v_{\mathbf{p}}]_{\left\{\frac{\partial}{\partial x^j}\big|_{\mathbf{p}}\right\}_{j=1}^n})^i = v_{\mathbf{p}}(x^i) = \phi_{x^i}(v_{\mathbf{p}}) \\ &([\phi_{\mathbf{p}}]_{\left\{dx^j|_{\mathbf{p}}\right\}_{j=1}^n})_i = \phi_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big) \end{aligned}}$$

In the second equation of the first line, ϕ_{x^i} is the element of $C^\infty(U \subseteq M \to \mathbb{R})^{**} = T^*_{\mathbf{p}}(M)$ that is identified with the ith coordinate function $x^i \in C^\infty(U \subseteq M \to \mathbb{R})$ of $\mathbf{x}$. Recall Theorem 4.23 to see that $\phi_f : C^\infty(U \subseteq M \to \mathbb{R})^* = T_{\mathbf{p}}(M) \to \mathbb{R}$, where $f \in C^\infty(U \subseteq M \to \mathbb{R})$, is defined by $\phi_f(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)$.

The second equation of the first line is not of much practical use, but it helps formalize the precise way in which $x^i|_{\mathbf{p}}$ and $dx^i|_{\mathbf{p}}$ are "the same": $dx^i|_{\mathbf{p}} = (\phi_{x^i})|_{\mathbf{p}}$.

Remark 11.39. (Interpretations of x^i).

Let M be a smooth n-manifold, and consider a smooth chart $(U, \mathbf{x})$ about $\mathbf{p} \in M$.

When considering the basis $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}$ for $T_{\mathbf{p}}(M)$, the x^i in the "denominator" is *not* a coordinate function of the smooth chart $\mathbf{x}$ that is involved in the definition $\left(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right)(f) = \frac{\partial(f \circ \mathbf{x}^{-1})}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$. (Recall Definition [...]). In this context, the x^i in the "denominator" on the left hand side is simply notation that evokes the mental imagery of the meaning of the x^i on the right hand side (on the right hand side, the x^i in the "denominator" is used in the notation $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} := \frac{\partial}{\partial \hat{\mathbf{e}}^i}\Big|_{\mathbf{p}}$ for directional derivatives which act on smooth functions defined on $\mathbb{R}^n$). The purpose of the function $\mathbf{x}$, from the smooth chart $(U, \mathbf{x})$, is unrelated to this mental imagery; $\mathbf{x}$ serves the different purpose of producing the coordinate representation $\widetilde{\mathbf{f}}_{(U,x)} = (\mathbf{f} \circ \mathbf{x}^{-1}) : C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}$ of $\mathbf{f}$.

However, we have also seen that it is useful to use x^i to denote a coordinate function of $\mathbf{x}$ when we are interested in the coordinates of $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ relative to $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}$, since $x^i(v_{\mathbf{p}})$ is the ith coordinate of $v_{\mathbf{p}}$ relative to $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}$.

A general rule is that when x^i appears in a "numerator" or "by itself", then x^i is a coordinate function that is the argument of a directional derivative, and, when x^i appears in a "denominator", it is because that "denominator" is part of the basis vector $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ of $T_{\mathbf{p}}(M)$. (In the special case of $M = \mathbb{R}^n$, then $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ is a directional derivative that acts on smooth functions defined on $\mathbb{R}^n$. In this special case, the "mental imagery" mentioned above has been realized, because the chart $\mathbf{x}$ is the identity).

Theorem 11.40. (Change of coordinates for derivations in terms of basis vectors of $T_{\mathbf{p}}(M)$).

Theorem 2.60 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-dimensional manifold, and consider smooth charts $(U, \mathbf{x})$ and $(V, \mathbf{y})$, where $\mathbf{p} \in U \cap V$. Let x^i and y^i denote the ith coordinate functions of $\mathbf{x}$ and $\mathbf{y}$, respectively. Applying the above theorem to the vector space $T_{\mathbf{p}}(M)$ and its bases $E = \left\{\frac{\partial}{\partial x^j}\Big|_{\mathbf{p}}\right\}_{j=1}^n$ and $F = \left\{\frac{\partial}{\partial y^j}\Big|_{\mathbf{p}}\right\}_{j=1}^n$, we have

$$[\mathbf{F}]_E = \left(\left[\frac{\partial}{\partial y^1}\Big|_{\mathbf{p}} \right]_F \quad \cdots \quad \left[\frac{\partial}{\partial y^n}\Big|_{\mathbf{p}} \right]_E \right) = \begin{pmatrix} \frac{\partial x^1}{\partial y^1}\Big|_{\mathbf{p}} & \cdots & \frac{\partial x^1}{\partial y^n}\Big|_{\mathbf{p}} \\ \vdots & & \vdots \\ \frac{\partial x^n}{\partial y^1}\Big|_{\mathbf{p}} & \cdots & \frac{\partial x^n}{\partial y^n}\Big|_{\mathbf{p}} \end{pmatrix} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}},$$

where $\frac{\partial x^i}{\partial y^j}\Big|_{\mathbf{p}} = \left(\frac{\partial}{\partial y^j}\Big|_{\mathbf{p}} \right)(x^i)$, and where x^i is the ith coordinate function of $\mathbf{x}$. The matrix $\frac{\partial \mathbf{x}}{\partial \mathbf{y}}$ is the Jacobian matrix described in Theorem [jacobian].

Applying the fact $\mathbf{f}_i = \sum_{j=1}^n ([\mathbf{F}]_E)_{ji} \mathbf{e}_j$ from above, we have

$$\boxed{\frac{\partial}{\partial y^i}\Big|_{\mathbf{p}} = \sum_{j=1}^n \frac{\partial x^j}{\partial y^i}\Big|_{\mathbf{p}} \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}.}$$

This change of basis equation strongly resembles the chain rule, and indeed simplifies to the chain rule when $M = \mathbb{R}^n$. In the general case when M is not necessarily $M = \mathbb{R}^n$, be sure to interpret the x^i's and y^i's as described in the previous remark.

