Tensors, differential forms, and computer graphics

Ross Grogan-Kaylor

November 2020

Acknowledgements

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

About this book

Main goals

There are two main goals of this book: (1) generalize vector calculus concepts such as div, grad, curl, the derivative of a vector valued function of a vector variable, and vector calculus integral theorems, and (2) apply these generalizations to the discrete setting of computer graphics. The goals of (1) are accomplished in Chapter 8 with the algebraic machinery of *differential forms* and with a statement of the culminating generalization of vector calculus, the *generalized Stokes' theorem*. Relying on the previous chapters, the goals of (2) are accomplished in the last chapter, Chapter 9. Specifically, in Chapter 9, we "smooth out a surface" in a discrete setting by using the Laplacian operator, and learn how to "unwrap" a manifold in a manner that is as angle-preserving as possible.

Tentative prerequisites and reading advice for a student

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

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

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

Reading advice for reviewers

I recognize that there are multiple sections of this book where skimming will be necessary. Here I outline what I think should be skimmed, and what I think shouldn't.

- Chapter 1 is the review chapter on linear algebra. Obviously, you all know linear algebra, but I would appreciate it if some attention was paid to this chapter, because I spent a lot of work in coming up with original ways to state results from linear algebra in ways I believe are more intuitive than the classical methods. More details on where I have been "original" with regards to linear algebra are at the beginning of the linear algebra chapter.

- Chapter 2, "A motivated introduction to tensors", is relatively short. I hope you enjoy it. I'm especially proud of Section 2.2, "A motivated introduction to $\binom{p}{q}$ tensors".

- I anticipate that Chapter 3, "Bilinear forms, metric tensors, and coordinates of tensors", will contain the most skim-worthy material. More advice is given in that chapter's heading.

- Chapter 4, "Exterior powers, the determinant, and orientation", has some substantial material. Some of it took a long time to write, but likely won't take too long to read, though, so I'm hopeful.

- Chapter 5, the review chapter on calculus, can be skimmed.

- Chapter 6, the chapter on topology, attempts some originality in motivation, so I would appreciate it if the first section or so were read, but once the motivation ends and the statement of standard topological results begins, you can safely stop reading.

- Chapter 7, "Manifolds", presents a lot of cool generalizations of multivariable calculus, but if it's becoming a slog, go ahead and skim.

- Chapters 8 - 9 are where the goals (stated above) are achieved.

I also want to give advance notice of my unconventional approach of striving to treat tensors and differential forms as elements or pointwise elements of tensor product spaces whenever possible. There are times when it is necessary for a differential form to act on tangent vectors, and this is done, but this view of differential forms as "pointwise antisymmetric multilinear elements" is emphasized.

Contents

1

Review of linear algebra

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.

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. 4. 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 100% expected. Covariance and contravariance are explained later).

1.1 Vector spaces, span, and linear independence

Definition 1.1. (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 1.2. (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 1.3. (Vector space over a field).

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 1.4. ($\emptyset$ 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 1.5. (Vector subspace).

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

Span and linear independence

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

Definition 1.6. (Linear combination).

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 1.7. (Span of vectors).

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 1.8. (Linear independence of vectors, intuitive version).

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 1.9. The above definition does not apply to a "list" of just one vector, $\mathbf{v}_1$.

Definition 1.10. (Linear independence of vectors).

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 1.11. 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 1.12. (Equivalence of definitions of linear independence).

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

Bases and dimension of vector spaces

Definition 1.13. (Finite-dimensional vector space).

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 1.14. (Basis, dimension of a vector space).

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 1.15. (Standard basis for K^n).

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 1.16. (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 defined "infinite-dimensional" in terms of bases, 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 1.17. (0-dimensional spaces).

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 1.18. 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 1.19. The statement "every vector space, including infinite-dimensional vector spaces, has a basis" is equivalent to the Axiom of Choice.

Lemma 1.20. (Linear dependence lemma).

Let V and W be vector spaces over a field K. If $\mathbf{v}_1, ..., \mathbf{v}_k \in V$ and $\mathbf{w}_1, ..., \mathbf{w}_\ell \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$ 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 $\mathbf{w}_1, ..., \mathbf{w}_\ell \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$, 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 $\mathbf{w}_1, ..., \mathbf{w}_\ell \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$ 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 $\mathbf{w}_1, ..., \mathbf{w}_\ell \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_{k+1})$, 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 $\mathbf{v}_{k+1} \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_k, \mathbf{w}_\ell)$; explicitly write out the linear combination to see this. Then, since each $\mathbf{w}_i \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_{k+1})$, and $\mathbf{v}_{k+1} \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_k, \mathbf{w}_\ell)$, then each $\mathbf{w}_i \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_k, \mathbf{w}_\ell)$; 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 $\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)$ to be the unique vector $\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2) := c_2\mathbf{v}_2$ for which $\mathbf{v}_1 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$.

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

Claim: $\{\mathbf{w}_i - \text{proj}(\mathbf{w}_i \to \mathbf{w}_\ell) \mid i = 1, ..., \ell - 1\}$ 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 $c_1(\mathbf{w}_1 - \text{proj}(\mathbf{w}_1 \to \mathbf{w}_\ell)) + ... + c_{\ell-1}(\mathbf{w}_{\ell-1} - \text{proj}(\mathbf{w}_{\ell-1} \to \mathbf{w}_\ell)) = \mathbf{0}$. Since, for each i, $\text{proj}(\mathbf{w}_i \to \mathbf{w}_\ell) = q_i \mathbf{w}_\ell$ for some $q_i \in F$, we can distribute the c_i's over the $\mathbf{w}_i - \text{proj}(\mathbf{w}_i \to \mathbf{w}_\ell)$, combine the $c_i q_i \mathbf{w}_\ell$ terms, and obtain an equation of the form $d_1 \mathbf{w}_1 + ... d_\ell \mathbf{w}_\ell + d_{\ell+1} \mathbf{w}_{\ell+1} = \mathbf{0}$ 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 1.21. (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 $\mathbf{v}_i \notin \text{span}(\mathbf{w}_1, ..., \mathbf{w}_\ell)$.