11.8 Vector fields and covector fields

- A *vector field* is a section of the map $\boldsymbol{\pi} : T(M) \to M$. That is, a vector field is a continuous map $\mathbf{V} : M \to T(M)$. A smooth vector field is a smooth such map. The set of all smooth vector fields on M is denoted $\mathfrak{X}(M)$. $\mathfrak{X}(M)$ is a module over $C^\infty(M \to \mathbb{R})$.
- A *local frame for M over U* is an ordered n-tuple of vector fields $(\mathbf{E}_1, ..., \mathbf{E}_n)$ where each $\mathbf{E}_i : U \subseteq M \to \mathbb{R}^n$ is defined on an open subset $U \subseteq M$ and where at each $\mathbf{p} \in U \subseteq M$, the $\mathbf{E}_i$'s are linearly independent and span $T_{\mathbf{p}}(M)$. $(\mathbf{E}_1, ..., \mathbf{E}_n)$ 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 for M over U* is an ordered n-tuple of covector fields $(\boldsymbol{\mathcal{E}}^1, ..., \boldsymbol{\mathcal{E}}^n)$ where each $\boldsymbol{\mathcal{E}}^i : U \subseteq M \to \mathbb{R}^n$ is defined on an open subset $U \subseteq M$ and where at each $\mathbf{p} \in U \subseteq M$, the $\boldsymbol{\mathcal{E}}^i$'s are linearly independent and span $T^*_{\mathbf{p}}(M)$. $(\mathbf{E}_1, ..., \mathbf{E}_n)$ is a *smooth frame* iff each $\boldsymbol{\mathcal{E}}^i$ is smooth, and is a *global frame* iff $U = M$. A local frame is essentially a "basis" of covector fields.
- Given a smooth chart $(U, \mathbf{x})$ on M, the smooth vector field defined by $\mathbf{p} \mapsto \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ is denoted $\frac{\partial}{\partial x^i}$ and is called the *ith coordinate vector field.* [Lee, p.176]
  - Won't formally show it's smooth
- The smooth local frame $\{\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}\}$, where $\frac{\partial}{\partial x^i}$ is the ith coordinate vector field, is called a *coordinate frame* [Lee, p.178]
- $\frac{\partial}{\partial y^i} = \sum_{i=1}^n \frac{\partial x^i}{\partial y^i} \frac{\partial}{\partial x^i}$ due to earlier result [Lee, p.275]
- The smooth covector field defined by $\mathbf{p} \mapsto dx^i|_{\mathbf{p}}$ is denoted by dx^i and is called the *ith coordinate covector field.*
  - Won't formally show it's smooth
- The smooth (see [Lee, p.278]) coframe $\{dx^1, ..., dx^n\}$ is called the *coordinate coframe.* [Lee, p.278]
- A frame and coframe are said to be *dual* to each other iff, pointwise, one is the induced dual basis of the other. So, $\frac{\partial}{\partial x^i}$ and dx^i are dual. [Lee, p.278]
- The differential of a smooth function $f : M \to \mathbb{R}$ is a smooth covector field. Won't show smoothness formally. [Lee, p.281]

11.9 The boundary of a manifold

- $\dim(\partial M) = \dim(M) - 1$. We say ∂M has a *codimension* of 1.
- If M is a smooth manifold with boundary, then the boundary ∂M of M is a smooth manifold properly embedded in M. (A manifold N is said to be *embedded* in a manifold M iff there exists an embedding, which is an injective smooth immersion, $N \to M$. An *immersion* is a smooth map between manifolds whose derivative is everywhere injective. See Lee Theorem 5.11).
- "slice charts for embedded submanifolds"

11.10 Oriented manifolds and their oriented boundaries

Definition 11.41. [Lee, p.380] (Oriented manifolds).
Define pointwise orientation of a manifold in terms of orientation of its tangent spaces.
...
Define oriented manifold

Definition 11.42. [Lee, p.381, 382] (Orientation of a smooth chart on an oriented manifold).

Inward- and outward- pointing vectors

-
-
- There exists a global smooth vector field on M whose restriction to ∂M is everywhere inward-pointing, and one whose restriction to ∂M is everywhere outward-pointing. (See Lee problem 8-4).
- Let $(U, \mathbf{x})$ be a smooth chart on ∂M with $\mathbf{p} \in U$. We classify vectors in $T_{\mathbf{p}}(M)$ as follows.
  - A vector is *inward-pointing (on ∂M)* iff it has positive-x^n component.
  - A vector is *tangent to ∂M* iff it has an x^n-component of zero.
  - A vector is *outward-pointing (on ∂M) iff it has negative-x^n component*, i.e., $\mathbf{v} \in T_{\mathbf{p}}(M)$ is outward pointing iff $-\mathbf{v}$ is inward pointing.
  - Example. Consider the unit disk D in $\mathbb{R}^2$, and pick a point $\mathbf{p}$ on the boundary. Any vector thought of as anchored at $\mathbf{p}$ is either in $T_{\mathbf{p}}(D)$, which is a line, or in one of the two half-spaces resulting from the splitting of $\mathbb{R}^2$ by the line $T_{\mathbf{p}}(\partial D)$. Both halves of $\mathbb{R}^2$ are homeomorhpic to half planes. One of these half planes contains $D - \{\mathbf{p}\}$ and the other one does not, so it makes sense to call vectors in the first half-plane "inward pointing" and vectors in the second half-plane "outward pointing."
- Let ω be an orientation form on M in the actual function sense. Since there exists a vector field $\mathbf{N}$ that is nowhere tangent to ∂M, then there is an induced *induced orientation form ω' on the boundary (due to interior multiplication*[3]) defined by $\omega'(\mathbf{v}_1, ..., \mathbf{v}_{n-1}) = \omega(\mathbf{N}(\mathbf{p}), \mathbf{v}_1, ..., \mathbf{v}_{n-1})$. (Lee Proposition 15.21). The orientation induced by ω' does not depend on the vector field $\mathbf{N}$ that is nowhere tangent to ∂M. (Lee Proposition 15.24).

  Proof. We need to show (1) that ω' is indeed an orientation form on ∂M and (2) that the orientation induced by ω' is independent of the choice of the nowhere tangent vector field $\mathbf{N}$.

[3] We could have put $\mathbf{N}(\mathbf{p})$ in any of ω's n argument slots, but we chose to use the first. This choice corresponds to the operation called *interior multiplication*, which you can read about in Lee's book. See p. 358 and Corollary 14.21 on p. 362.

1. We show $\omega' > 0$ on [tangent bundle of ∂M]; it suffices to show $\omega' \neq 0$ on [tangent bundle] because $\omega > 0$ on [tangent bundle of M].

   Let $\{\mathbf{e}_1, ..., \mathbf{e}_{n-1}\}$ be an ordered basis for $T_{\mathbf{p}}(\partial M)$. For all $\mathbf{p} \in M$, $\mathbf{N}(\mathbf{p})$ is not tangent to M, so $\{\mathbf{N}(\mathbf{p}), \mathbf{e}_1, ..., \mathbf{e}_{n-1}\}$ is a basis for $T_{\mathbf{p}}(M)$, since $T_{\mathbf{p}}(M)$ is identifiable with $\mathbb{R}^n$. ω is nonvanishing, so it cannot be zero on any ordered basis. This means ω' nonvanishing as well.

2. Let $\mathbf{N}$ and $\mathbf{N}'$ be two vector fields that are both nowhere tangent to ∂M. We need to show that the ordered bases $E = \{\mathbf{N}(\mathbf{p}), \mathbf{v}_1, ..., \mathbf{v}_{n-1}\}$ and $F = \{\mathbf{N}'(\mathbf{p}), \mathbf{v}_1, ..., \mathbf{v}_{n-1}\}$ have the same orientation. To do so, we prove that the determinant of the change of basis matrix between the two ordered bases is positive.

   $\mathbf{N}$ and $\mathbf{N}'$ are both outward-pointing, so the 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 nth entry is $\frac{(\mathbf{N}'(\mathbf{p}))_n}{(\mathbf{N}(\mathbf{p}))_n}$, and for $i > 1$, the ith column of this matrix is $\hat{\mathfrak{e}}_i$. The change of basis matrix is therefore upper triangular, so its determinant is the product of the diagonal entries, i.e., the determinant is $\frac{(\mathbf{N}'(\mathbf{p}))_n}{(\mathbf{N}(\mathbf{p}))_n} > 0$.

□

12

Tangent vs. cotangent

- Tangent space at $\mathbf{p}$
- Cotangent space at $\mathbf{p}$, $T^*_{\mathbf{p}}(M) := T_{\mathbf{p}}(M)^*$. An element of $T_{\mathbf{p}}(M)^*$ is called a *covector (at* $\mathbf{p}$*)*.
- Vector fields, frames $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$
- Covector fields, coframes $\{\boldsymbol{\mathcal{E}}^1, ..., \boldsymbol{\mathcal{E}}^n\}$
- Tangent bundle: $T(M) := \bigsqcup_{\mathbf{p}\in M} T_{\mathbf{p}}(M)$. Note that a disjoint union is *not* a "pairwise disjoint union".
- Cotangent bundle: $T^*(M) := \bigsqcup_{\mathbf{p}\in M} T^*_{\mathbf{p}}(M)$.
- Tangent bundle and cotangent bundle are smooth vector bundles.
- Tensor bundle: $T^p_q(T(M)) := \bigsqcup_{\mathbf{p}\in M} T^p_q(T_{\mathbf{p}}(M))$.
  - Have $T^p_0(T(M)) = \bigsqcup_{\mathbf{p}\in M}(T_{\mathbf{p}}(M))^{\otimes p}$ and $T^0_q(T(M)) = \bigsqcup_{\mathbf{p}\in M}(T^*_{\mathbf{p}}(M))^{\otimes p}$. So in particular, $T^1_0(T(M)) = T(M)$, $T^0_1(T(M)) = T^*(M)$.
  - What about $T^0_0(M)$?
  - A map $A \subseteq T^p_q(T(M)) \to M$ is a *tensor field*. A *smooth tensor field* is a smooth such map. A *(smooth) vector field* is a (smooth) tensor field with $q = 0$, and a *(smooth) covector field* is a (smooth) tensor field with $p = 0$.
  - The subset of $T^0_k(T(M))$ consisting of alternating tensors is denoted $\Lambda^k(T(M))$, and is $\Lambda^k(T(M)) = \bigsqcup_{\mathbf{p}\in M} \Lambda^k(T^*_{\mathbf{p}}(M))$. Is $\binom{n}{k}$ dimensional, and has basis [...]. A *differential k-form on M* is a smooth map $\Lambda^k(T(M)) \to M$. The set of differential k-forms on M is denoted $\Omega^k(M)$.
- Takeaways from tangent vectors section: no way to directly pull back geometric vectors onto a manifold, but can pull back directional derivatives, so we use derivations to formalize tangent vectors, $\frac{\partial}{\partial x^i}$ is a basis for $T_{\mathbf{p}}(M)$, x^i is a coordinate function of a smooth chart
- coords of tangent vector, coords of cotangent vector in analogy to coords of vector and coords of dual vector
- use word tangent vector more

Given a smooth chart $(U, \mathbf{x})$ on M,

- The 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 coordinate vector field*. [Lee, p.176]
- The smooth local frame $\{\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}\}$, where $\frac{\partial}{\partial x^i}$ is the ith coordinate vector field, is called a *coordinate frame* [Lee, p.178]

- $\frac{\partial}{\partial y^i} = \sum_{i=1}^{n} \frac{\partial x^i}{\partial y^i} \frac{\partial}{\partial x^i}$ due to earlier result [Lee, p.275]
- $\{\lambda_1|_{\mathbf{p}}, ..., \lambda_n|_{\mathbf{p}}\}$ denotes the dual basis for $T^*_{\mathbf{p}}(M)$ induced by the basis $\left\{ \frac{\partial}{\partial x^1}\Big|_{\mathbf{p}}, ..., \frac{\partial}{\partial x^1}\Big|_{\mathbf{p}} \right\}$ for $T_{\mathbf{p}}(M)$ [Lee, p.275]
- The covector field defined by $\mathbf{p} \mapsto \lambda_i|_{\mathbf{p}}$ is denoted (in an abuse of notation) by λ_i and is called the *ith coordinate covector field*
- If $\phi_{\mathbf{p}}$ is a covector, $\phi_{\mathbf{p}} \in T_{\mathbf{p}}(M)$, then $([\phi]_{\{\lambda_1|_{\mathbf{p}},...,\lambda_n|_{\mathbf{p}}\}})_i = \phi\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \Big)$ [Lee, p.275]
- The components of a covector field ϕ at $\mathbf{p}$ relative to $\{\lambda_1, ..., \lambda_n\}$ are the covector fields $([\phi]_{\{\lambda_1,...,\lambda_n\}})_i$ defined by $([\phi]_{\{\lambda_1,...,\lambda_n\}})_i|_{\mathbf{p}} = \phi_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \Big)$ [Lee, p.277]
- The smooth (see [Lee, p.278]) coframe $\{\lambda_1, ..., \lambda_n\}$ is called the *coordinate coframe.* [Lee, p.278]
- The differential of a smooth function $f : M \to \mathbb{R}$ is a smooth covector field. [Lee, p.281]
- We have $\lambda_i = d(x^i)$, where x^i is a coord fn and d is its differential in sense of Theorem [...]