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

Theorem 1.22. (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 1.20), the vectors in F are linearly dependent. This is a contradiction because F is a basis. □

1.2 Linear functions

Definition 1.23. (Linear function).

Let V and W be vector spaces over a field K. 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 $c \in K$,

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

Another equivalent condition is

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

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 1.24. 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 1.25. 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, $\mathbf{f}(\mathbf{v})$. $\square$

Kernel and image of a linear function

Definition 1.26. (Kernel, image of a linear function).

Let V and W be vector spaces, and let $\mathbf{f} : V \to W$ be a linear function. The *kernel* of $\mathbf{f}$ is $\ker(\mathbf{f}) := \mathbf{f}^{-1}(\mathbf{0}) = \{\mathbf{v} \in V \mid \mathbf{f}(\mathbf{v}) = \mathbf{0}\}$. The *image* of V is $\text{im}(\mathbf{f}) := \mathbf{f}(V) = \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V \text{ s.t. } \mathbf{w} = \mathbf{f}(\mathbf{v})\}$.

Definition 1.27. (Rank of a linear function).

Let V and W be vector spaces, and let $\mathbf{f} : V \to W$ be a linear function. The *rank* of $\mathbf{f}$ is defined to be $\dim(\mathbf{f}(V))$, the dimension of the image of $\mathbf{f}$.

Theorem 1.28. (Kernel and image are subspaces).

Let V and W be vector spaces, and let $\mathbf{f} : V \to W$ be a linear function. The kernel of $\mathbf{f}$ is a vector subspace of V and the image of $\mathbf{f}$ is a vector subspace of W. (Proof left as exercise).

Theorem 1.29. (One-to-one linear functions have trivial kernels). $\mathbf{f}$ is one-to-one iff $\mathbf{f}^{-1}(\mathbf{0}) = \{\mathbf{0}\}$.

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

Remark 1.30. 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 1.31. (Main dimension theorem).

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

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 $d_{k+1}\mathbf{e}_{k+1} + ... + d_n\mathbf{e}_n \in \mathbf{f}^{-1}(\mathbf{0}) = \mathrm{span}(\mathbf{e}_1, ..., \mathbf{e}_k)$, 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. □

Linear isomorphisms

Definition 1.32. (Linear isomorphism).

Let V and 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)*.

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

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

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

Lemma 1.33. (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

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

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.

($\Longrightarrow$). Suppose that if $\mathbf{v}_1, ..., \mathbf{v}_k$ are linearly independent, then $\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_n)$ are linearly independent. We need to show $\mathbf{f}$ is one-to-one; it suffices to show that $\mathbf{f}$ has a trivial kernel. Let $\mathbf{v} \in \mathbf{f}^{-1}(\mathbf{0})$, so $\mathbf{f}(\mathbf{v}) = \mathbf{0}$. We want to show $\mathbf{v} = \mathbf{0}$. Since V is finite-dimensional, there is a basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ for V. 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 1.34. (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 1.35. (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 1.36. 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 1.37. (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. Since $\dim(\mathbf{f}(V)) = \dim(W)$, it is straightforwardly shown that $\mathbf{f}(V) \supseteq W$. Then $\mathbf{f}(V) \subseteq W$ and $W \subseteq \mathbf{f}(V)$, so $\mathbf{f}(V) = W$, which means $\mathbf{f}$ is onto, i.e., $\mathbf{f}$ is a linear isomorphism. □

Definition 1.38. (Natural linear isomorphism).

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

Matrices and coordinatization

Coordinates relative to a basis

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

Definition 1.39. (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 1.40. (Linear function acting on a list of vectors).

Consider a linear function $\mathbf{f} : V \to K^m$. We define the notation

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

Note that when $\mathbf{f}$ is invertible and E is a basis of V, then the columns of the matrix $\mathbf{f}(E)$ are a basis of K^m, since invertible linear functions preserve linear independence (see Theorem 1.33).

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

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

Theorem 1.41. 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 1.29): 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 1.42. (Primitive matrix of a linear function $V \to K^m$).

The fundamental idea behind this theorem is the definition of a linear function. Recall from Definition 1.23 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 1.39) 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 1.43. (Matrix of a linear function relative to bases).

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:

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

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 1.40 to define $[\mathbf{f}(E)]_F := [\cdot]_F(E)$. That is,

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

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 1.44. (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 1.45. (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 1.23, 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 1.42.

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 1.42 should produce the output of Derivation 1.42 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 1.52 and 1.53). Upon seeing these index-notation formulas, remember that they are consequences, and are not "just the way things are"!

Remark 1.46. (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 1.40).

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 1.47. (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 1.43 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 1.42), 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

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

Remark 1.48. ($\mathbf{f}_{E,F}$ 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 1.49. (Composition of linear functions that map into K^m and K^p).

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 1.42) is

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

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 1.46). The above then becomes

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

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

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

so that

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

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 1.50. (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 1.51. (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 1.47). 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 1.52. (ith entry of matrix-vector product).

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix with entries in a field K and let $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in K^n$ be a column vector. Referring to the definition of matrix-vector product in Derivation 1.42, 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 $\mathbf{A}$."