$$\boxed{\begin{aligned} df_{\mathbf{p}} &= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}} \\ df &= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx^i \end{aligned}}$$

are the coordinate functions x^i a dual basis in some sense? they are when $M = \mathbb{R}^n$

YES: use $T_{\mathbf{p}}(M) = C^\infty(M \to \mathbb{R})^*$, $T^*_{\mathbf{p}}(M) = C^\infty(M \to \mathbb{R})^{**}$. Use id $V \cong V^{**}$ to formalize this. Also check Reddit comment

13

Differential forms on manifolds

13.1 Differential forms

Definition 13.1. [Lee, p.360] (Alternating $\binom{0}{k}$ tensor field, differential k-form).

We denote the set of $\binom{0}{k}$ tensor fields that yield an alternating $\binom{0}{k}$ tensor at each point by $\Lambda^k(T^*(M))$.

A *differential k-form* is a continuous function $\lambda^k(T^*(M)) \to M$. So, you might say that a differential k-form is a "alternating $\binom{0}{k}$ tensor field" (remember, all tensor fields are continuous maps).

Theorem 13.2. [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, a differential k-form ω can be expressed as

$$\omega = \sum_{i_1 < ... < i_k} f_{i_1 ... i_k} dx^{i_1} \overline{\wedge} ... \overline{\wedge} dx^{i_k},$$

where each $f_{i_1 ... i_k} : U \to \mathbb{R}$.

Theorem 13.3. [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.

13.2 Integration of differential forms on manifolds

This section follows Ch. 16 of [Lee].

There is not always a integrate *functions* on manifold in a coordinate-independent way. For example, if B is the unit ball in $\mathbb{R}^n$ and $f \equiv 1$, then $\int_B f dV = \text{vol}(B)$; using the change of coordinates $\mathbf{x} \mapsto c\mathbf{x}$, we see the integral changes to $c\text{vol}(B)$. (Or $c^n\text{vol}(B)$)?

Differential forms will be the objects whose integrals on manifolds are coordinate independent.

We will consider only compactly supported differential forms; this means the integrals we will consider are analogous to "proper" integrals of functions $f : \mathbb{R}^n \to \mathbb{R}$. Recall, we say that an integral of a smooth function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is *improper* iff D does not contain all of its limit points (so D wouldn't be closed, and thus wouldn't be compact due to Theorem [...]) or is unbounded (so D wouldn't be compact due to Theorem [...]).

[Lee, p.422] We can define a coordinate independent integral of a smooth function $\mathbf{f} : M \to \mathbb{R}$ a Riemannian manifold (M, g): define $\int_M f := \int_M f \omega_g$, where ω_g is the Riemannian volume form induced by g. Riemannian volume form is denoted ds, dA, or dV. (But the d next to s, A, or V does not indicate that ω_g can be interpreted as an exterior derivative!)

In order to define a coordinate-independent notion of integrating differential forms on manifolds, we will first need to compute the pullback of a differential k-form.

Theorem 13.4. [GP74, p.163-165] (Pullback of a differential k-form).

Let M and N be smooth manifolds, and let $\mathbf{f} : M \to N$ be a smooth map. Consider a differential k-form on N, $\omega \in \Omega^k(N)$. In Theorem [...], we showed how to pull back an element of $\Lambda^k(W)$ to obtain an element of $\Lambda^k(V)$. Since a differential form evaluated at a point yields an element of an exterior power, we can pull back differential forms by using the pullback map on exterior powers.

We have $d\mathbf{f} : \Omega^k(M) \to \Omega^k(N)$, $d\mathbf{f}_{\mathbf{x}} : T_{\mathbf{x}}(M) \to T_{\mathbf{x}}(N)$, and $(d\mathbf{f}_{\mathbf{x}})^* : T_{\mathbf{f}(\mathbf{x})}(N)^* \to T_{\mathbf{x}}(M)^*$.

Any id with $(d\mathbf{f}_{\mathbf{x}})$ that needs to be mentioned?

The pullback $\overline{\Omega^k}\mathbf{f}^* : \Omega^k(N) \to \Omega_k(N)$ of a differential k-form ω on N is defined by

$$\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)_{\mathbf{x}} = \Big(\Lambda^k(d\mathbf{f}_{\mathbf{x}})^*\Big)(\omega_{\mathbf{f}(\mathbf{x})}).$$

Repeating the argument made for the pullback of an element of $\Lambda^k(W^*)$ (see [...]), the above becomes

$$\Big(\Lambda^k(d\mathbf{f}_{\mathbf{x}})^*\Big)(\omega_{\mathbf{f}(\mathbf{x})})) = \omega_{\mathbf{f}(\mathbf{x})} \circ d\mathbf{f}_{\mathbf{x}} = (d\mathbf{f}_{\mathbf{x}})^*(\omega_{\mathbf{f}(\mathbf{x})})).$$

Therefore

$$(\overline{\Omega^k}\mathbf{f}^*)(\omega) = \omega_{\mathbf{f}(\mathbf{x})} \circ d\mathbf{f}_{\mathbf{x}}.$$

More explicitly, $(\overline{\Omega^k}\mathbf{f}^*)(\omega)$ is defined at each point $\mathbf{x} \in M$ as

$$\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)_{\mathbf{x}}(\mathbf{v}_1, ..., \mathbf{v}_k) = \omega_{\mathbf{f}(\mathbf{x})})(d\mathbf{f}_{\mathbf{x}}(\mathbf{v}_1), ..., d\mathbf{f}_{\mathbf{x}}(\mathbf{v}_k)).$$