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

Theorem 1.53. (ij entry of matrix-matrix product).

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 1.51) 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 1.54. (Change of basis for vectors).

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 1.49), 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 1.40).

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 1.42) 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 1.55. ($(\mathbf{I}_V)_{E,F}^{-1} = (\mathbf{I}_V)_{F,E}$).

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 1.47 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 1.56. (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 1.47). 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 1.51), 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 1.57. (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 1.58. (Change of basis in terms of basis vectors).

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

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

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

Remark 1.59. (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 1.54. This approach for proving Theorem 1.54 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.

1.3 The dot 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 4. In the future, perhaps I will add a discussion of how the cross product relates to the second exterior power of a 3-dimensional vector space (hint: $\dim(\Lambda^2(V)) = \dim(V)$ when V is 3-dimensional, so $\Lambda^2(V) \cong V$); as of now, there's no time for that.

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

Definition 1.60. (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 1.61. (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 1.62. (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 1.63. ($\perp$ operator on $\mathbb{R}^2$).

We define $\perp : \mathbb{R}^2 \to \mathbb{R}^2$ 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 1.64. When we consider $\wedge : \mathbb{R}^2 \to \mathbb{R}^2$ and $\perp : \mathbb{R}^2 \to \mathbb{R}^2$, we have $\perp \circ \wedge = \wedge \circ \perp$, so writing $\hat{\mathbf{v}}_\perp$ is unambiguous when $\mathbf{v} \in \mathbb{R}^2$.

Definition 1.65. (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 1.66. Rotations are orthogonal linear functions.

Definition 1.67. (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 1.68. 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 1.69. (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 1.70. (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 $||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = ||\mathbf{v}_1|| \cos(\theta)$. $\square$

Definition 1.71. (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 1.89!). 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 1.72. 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 1.73. (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 1.74. (The geometric dot product on $\mathbb{R}^2$ is a bilinear function).

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

Proof. The geometric dot product is symmetric, so it suffices to show that it is a linear function in either argument; it suffices to show that $\mathbf{f} : V \to K$ 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 1.75. 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 1.76. 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 1.77. (Algebraic dot product on $\mathbb{R}^2$).

We can now derive an "algebraic" formula for the dot product, using its bilinearity (Theorem 1.74) 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 1.78. (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 $\mathbf{f}$. $\square$

Theorem 1.79. (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 1.69.
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 $\mathbb{R}^2$. $\square$

Remark 1.80. 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 1.87.

Definition 1.81. (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 1.82. (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 an exercise. $\square$

Theorem 1.83. (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 1.84. (Angle in $\mathbb{R}^n$).

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

The angle $\theta := \cos^{-1}\left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_1|| ||\mathbf{v}_2||}\right)$ does indeed satisfy the condition $\theta \in [0, 2\pi)$. In Theorem 3.16, we prove this in a more general setting.

Remark 1.85. (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 1.86. (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 1.87. (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 1.80.

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 1.88. 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 1.89. (Vector projection in terms of algebraic dot product).

Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$. We can express the projection of $\mathbf{v}_1$ onto $\mathbf{v}_2$ in terms of the dot product $\mathbf{v}_1 \cdot \mathbf{v}_2$. (Recall Definition 1.67 for the definition of vector projection).

Recall from Lemma 1.70 that $||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = ||\mathbf{v}_1|| \cos(\theta)$. Since $\mathbf{v}_1 \cdot \mathbf{v}_2 = ||\mathbf{v}_1|| ||\mathbf{v}_2|| \cos(\theta)$, then $||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_2||}$. Thus

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

Theorem 1.90. (Vector projection in $\mathbb{R}^n$ is a bilinear function).

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

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

In general, vector projection in any finite-dimensional vector space *with inner product* is a bilinear function, because an inner product is by definition a bilinear function. We will discuss inner products and vector spaces with inner product in Chapter 3. □

Part I

Multilinear algebra and tensors

2

A motivated introduction to tensors

In this chapter, we introduce the idea of a "tensors", since tensors underpin differential forms. There are two key ideas that we must formalize before we define what a "tensor" is.

One of the ideas is that of a "multilinear element". Recall that elements of vector spaces (vectors) can be thought of as "linear elements" 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".

2.1 Multilinear functions and tensor product spaces

Definition 2.1. (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 2.2. (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 2.3. (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 1.23). 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 2.4. (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 2.5. (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 2.28). 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 2.6. (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 2.7. (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 1.32 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 1.38).

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 2.8. (Basis and dimension of a tensor product space).

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

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

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

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

is a basis of $V \otimes W$.

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

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

by the seeming-bilinearity of $\otimes$.

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

Theorem 2.9. (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 $\mathbf{f}$, and 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. □

---

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

Theorem 2.10. (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 2.9. We already know this map is an injection because $\mathbf{h}$ is uniquely determined by $\mathbf{f}$ (see the proof of Theorem 2.9). 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 3.41). 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. $\square$

2.2 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 1.53). 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 1.49 and possibly Theorem 1.51). 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 1.27).

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 1.42)) 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 2.11. 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, $\mathbf{f}(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{f}(\mathbf{v}_1) + \mathbf{f}(\mathbf{v}_2)$. $\square$

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 2.12. (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:

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

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 $\mathbf{w}_i \otimes \phi_j \in W \otimes V^*$:

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

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

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

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

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

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

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

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

As was alluded to before, the above is achieved due to the bilinearity of $\circ$ and the seeming-bilinearity[2] of $\otimes$:

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

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

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 2.13. ($\mathcal{L}(V \to W) \cong W \otimes V^*$ 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 1.38).

Remark 2.14. (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 2.15. (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 2.16. (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 $\otimes$.