Again referring back to how the definition of $\Lambda^k\mathbf{f}^*$ translates over to $\overline{\Lambda^k}\mathbf{f}^*$, we see that if $\omega(\mathbf{x}) = f(\mathbf{x})\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k$, then the determinant theorem [...] yields

$$\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)_{\mathbf{x}} = \det((d\mathbf{f}_{\mathbf{x}})^*) f(\mathbf{f}(\mathbf{x}))\ (d\mathbf{f}_{\mathbf{x}})^*(\epsilon^1) \overline{\wedge} ... \overline{\wedge} (d\mathbf{f}_{\mathbf{x}})^*(\epsilon^k).$$

Set $\delta^i = (d\mathbf{f}_{\mathbf{x}})^*(\epsilon^i)$, i.e., $\delta^i = \epsilon^i \circ d\mathbf{f}_{\mathbf{x}}$ to restate this as

$$\Big(\omega_{\mathbf{f}(\mathbf{x})}(\omega)\Big)_{\mathbf{x}} = \det((d\mathbf{f}_{\mathbf{x}})^*) f(\mathbf{f}(\mathbf{x}))\ \delta^1 \overline{\wedge} ... \overline{\wedge} \delta^k.$$

Therefore, using that $\det((d\mathbf{f}_{\mathbf{x}})^*) = \det(d\mathbf{f})$ and suppressing dependence on $\mathbf{x}$, we have

$$(\overline{\Omega^k}\mathbf{f}^*)(\omega) = \det(d\mathbf{f})(f \circ \mathbf{f})\ \delta^1 \overline{\wedge} ... \overline{\wedge} \delta^k.$$

Here $d\mathbf{f}$ denotes the function which sends $\mathbf{x} \mapsto d\mathbf{f}_{\mathbf{x}}$.

Using the convention of denoting $\mathbf{f}^* := \overline{\Omega^k}\mathbf{f}^*$, the above is restated as

$$\boxed{\mathbf{f}^*(\omega) = \det(d\mathbf{f})(f \circ \mathbf{f})\ \delta^1 \overline{\wedge} ... \overline{\wedge} \delta^k.}$$

Definition 13.5. [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 13.6. (Support of a function on a manifold).

Let M be a (not necessarily smooth) manifold and $\mathbf{f} : M \to \mathbb{R}^n$. The *support* of $\mathbf{f}$ is defined to be the closure of the set of points where $\mathbf{f}$ is nonzero, $\text{supp}(\mathbf{f}) := \text{cl}(M - \mathbf{f}^{-1}(\mathbf{0}))$. Iff $\text{supp}(\mathbf{f}) \subseteq A$, then we say $\mathbf{f}$ *is supported in* A.

We say $\mathbf{f}$ is *compactly supported* iff $\text{supp}(\mathbf{f})$ is compact.

Can similarly (by extension?) define the support of a differential k-form.

Derivation 13.7. [GP74, p.166] [Lee, p.403] (Change of variables theorem in light of the pullback of a map $M \to N$ and integral of a differential k-form).

The change of variables theorem says that if U and V are open domains of integration in either $\mathbb{R}^n$ or $\mathbb{H}^n$ and $\mathbf{f} : \text{cl}(U) \to \text{cl}(V)$ is a smooth map that restricts to a diffeomorphism $U \to V$, then, for every continuous function $f : \text{cl}(V) \to \mathbb{R}$, we have

$$\int_U f = \int_V (f \circ \mathbf{f}) |\det(d\mathbf{f})|.$$

Notice that, when $|\det(d\mathbf{f})| = \det(d\mathbf{f})$, i.e., when $\det(d\mathbf{f}) > 0$, then the integrand of the right hand side is exactly the pullback of $f \ \delta^1 \overline{\wedge} ... \overline{\wedge} \delta^k$. When $\det(d\mathbf{f}) < 0$, the integrand of the right hand side is the negation of this pullback. So, *in the case that* $\mathbf{f}$ *is orientation-preserving or orientation-reversing*, the change of variables theorem can be restated as

$$\int_U f = \begin{cases} \int_V \mathbf{f}^*(f \ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n) & \mathbf{f} \text{ is orientation-preserving} \\ -\int_V \mathbf{f}^*(f \ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n) & \mathbf{f} \text{ is orientation-reversing} \end{cases}.$$

(It is possible for $\mathbf{f}$ to be neither orientation-preserving nor orientation-reversing. In this case the integral of the pullback over V is likely unrelated to the integral of $\mathbf{f}$ over U).

Therefore, we define the *integral of a compactly supported n-form $\omega = f \ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n$ on a domain of integration D, where $D \subseteq \mathbb{R}^n$ or $D \subseteq \mathbb{H}^n$*, to be

$$\boxed{\int_D \omega = \int_D f \ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n := \int_D f}$$

With this definition, the above restatement of the change of variables theorem is further restated as

$$\boxed{\int_U \omega = \begin{cases} \int_V \mathbf{f}^*(\omega) & \mathbf{f} \text{ is orientation-preserving} \\ -\int_V \mathbf{f}^*(\omega) & \mathbf{f} \text{ is orientation-reversing} \end{cases}}$$

(Recall that earlier we said we would consider compactly supported differential n-forms supported so that we are not considering what would essentially be improper integrals).

The notion of pullback not only simplifies the change of variables theorem. It also 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

$$\int_U f = \int_U f dx^1 ... dx^n,$$

then the definition of the integral of a differential n-form on an open domain of integration becomes

$$\int_U f \ dx_1 \overline{\wedge} ... \overline{\wedge} dx_n := \int_U f dx^1 ... dx^n.$$

In some sense, the placeholder $dx^1 ... dx^n$ is "secretly" $dx_1 \overline{\wedge} ... \overline{\wedge} dx_n$. So, while the definition technically defines the left hand side in terms of the right hand side, you might think of it as giving algebraic meaning to the old placeholder notation of the right hand side.

Remark 13.8. Could discover def of differential forms by starting with the change of variables theorem and trying to contrive the notion of a pullback

Definition 13.9. [Lee, p.404] (Integral of a compactly supported differential n-form over an open subset of $\mathbb{R}^n$ or $\mathbb{H}^n$).

Let U be an open subset of either $\mathbb{R}^n$ 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 containing $\text{supp}(\omega)$. This definition does not depend on the choice of D; see [Lee, p. 403] for the details. The right hand side is interpreted with the definition of the previous derivation.

Theorem 13.10. [Lee, p.404] (Change of variables theorem in light of pullback for open subsets).

The restatement of the change of variables theorem in light of the pullback (stated in Derivation 13.7). We now state essentially the same result, but use open subsets rather than open domains of integration.

Let U and V be open subsets of either $\mathbb{R}^n$ or $\mathbb{H}^n$, and consider a diffeomorphism $\mathbf{f} : U \to V$. If ω is a compactly supported differential n-form on V, then

$$\int_U \omega = \begin{cases} \int_V \mathbf{f}^*(\omega) & \mathbf{f} \text{ is orientation-preserving} \\ -\int_V \mathbf{f}^*(\omega) & \mathbf{f} \text{ is orientation-reversing} \end{cases}$$

Proof. The proof is a matter of using technicalities about diffeomorphisms. See [Lee, p. 404]. □

Definition 13.11. [Lee, p.404] (Integral of a differential n-form that is compactly supported in a single chart over a smooth n-manifold).

Let M be an oriented smooth n-manifold with or without boundary, let $(U, \mathbf{x})$ be a positively oriented or negatively oriented chart on M, and let ω be a differential n-form on M with compact support in U. We define the *integral of ω over M* to be

$$\int_M \omega := \begin{cases} \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega) & (U, \mathbf{x}) \text{ is positively oriented} \\ -\int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega) & (U, \mathbf{x}) \text{ is negatively oriented} \end{cases}.$$

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 details.

Recall from Definition 11.42 that a smooth chart is positively oriented iff ... and ...

Definition 13.12. [Lee, p.43] (Partitions of unity).

If $\{U_\alpha\}$ is an open cover of M and $\{f_\alpha\}$ is a set of functions $M \to \mathbb{R}$, we say that $\{f_\alpha\}$ is a *partition of unity of M (subordinate to $\{U_\alpha\}$)* iff

- $\text{supp}(f_\alpha) \subseteq U_\alpha$ for all α
- $f_\alpha(\mathbf{p})$ is nonzero for only finitely many α, for each $\mathbf{p} \in M$
- $(\sum_\alpha f_\alpha(\mathbf{p}) = 1 \text{ for all } \mathbf{p} \in M) \iff (\sum_\alpha f \text{ is the multiplicative identity of } \{\text{functions } M \to \mathbb{R}\})$.

A *smooth partition of unity* is a partition of unity in which each function is smooth.

Every open cover of a smooth manifold with or without corners admits a smooth partition of unity ([Lee, p. 43]). Since every manifold has an open cover, this means that every smooth manifold with or without corners admits a smooth partition of unity.

Remark 13.13. Partitions of unity are the theoretical way to break up a region.

The in-practice way to use Stokes' thm is "Integral of a differential n-form over parameterizations".

Definition 13.14. [Lee, p.405] (Integral of a differential n-form over a smooth n-manifold).

Let M be a smooth n-manifold with or without boundary and let ω be a compactly supported differential n-form on M. Let $\{(U_i, \mathbf{x}_i)\}_{i=1}^n$ be a finite atlas of M such that $\{U_i\}_{i=1}^n$ is a finite open cover of $\mathrm{supp}(\omega)$ [prove this exists: use Prop 15.6 on p. 382], and let $\{f_i\}_{i=1}^n$ be a smooth partition of unity subordinate to $\{U_i\}_{i=1}^n$. We define the *integral of* ω *over* M to be

$$\int_M \omega := \sum_{i=1}^n \int_M f_i \omega.$$

This definition does not depend on the choice of open cover or partition of unity. See [Lee, 405] for the details.

Theorem 13.15. [Lee, p.407] (Properties of integrals of differential forms).

Theorem 13.16. [Lee, p.408] (Integral of a differential n-form over parameterizations).

Let M be an oriented smooth n-manifold with or without corners and let ω be a compactly supported n-form on M.

Additionally, suppose that we are given adequate parameterizations of the support of ω. That is, suppose that $D_1, ..., D_k$ are open domains of integration in $\mathbb{R}^n$ and that we are given orientation-preserving diffeomorphisms $\mathbf{f}_i : \mathrm{cl}(D_i) \to M$ such that $\{\mathbf{f}_i(D_i)\}$ is a pairwise disjoint collection whose union contains $\mathrm{supp}(\omega)$.

We define *the integral of* ω *over* M *to be*

$$\int_M \omega = \sum_{i=1}^k \int_{\mathbf{f}_i(D_i)} \mathbf{f}_i^*(\omega).$$

Remark 13.17. Reason for the orientation preserving diffeos is because aren't guaranteed that all charts in an atlas will be orientation preserving? need external structure?

Does there always exist a collection of orientation preserving diffeos? if not, defn in terms of partitions of unity makes sense

Remark 13.18. We can make the above definition slightly more general by requiring that $\mathbf{f}_i$: $\mathrm{cl}(D_i) \to M$ be smooth maps whose restrictions $\mathbf{f}_i|_{D_i}$ are orientation-preserving diffeomorphisms.

Proof. As was the case in the definition of the integral of a differential n-form over an n-manifold (Definition [...]), let $\{(U_i, \mathbf{x}_i)\}_{i=1}^n$ be a finite atlas of M such that $\{U_i\}_{i=1}^n$ is a finite open cover of $\mathrm{supp}(\omega)$. We have

$$\int_M \omega = \sum_{i=1}^n \int_M f_i \omega = \int_M$$

The first equality is due to the definition of the integral of a differential n-form over an n-manifold (Definition [...]). The second equality is due to the linearity of this integral.

Then $\omega = \sum_i f_i \omega$. Since ω is compactly supported on M and as $\mathrm{supp}(f_i \omega) \subseteq U_i$ is supported on U_i, then $f_i \omega$ is compactly supported on U_i.

[Lee, p.654] If $D_1, ..., D_k$ are domains of integration whose pairwise intersections have measure zero and whose union is D, then for any continuous bounded function $f : D \to \mathbb{R}$, $\int_D f = \sum_{i=1}^k \int_{D_i} f$. $\square$

Definition 13.19. [HH, p.356] (Dyadic paving).

also, assumption on Riemann sums

Definition 13.20. (Integral of a k-form, $k \leq n$, over an n-manifold)

Use inclusion map

Theorem 13.21. (Stokes' theorem on a smooth chart).