2.3 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 2.17. (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 2.18. (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 1.42 and Remark 1.46).

Recall from Definition 1.40 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 $\hat{\mathfrak{e}}_i^\top$:

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

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 2.19. (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 $\phi_{\mathbf{e}_i}$. □

Remark 2.20. (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 1.38) 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 2.21. (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 2.18 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 2.22. (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.

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

Theorem 2.23. ($V \cong V^{**}$ 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 2.18. 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 2.24. (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 2.25. (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 2.26. (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 2.24, 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 2.27. (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.

Proof. Left as an exercise. □

2.4 $\binom{p}{q}$ tensors

Definition 2.28. ($\binom{p}{q}$ 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 $T^p_q(V)$ to denote the vector space of $\binom{p}{q}$ tensors on V.

Definition 2.29. (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 $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^*. The *coordinates of a* $\binom{p}{q}$ *tensor* $T \in T^p_q(V)$ *relative to* E *and* E^* are the scalars $T^{i_1 ... i_p}_{j_1 ... j_q} \in K$ for which

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

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

Remark 2.30. (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 3.39). 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 2.33), 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[3] 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 $T^2_0(V) \cong T^1_1(V) \cong T^0_2(V)$ due to the (unnatural) isomorphism $V \cong V^*$ that is obtained by choosing a basis (see Remark 2.20).

Definition 2.31. (Covariance and contravariance, coordinates of a $\binom{p}{q}$ tensor).

Let V be a vector space. For reasons that will be explained later, in Remark 3.37, 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 $c^1 \mathbf{v}_1 + ... + c^n \mathbf{v}_n$.

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

Definition 2.32. (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 2.33. (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

$$\boxed{T^p_q(V) \cong T^q_p(V^*)}$$

Proof. The first line in the first box is Theorem 2.10, and the second line in the first box is Theorem 2.13. 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. $\square$

Remark 2.34. (The four-fold nature of $\binom{p}{q}$ tensors).

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

$$\begin{tikzcd}
\substack{\text{multilinear element} \\ \text{(element of a tensor product space)}} \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \text{multilinear function} \arrow[d, leftrightarrow] \\
\substack{\text{linear element} \\ \text{(element of a vector space)}} \arrow[r, leftrightarrow] & \text{linear function}
\end{tikzcd}$$

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

Theorem 2.35. (Other useful natural isomorphisms for $\binom{p}{q}$ tensors).

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

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

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

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

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

We have $\sum_{ij} a^j_i \phi^{\mathbf{e}_i} \otimes \mathbf{f}_j = \sum_{ij} \phi^{\mathbf{e}_i} \otimes \mathbf{f}(\mathbf{e}_i)$. The explicit check that this is a linear isomorphism is left to the reader. It is enough to show linearity and surjectivity because $\mathcal{L}(V \to W) \cong W \otimes V^*$ is a finite-dimensional vector space; injectivity automatically follows due to Theorem 1.37.

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

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 1.37. □

3

Bilinear forms, metric tensors, and coordinates of tensors

The goal of this chapter is to present results regarding the coordinates of $\binom{p}{q}$ tensors relative to bases. This is accomplished in the second part of this chapter. The first subsection of the second part, "Coordinates with a metric tensor", is the most important part of this entire chapter. Particularly important is Theorem 3.26, which describes how vectors and dual vectors can act on each other to produce each other's coordinates. The subsequent subsections of the second part of this chapter are less important, but still interesting. In these, we show how to change the bases of a $\binom{p}{q}$ tensor and how composition of linear functions generalizes to *tensor contraction*. Of least (direct) importance is the discussion of the convention of *slanted indices*; we have included this because it is handy to know of when exploring literature on tensors.

We build up to these ideas about coordinates by investigating *bilinear forms* in the first part of the chapter. A special type of bilinear form with which the reader may be familiar is an *inner product*; we will define and investigate various facts about these. If you wish to be economical about reading the first part of this chapter, you can skip the middle two sections ("Inner products" and "Symmetric and orthogonal linear functions"); those sections are included for completeness.

3.1 Bilinear forms and metric tensors

Definition 3.1. (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 3.2. (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 2.33, we have $\mathcal{L}(V^{\times k} \to K) \cong \mathcal{L}(V^{\otimes k} \to K) = (V^{\otimes k})^* \cong (V^*)^{\otimes k} = T_k^0(V)$. Therefore a linear k-form is naturally identified with a $\binom{0}{k}$ tensor.

Definition 3.3. (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 1.37).

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 1.38). 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 3.4. (The adjoint of a linear function).

Let V and W be finite-dimensional vector spaces, let B be a nondegenerate bilinear form on V and W, and consider the musical isomorphisms $\flat_1 : V \to W^*$ 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 3.5. (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 3.27.

Definition 3.6. (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 3.7. (The notation $\flat$ and $\sharp$).

When V is a finite-dimensional and there is a metric tensor g on V, then the musical isomorphisms $\flat_1 : V \to V^*$ and $\flat_2 : V \to V^*$ induced by g are the same because g is symmetric. This leads us to define $\flat := \flat_1 = \flat_2$ and $\sharp := \sharp_1 = \flat_1^{-1} = \sharp_2 = \flat_2^{-1}$.

Inner products

Definition 3.8. (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 $g(\mathbf{v}, \mathbf{v}) \geq 0$ for all $\mathbf{v} \in V$, with $g(\mathbf{v}, \mathbf{v}) = \mathbf{0}$ iff $\mathbf{v} = \mathbf{0}$. (We automatically have the reverse implication for any bilinear form g for the reasons discussed in Definition 3.3).