Let M be a smooth n-manifold and consider a smooth chart $(U, \mathbf{x})$ on M, where U has the standard orientation of $\mathbb{R}^k$ [?] and ∂U has the induced boundary [Stokes'?] orientation [shouldn't this be given to $\mathbf{x}(U)$?]. Let ω be a differential $(k-1)$-form with compact support in U of differentiability class C^2. Then

$$\int_U d\omega = \int_{\partial U} \omega.$$

Intuitively, this theorem is true because

$$\int_U d\omega \approx \sum_i d\omega(P_i) \approx \sum_i \int_{\partial P_i} \omega \approx \int_{\partial U} \omega.$$

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 exterior derivatives 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 [...].

Proof. Let $\mathbf{x}(U)$ be an open subset of $\mathbb{H}^k$. ??

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 associative. So, overall, we are showing [...]

($\int_U d\omega \approx \sum_i d\omega(P_i)$). Let $\epsilon > 0$. Then there exists an N large enough such that the dyadic paving of U of fineness 2^{-N} ensures $| \int_U d\omega - \sum_{C_i \in D_N(\mathrm{cl}(\mathbb{H}^k))} d\omega(P_i)| < \epsilon$.

($\sum_i d\omega(P_i) \approx \sum_i \int_{\partial P_i}$). Let $\epsilon > 0$ (forget about the previous ϵ). Take the dyadic decomposition 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 $C_i = P_{\mathbf{x}}(h\hat{\mathfrak{e}}_1, ..., h\hat{\mathfrak{e}}_k)$ for some $\mathbf{x}$, and where $h = 2^{-N}$. By the definition of the exterior derivative (Definition [...]), there exist $K, \delta > 0$ such that when $|h| = 2^{-N} < \delta$, we have $| \int_{C_i} d\omega - \sum_i d\omega(C_i)| < Kh^{k+1}$. By taking N sufficiently large, $|h|$ becomes sufficiently small, and we get $| \int_{C_i} d\omega - \sum_i d\omega(C_i)| < Kh^{k+1} < \epsilon$.

($\sum_i \int_{\partial P_i} \omega \approx \int_{\partial U} \omega$). Let $\epsilon > 0$ (forget about the previous ϵ).

Biggest thing I need to show is

$$\sum_{C \in D_N(\mathrm{cl}(\mathbb{H}^k))} \int_{\partial C} \omega = \sum_{C' \in D_N(\partial \mathrm{cl}(\mathbb{H}^k))} \int_{C'} \omega$$

□

Remark 13.22. The above theorem is the real "heart" of Stokes' theorem. The proof of the full-blown Stokes' theorem just stitches together the application of Stokes' theorem to all charts [in the smooth structure].

[HH, p.543] might be helpful, talks about boundary of a cube

Will also have to use notion of integrating over oriented manifold.

Theorem 13.23. [Lee, p.419] (The generalized Stokes' theorem).

Let M be an oriented smooth manifold with corners, and let ω be a compactly supported smooth [?] $(n-1)$ form on M. Then

$$\int_M d\omega = \int_{\partial M} \omega.$$

Proof. [Lee, p.]

□

Remark 13.24. empty boundary?

pullback of inclusion?

Geometric interpretation of linear functionals

Given a basis E for a finite-dimensional vector space V, a dual vector $\phi \in V^*$ acts on a vector by $\phi(\mathbf{v}) = [\phi]_E \cdot \mathbf{v}$. This is reminiscent of the equation for a plane.

Recall, the plane that contains the point $\mathbf{x}_0 \in V$ and has (not necessarily unit-length) normal $\mathbf{n}$ is the set $\{\mathbf{x} \mid (\mathbf{x} - \mathbf{x}_0) \cdot \mathbf{n} = 0\}$, which is the same set as $\{\mathbf{x} \mid \mathbf{x} \cdot \mathbf{n} = \mathbf{x}_0 \cdot \mathbf{n}\}$. This means that if we consider the level planes $\{\phi(\mathbf{v}) = [\phi]_E \cdot \mathbf{v} = k \mid c \in K\}$, then $[\phi]_E$ takes the role of $\mathbf{x}$, $\mathbf{v}$ takes the role of $\mathbf{n}$, and k takes the role of $\mathbf{x}_0 \cdot \mathbf{n}$ (for some $\mathbf{x}_0$).

Noticing this implies the following facts about $\{\phi(\mathbf{v}) = k \mid c \in K\}$.

1. The level planes are parallel to each other, because their normal vectors are all scalar multiples of $\mathbf{v}$.

2. $[\phi]_E$ is perpendicular to each level plane. (Consider that $[\phi]_E$ is perpendicular to the level plane corresponding to $k = 0$, since $[\phi]_E \cdot \mathbf{v} = 0$. This level plane is parallel to every other level plane, so $[\phi]_E$ is perpendicular to every level plane.)

Now, we construct a set of level planes analogous to a coordinate system. Consider the set of level planes for which $k = 0$, $k = 1$, and $k = 1 \cdot j$ for $j \in \mathbb{Z}$ (the homomorphism $j \mapsto j \cdot 1_F$ relates to ring characteristic!). Then to evaluate $h(\mathbf{v}) = [\phi]_E \cdot \mathbf{v}$, we count the number of these level planes that $[\phi]_E$ pierces.

This interpretation emphasizes thinking of dual vectors as corresponding to row-vectors, since row-vector times column-vector is the same as the dot product between the two corresponding column vectors.

Piponi visualization of diff forms

1-forms are integrated along paths.

To evaluate a linear functional on a vector, think of the linear functional as represented by a set of parallel $(n-1)$-dimensional planes, each a unit distance apart, and count how many planes the vector goes through.

To integrate a differential 1-form over a curve, think of the diff 1-form as represented by curving $(n-1)$-dimensional "leaves." Take the directed count of how many leaves the curve passes through.

A diff k-form on an n-manifold is represented by $(n-k)$-dimensional leaves.

A nonzero exterior product of a diff k form with a diff ℓ form is a $k+\ell$ form because intersecting a collection of $(n-k)$-dimensional leaves with a collection of $(n-\ell)$-dimensional leaves gives a collection of $(n-(k+\ell))$-dimensional leaves.

A differential n-form on an n-manifold is a "top" form whose leaves are points. Such a diff form is a volume density.