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 3.9. (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 3.10. (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 3.11. The dot product on $\mathbb{R}^n$ is an inner product on $\mathbb{R}^n$. (Proof left as exercise).

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

Length and orthogonality with respect to an inner product

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

Definition 3.12. (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 1.83), 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 3.13. (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 1.84), 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 3.14. (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 3.4). The condition on $\langle \cdot, \cdot \rangle$ induced by using the musical isomorphism $\flat : V \to V^*$ to identify $V \cong V^*$ (see the end of Definition 3.4) 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 3.15. (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 1.85).

Theorem 3.16. (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 3.17. (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 3.18. (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 3.19. (Gram-Schmidt algorithm).

Let V be an inner product space. Given any basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ for V, we can use the following *Gram-Schmidt algorithm* to convert E into an orthonormal basis $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$.

First, we "orthogonalize" the basis E into a basis G. Set $\mathbf{g}_1 := \mathbf{e}_1$, and, for $i \geq 2$, set

$$\mathbf{g}_i := \mathbf{e}_i - \text{proj}(\mathbf{f}_i \to \text{span}(\mathbf{e}_1, ..., \mathbf{e}_{i-1})) = \mathbf{e}_i - \sum_{j=1}^{n} \text{proj}(\mathbf{g}_i \to \mathbf{e}_j) = \mathbf{e}_i - \sum_{j=1}^{n} \frac{\langle \mathbf{g}_i, \mathbf{e}_j \rangle}{\langle \mathbf{e}_j, \mathbf{e}_j \rangle}, \quad i \geq 2.$$

(In the last equality in the line above, we've used an analogue of Theorem 1.89 to express vector projections in terms of inner products). To obtain the orthonormal basis $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$, we just normalize the orthogonal basis G, and set $\mathbf{f}_i := \frac{\mathbf{g}_i}{||\mathbf{g}_i||}$.

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 3.22 satisfied by an orthogonal linear function.

Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$. Consider a linear function $\mathbf{f} : V \to V$, and let $\mathbf{f}^* : V \to V$ be its adjoint (see Definition 3.4). The defining condition of the adjoint (see the end of Definition 3.4) 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 3.20. (Transpose of a matrix).

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

Definition 3.21. (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 $\widehat{U}$ 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 $\widehat{U}$, $[\mathbf{f}(\widehat{U})]_{\widehat{U}} = [\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$.

   (b) Moreover, the matrix of the adjoint $\mathbf{f}^* : V \to V$ relative to $\widehat{U}$ is $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$.

Proof.
($1 \iff 2$).

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

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

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

($\impliedby$). Let $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ be an orthonormal basis for V, and let the matrix $\mathbf{A}$ of $\mathbf{f}$ relative to $\widehat{U}$ satisfy $a_{ij} = a_{ji}$. Since $a_{ij} = \langle \mathbf{f}(\hat{\mathbf{u}}_i), \hat{\mathbf{u}}_j \rangle$, then $\langle \mathbf{f}(\hat{\mathbf{u}}_i), \hat{\mathbf{u}}_j \rangle = \langle \hat{\mathbf{u}}_i, \mathbf{f}(\hat{\mathbf{u}}_j) \rangle$. Extend with multilinearity to obtain the conclusion.
(3b). The jth column of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$ is $[\mathbf{f}(\hat{\mathbf{u}}_j)]_{\widehat{U}}$, so the ij entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$ is $\langle \mathbf{f}(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle$. Similarly, the ij entry of the matrix of $\mathbf{f}^*$ relative to $\widehat{U}$ is $\langle \mathbf{f}^*(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle$. Due to the condition on $\langle \cdot, \cdot \rangle$ induced by the identification $V \cong V^*$, we have $\langle \mathbf{f}^*(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle = \langle \hat{\mathbf{u}}_j, \mathbf{f}(\hat{\mathbf{u}}_i) \rangle = \langle \mathbf{f}^*(\hat{\mathbf{u}}_i), \hat{\mathbf{u}}_j \rangle$. But this is the ji entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$, i.e., the ij entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$. Thus $[\mathbf{f}(\widehat{U})]_{\widehat{U}} = [\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$. □

Definition 3.22. (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 $\widehat{U}$ is an orthonormal basis of V, then the matrix $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$ of $\mathbf{f}$ relative to $\widehat{U}$ has orthonormal columns.
7. If V is finite-dimensional and $\widehat{U}$ is an orthonormal basis of V, then $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top [\mathbf{f}(\widehat{U})]_{\widehat{U}} = \mathbf{I}$ and $[\mathbf{f}(\widehat{U})]_{\widehat{U}} [\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top = \mathbf{I}$, so $[\mathbf{f}(\widehat{U})]_{\widehat{U}}^{-1} = [\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$.

[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.

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 1.78 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 $\widehat{U}$ is orthonormal, $\langle \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j \rangle = \delta^i_j$. Therefore $\langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \delta^i_j$, so the columns $[\mathbf{f}(\hat{\mathbf{u}}_i)]_{\widehat{U}}$ of the matrix of $\mathbf{f}$ relative to $\widehat{U}$ are orthonormal.

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

(6 $\iff$ 7).

($\implies$). The ij entry of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}[\mathbf{f}(\widehat{U})]_{\widehat{U}}^\top$ is (ith row of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$) $\cdot$ (jth column of $[\mathbf{f}(\widehat{U})]_{\widehat{U}})^\top$ = (ith row of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$) $\cdot$ (jth row of $[\mathbf{f}(\widehat{U})]_{\widehat{U}}$) $= \langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \delta^i_j$.

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

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

Self-duality and choice of basis

Definition 3.23. (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 $\flat : V \to V^*$ induced by B sends $\mathbf{e}_i \overset{\flat}{\mapsto} \phi^{\mathbf{e}_i}$ for all i.

Theorem 3.24. (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 $\mathbf{F} = \flat : V \to V^*$, where $\flat$ is the musical isomorphism induced by the inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}$ defined by $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = [\mathbf{v}_1]_E \cdot [\mathbf{v}_2]_E$, and where $\cdot$ is the dot product on $\mathbb{R}^n$. (It is left to the reader to check that this is indeed an inner product).

This theorem sheds light on the "unnatural" isomorphism $V \to V^*$ which sends $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ that was discussed in Remark 2.20). We said that the isomorphism which sends $\mathbf{e}_i \mapsto \phi^{\mathbf{e}_i}$ is not natural because it depends on a choice of basis. (This is still true). Recently, we have seen that a nondegenerate bilinear form, such as an inner product, on V induces the natural musical isomorphism $\flat : V \to V^*$. 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 $\flat : V \to V^*$ 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 $\phi^{\mathbf{e}_i}(\mathbf{v}) = [\mathbf{e}_i]_E^\top [\mathbf{v}]_E = [\mathbf{e}_i]_E \cdot [\mathbf{v}]_E = \langle \mathbf{e}_i, \mathbf{v} \rangle = \mathbf{e}_i^\flat(\mathbf{v})$. Thus $\phi^{\mathbf{e}_i}(\mathbf{v}) = \mathbf{e}_i^\flat(\mathbf{v})$, so $\phi^{\mathbf{e}_i} = \mathbf{e}_i^\flat$. $\square$

Theorem 3.25. (Self-dual $\iff$ orthonormal).

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

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

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

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

Coordinates with a metric tensor

Theorem 3.26. (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 2.23 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 2.23 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 2.23), 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 3.27. (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 3.5). 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 3.32.

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:

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

□

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

Lemma 3.28. (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 1.46) 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 3.29. 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 1.42), 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 3.27 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 3.5). 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 3.27. 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. $\square$

Theorem 3.30. (Relationship between coordinates of vectors and dual vectors).

Consider the hypotheses of Theorem 3.27. 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 3.31. (The special case $g_{ij} = \delta_{ij}$).

When $V = W$, g is an inner product on V, and $E = \widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ is an orthonormal basis (relative to the inner product) for V, then $g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j) = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$.

Then the converse is true as well (so long as $V = W$). If $g_{ij} = \delta_{ij}$, then g is an inner product on V, and, since $g(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) = g_{ij} = \delta_{ij}$ for some basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ of V, then that basis is an orthonormal basis of V.

Remark 3.32. (Lower indices in g_{ij} and upper indices in g^{ij}).

We use lower indices in the notation $g_{ij} := g(\mathbf{e}_i, \mathbf{e}_j)$ because, if g is a metric tensor on V, then g is an element of $\mathcal{L}(V \times V \to K)$ and can thus be identified with a $\binom{0}{2}$ tensor on V (see Remark 3.2). We use lower indices for $g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j)$ to follow the index conventions of Definition 2.31.

What about the upper indices in g^{ij}? Suppose again that g is a metric tensor on V. Just as g can be identified with a $\binom{0}{2}$ tensor on V, the induced bilinear form $\widetilde{g}$ on V^* can be identified with a $\binom{2}{0}$ tensor on V. (An element of $\mathcal{L}(V^* \times V^* \to K)$ can be identified with a $\binom{0}{2}$ tensor on V^*, which can be identified with a $\binom{2}{0}$ tensor on V). Therefore, since $\mathbf{g}^{-1}$ stores the coordinates of a $\binom{2}{0}$ tensor, we use upper indices to denote the ij entry of $\mathbf{g}^{-1}$.

Remark 3.33. (δ^{ij} vs. δ^i_j vs. δ^i_j).

One may be confused when they consider that there are three ways to write the Kronecker delta: δ^{ij}, δ^i_j, and δ_{ij}. The difference between these functions is straightforward: δ^{ij} is used for coordinates of a $\binom{2}{0}$ tensor, δ^i_j is used for a coordinates of a $\binom{1}{1}$ tensor, and δ_{ij} is used for coordinates of a $\binom{0}{2}$ tensor. All three functions are defined, as usual, to be 1 when $i = j$ and 0 otherwise.

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

Definition 3.34. (Coordinates of a $\binom{p}{q}$ tensor).

For convenience, we now restate Definition 2.29.

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let E^* be a basis for V^*. (Note that we have not assumed that E^* is induced by E). The *coordinates of a* $\binom{p}{q}$ *tensor* $T \in T^p_q(V)$ *relative to* E *and* E^* are the scalars $T^{i_1 \dots i_p}_{j_1 \dots j_q}$ for which

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

Lemma 3.35. (Lemma for change of basis for vectors and dual vectors).

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

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

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 3.26), we have

$$\begin{aligned} ([\phi]_{F^*})_i &= \phi(\mathbf{f}_i) = \phi\Big(\sum_{j=1}^n ([\mathbf{f}_i]_E)_j \mathbf{e}_j \Big) = \sum_{j=1}^n ([\mathbf{f}_i]_E)_j \phi(\mathbf{e}_j) = \sum_{j=1}^n ([\mathbf{f}_i]_E)_j ([\phi]_{E^*})_i \\ &= \Big(([\phi]_{E^*})_1 \quad \dots \quad ([\phi]_{E^*})_n \Big) [\mathbf{f}_i]_E = [\phi]^\top_{E^*} [\mathbf{f}_i]_E. \end{aligned}$$

□

Theorem 3.36. (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

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

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

Proof. The first line of the boxed equation is Theorem 1.54, and the third line is Theorem 1.55. We will prove $[\mathbf{E}^*]_{F^*} = [\mathbf{E}]^{-\top}_F$, which is the fourth line. The second line then follows by applying the fourth line to the equation $[\phi]_{F^*} = [\mathbf{E}^*]_{F^*} [\mathbf{v}]_{E^*}$ (this last equation is implied by the first line).