To integrate $dx \wedge dy = \epsilon^1 \wedge \epsilon^2$ over an area, draw vertical grid lines (corresponding to ϵ^1), draw horiztonal grid lines (coorresponding to ϵ^2),

Part III

Computational applications of differential forms

14

Discrete differential geometry and computer graphics

14.1 Smoothing with the Laplacian

- (End goal). Solve the problem $\frac{\partial \mathbf{f}}{\partial t} = \Delta \mathbf{f}$. Application to graphics: "smooth out geometric detail on a surface mesh (also known as *fairing* or *curvature flow*)"
- (Theory outline). First subproblem: solve the problem $\Delta \mathbf{f} = \mathbf{g}$, where $\mathbf{f}, \mathbf{g} : M \to \mathbb{R}^3$. (The previous problem uses $\mathbf{g} = \frac{\partial \mathbf{f}}{\partial t}$. Second subproblem: solve the problem $\Delta u = f$, where $u, f : M \to \mathbb{R}$. After we solve the second subproblem, we can use its solution to solve first subproblem, because $\Delta \mathbf{f} = \Delta \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} \Delta f_1 \\ \Delta f_2 \\ \Delta f_3 \end{pmatrix}$. (Define vector of functions notation).
- Two approaches to theory. Both approaches discretize and find approximate solutions. FEM discretizes by only only considering solutions that live in a *discrete approximation space*. DEC discretizes by using the discrete exterior derivative [and it probably discretizes in other ways too- investigate]. Both approaches solve the second subproblem by yielding the "cotan formula", which is an expression for the Laplacian of $u : M \to \mathbb{R}$ at vertex i.
  - (6.2) (FEM). Let $\{\phi_i\}$ be a basis for $\{\text{approx space}\}$. Let f be the true solution (f is not necessarily in the approximation space). The best approximation $u \in \{\text{approx space}\}$ is the $\widetilde{u}$ for which $\langle \Delta \widetilde{u} - f, \phi_i \rangle = 0$ for all i, where $\langle \cdot, \cdot \rangle$ is the L^2 inner product.
  - (6.3) (Discretization via discrete exterior calculus).
    - $\Delta u = f$ is the same as $*d*du = f$.
      - "The Hodge star converts a circulation along the edge e_{ij} into the flux through the corresponding dual edge e^*_{ij}."
    - Something of relevance in this discussion is the same as discretizing the system $(d{*}d)u = *f$. $(d*d)$ is called the conformal Laplacian.

14.1.1 To figure out

- Figure out how Hodge star works, so I can use Stokes' theorem on $\int_\Omega g * df$. probably read the differential operators section
- "Discrete Hodge star" and "dual mesh" (p. 81)

14.2 Surface parameterization

- Let $\mathbf{f} : M \to \mathbb{R}^3$ be a smooth immersion. For smooth $z : M \to \mathbb{C}$, we want to require $\hat{\mathbf{n}} \times d\mathbf{f}(v_\mathbf{p}) = i dz(v_\mathbf{p})$ for all $v_\mathbf{p} \in T_\mathbf{p}(M)$, so that the notion of angle inherited from $\mathbb{R}^3$ by the immersed surface $\mathbf{f}(M) \subseteq \mathbb{R}^3$ is the same as the notion of angle in $\mathbb{C}$. Notice that there must be some $\mathcal{J} : T_{\mathbf{p}}(M) \to T_{\mathbf{p}}(M)$ for which $\hat{\mathbf{n}} \times d\mathbf{f}(v_{\mathbf{p}}) = d\mathbf{f}(\mathcal{J}(v_{\mathbf{p}}))$. Thus, we can express the condition we wanted, "$\hat{\mathbf{n}} \times df(v_{\mathbf{p}}) = idz(v_{\mathbf{p}})$ for all $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$", as "$d\mathbf{f}(\mathcal{J}(v_{\mathbf{p}})) = idz(v_{\mathbf{p}})$ for all $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$". This is the *Cauchy-Riemann equation.*

- Aside: $\omega(v_{\mathbf{p}}, ?) = *\omega(v_{\mathbf{p}}, \mathcal{J}(v_{\mathbf{p}}))$, so $*$ somehow takes the place of $\mathcal{J}$
- Now, the goal is to compute z's that solve the Cauchy-Riemann equation. Leads to minimization problem $\min_z ||*dz - idz||$. We call $E_C(z) = ||*dz - idz||$ the *conformal energy.* We show that for a piecewise linear map z on a simplical disc M, we have $||*dz - idz|| = \frac{1}{2}\langle \Delta z, z\rangle - A(z)$, where $A(z)$ is the signed area of the image $z(M)$.
  - The inner product of complex-valued one forms on M is $\langle \alpha, \beta \rangle = \mathrm{Re} \int_M *\overline{\alpha} \wedge \beta$
    - How does this inner product work for real-valued zero forms?
    - I think I've shown $||a + bi|| = ||a||^2 + 2\mathrm{Re}(\langle a, i\rangle) + ||b||^2$
- Somehow, we decide we need more constraints. So we solve the problem

$$\begin{aligned} &\min_z E_C(z) \\ &\text{s.t. } ||z|| = 1 \text{ for all } z \\ &\quad\ \langle z, \mathbf{1}\rangle = 0 \text{ for all } z \end{aligned}$$

where $E_C(z) = c||*dz - idz|| = c(\frac{1}{2}\langle \Delta z, z\rangle - A(z))$ for some scalar $c \in \mathbb{R}$. If we ignore the third condition, this problem is an eigenvalue problem. (Not obvious that E_C is linear- *is* it?)

Part IV

Appendix

15

Review of pedagogy

Test citation 1 [Lee], test citation 2 [HH], test citation 3 [Cra20], test citation 4 [War]

Bibliography

[BW97] Javier Bonet and Richard D. Wood. *Nonlinear Continuum Mechanics for Finite Element Analysis*. Cambridge Univeristy Press, The address, 1 edition, 1997.

[Cra20] Keenan Crane. *Discrete Differential Geometry: An Applied Introduction*. 1 edition, 4 2020. `https://www.cs.cmu.edu/~kmcrane/Projects/DDG/`.

[GP74] Victor Guillemin and Alan Pollack. *Differential Topology*. Prentice-Hall, 1 edition, 1974.

[HH] Barbara Burke Hubbard and John Hamal Hubbard. *Vector Calculus, Linear Algebra, And Differential Forms*. Prentice Hall, 1 edition.

[Lee] John M. Lee. *Introduction to Smooth Manifolds*. Springer, 2 edition.

[War] Frank W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. 1 edition.