As usual, set $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$, and $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$. We show $[\mathbf{E}^*]_{F^*} = [\mathbf{E}]^{-\top}_F$ by examining the ij entry of $[\mathbf{E}^*]_{F^*}$ and making use of the previous lemma. We have

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

□

Remark 3.37. (What covariance and contravariance *really* mean).

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

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

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

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

Theorem 3.38. (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 $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_n}\}$ be the corresponding induced dual bases for V^*. We have

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

Proof. The first line in the boxed equation follows from the definition of $[\cdot]_F$ (and was stated earlier in Theorem 1.58). 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 $\delta^i = \sum_{j=1}^n ([\mathbf{F}^*]_{E^*})^j_i \phi^{\mathbf{e}_j}$, where we have $[\mathbf{F}^*]_{E^*} = [\mathbf{F}]_E^{-\top}$ due to the previous theorem. □

Theorem 3.39. (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 $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ and $F^* = \{\psi^{\mathbf{f}_1}, ..., \psi^{\mathbf{f}_n}\}$ be the corresponding induced dual bases for V^*.

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

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

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

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

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

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

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

(It is possible to 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 $\binom{i_1 \dots i_p}{j_1 \dots j_q}$ component of the previously mentioned $\binom{p}{q}$ tensor relative to E and E^* as

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

Remark 3.40. (Tensors as "multidimensional matrices" that "transform like tensors").

As was mentioned in Remark 2.30, physicists often define tensors to be "multidimensional matrices" that follow the change of basis formula of the previous theorem.

Tensor contraction

Derivation 3.41. (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 this as an exercise!). Also recall from Section 2.2 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 2.2 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 2.2 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 3.42. (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:

$$\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_p \otimes \phi^1 \otimes \dots \otimes \phi^q$$

$$\overset{\binom{k}{\ell} \text{ contraction}}{\longmapsto}$$

$$C(\mathbf{v}_k, \phi^\ell)(\mathbf{v}_1 \otimes \dots \otimes \cancel{\mathbf{v}_k} \otimes \dots \mathbf{v}_p \otimes \phi^1 \otimes \dots \otimes \cancel{\phi^\ell} \otimes \dots \otimes \phi^q)$$

$$=$$

$$\phi^\ell(\mathbf{v}_k)(\mathbf{v}_1 \otimes \dots \otimes \cancel{\mathbf{v}_k} \otimes \dots \mathbf{v}_p \otimes \phi^1 \otimes \dots \otimes \cancel{\phi^\ell} \otimes \dots \otimes \phi^q).$$

Remark 3.43. (Contraction with upper and lower indices).

The convention we laid out in 2.31 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 3.44. (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 2.33, 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 3.41, 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 3.45. (Coordinates of a contracted tensor).

Let V be an n-dimensional vector space, let E be a basis for V, and let E^* be a basis for V^*. Consider a $\binom{p}{q}$ tensor $\mathbf{T} \in T^p_q(V)$. If the $\binom{i_1 \ldots i_p}{j_1 \ldots j_q}$ coordinate of $\mathbf{T}$ relative to E and E^* is $T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$, then the $\binom{i_1 \ldots i_{p-1}}{j_1 \ldots j_{q-1}}$ component of the $\binom{k}{\ell}$ contraction of $\mathbf{T}$ relative to E and E^* is $\sum_{r=1}^n T^{i_1 \ldots i_{k-1}\ r\ i_k \ldots i_{p-1}}_{j_1 \ldots j_{\ell-1}\ r\ j_\ell \ldots j_{q-1}}$.

Proof. Assume $\mathbf{T}$ has $\binom{i_1 \ldots i_p}{j_1 \ldots j_q}$ component $T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$, so $\mathbf{T} = \sum_{I_p, J_q} T^{I_p}_{J_q} \mathbf{e}^{I_p} \otimes \phi_{J_q}$, where $I_p = \{i_1, ..., i_p\}$, and $J_q = \{j_1, ..., j_q\}$. Then, using linearity of the contraction map, the $\binom{k}{\ell}$ contraction of $\mathbf{T}$ is

$$\sum_{I_p, J_q} \epsilon_{j_\ell}(\mathbf{e}^{i_k}) T^{I_p}_{J_q} \mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}},$$

where $I_p - \{i_k\} = \{i_1, ..., \cancel{i_k}, ..., i_p\}$, $J_q - \{j_\ell\} = \{j_1, ..., \cancel{j_\ell}, ..., j_p\}$.

Because $\epsilon_{j_\ell}(\mathbf{e}^{i_k}) = \delta_{i_k j_\ell}$, the above sum reduces to

$$\sum_{I_p - \{i_k\}, J_q - \{j_\ell\}} \sum_r \Big(T^{I'_p}_{J'_q} \mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}} \Big),$$

where $I'_p = \{i_1, ..., i_{k-1}, r, i_{k+1}, ..., i_p\}$, $J'_p = \{j_1, ..., j_{\ell-1}, r, j_{\ell+1}, ..., j_p\}$.

Since $\mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}}$ does not vary with r, this is the same as

$$\begin{aligned} &\sum_{I_p - \{i_k\}, J_q - \{j_\ell\}} \Big(\mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}} \sum_r T^{I'_p}_{J'_q} \Big) \\ = &\sum_{I_p - \{i_k\}, J_q - \{j_\ell\}} \Big(\sum_r T^{I'_p}_{J'_q} \Big) \mathbf{e}^{I_p - \{i_k\}} \otimes \phi_{J_q - \{j_\ell\}} \end{aligned}$$

So, we can see that the $\binom{k}{\ell}$ contraction of T is a $\binom{p-1}{q-1}$ tensor, with a $\binom{I'_p}{J'_q}$ component of $\sum_r T^{I'_p}_{J'_q}$. In other words, the $\binom{i_1 \ldots i_{k-1} i_{k+1} \ldots i_p}{j_1 \ldots j_{\ell-1} j_{\ell+1} \ldots j_p}$ component of the $\binom{k}{\ell}$ contraction of T is

$$\sum_r T^{i_1 \ldots i_{k-1}\ r\ i_{k+1} \ldots i_p}_{j_1 \ldots j_{\ell-1}\ r\ j_{\ell+1} \ldots j_q}.$$

Equivalently, after shifting the subindices of $i_{k+1}, ..., i_p$ down by one (which is valid because the subindices k, ℓ of i, j are no longer "occupied") we see the $\binom{i_1 \ldots i_{p-1}}{j_1 \ldots j_{q-1}}$ component of the $\binom{k}{\ell}$ contraction of T is

$$\sum_r T^{i_1 \ldots i_{k-1}\ r\ i_k \ldots i_{p-1}}_{j_1 \ldots j_{\ell-1}\ r\ j_\ell \ldots j_{q-1}}.$$

$\square$

Theorem 3.46. Taking any $\binom{k}{\ell}$ contraction is basis-independent.

Proof. This follows from the fact that the natural bilinear form C on V and V^* defined by $C(\mathbf{v}, \phi) = \phi(\mathbf{v})$ is basis-independent, as no bases are required to define C. $\square$

Theorem 3.47. (The trace is the $\binom{1}{1}$ contraction of a $\binom{1}{1}$ tensor).

Let V be a finite-dimensional vector space over a field K.

The *trace* of a square matrix (a^i_j) with entries in K is defined to be the sum of the matrix's diagonal entries: $\mathrm{tr}(a^i_j) := \sum_{i=1}^n a^i_i$. We have that $\mathrm{tr}(a^i_j)$ is the $\binom{1}{1}$ contraction of the $\binom{1}{1}$ tensor corresponding to (a^i_j).

Thus, we see the trace is a special case of tensor contraction.

Proof. Recall from Theorem 2.36 that when (a^i_j) is interpreted to be the matrix relative to bases of a linear function $V \to V$, then (a^i_j) is identified with the $\binom{1}{1}$ tensor $\sum_{ij} a^j_i \phi^{\mathbf{e}_j} \otimes \mathbf{e}^i$, where $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ is the basis for V^* induced by the basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ for V. The $\binom{1}{1}$ contraction of this $\binom{1}{1}$ tensor is $\sum_{ij} a^j_i \phi^{\mathbf{e}_j}(\mathbf{e}^i) = \sum_{ij} a^j_i \delta^j_i = \sum_i a^i_i = \mathrm{tr}(a^i_j)$. $\square$

Slanted indices

Theorem 3.48. Correspondence between multidimensional matrices and $\binom{p}{q}$ tensors

Theorem 3.49. (Converting vectors to dual vectors and dual vectors to vectors in a $\binom{p}{q}$ tensor).

Let V be a finite-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\epsilon^1, ..., \epsilon^n\}$ be a basis for V^*.

We can "convert" a $\binom{p}{q}$ tensor $\mathbf{T} \in T^p_q(V)$ to a $\binom{p-1}{q+1}$ tensor as follows:

$$\begin{gathered}\mathbf{T} = \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_k} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \\ \longmapsto \\ \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}^{\flat_1}_{i_k} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q}.\end{gathered}$$

Using Theorem 3.30, we compute $\mathbf{e}^{\flat_1}_{i_k}$ to be

$$\mathbf{e}^{\flat_1}_{i_k} = \sum_{r=1}^n ([\mathbf{e}^{\flat_1}_{i_k}]_{E^*})_r \epsilon^r = \sum_{r=1}^n \Big(\sum_{j=1}^n g_{rj}([\mathbf{e}_{i_k}]_E)^j \Big) \epsilon^r = \sum_{r=1}^n \Big(\sum_{j=1}^n g_{rj} \delta^j_{i_k} \Big) \epsilon^r = \sum_{r=1}^n g_{i_k r} \epsilon^r,$$

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 the $\binom{p}{q}$ tensor $\mathbf{T}$ to a $\binom{p-1}{q+1}$ tensor as follows:

$$\mathbf{T} = \mathbf{e}_{i_1} \otimes ... \otimes \mathbf{e}_{i_k} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q}$$
$$\longmapsto$$
$$\sum_{r=1}^{n} g_{i_k r} \Big(\mathbf{e}_{i_1} \otimes ... \otimes \cancel{\mathbf{e}_{i_k}} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{r} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \Big).$$

If the coordinates of $\mathbf{T}$ relative to E and E^* were originally $T^{i_1 \dots i_p}_{j_1 \dots j_q}$, then they get sent to $\sum_{r=1}^{n} g_{i_k r} T^{i_1 \dots \cancel{i_k} \dots i_p}_{r j_1 \dots \dots \dots j_q}$. 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 $\mathbf{T}$ relative to E and E^* were originally $T^{i_1 \dots i_p}_{j_1 \dots j_q}$, then they get sent to $\sum_{r=1}^{n} g^{i_k r} T^{r i_1 \dots \dots \dots i_p}_{j_1 \dots \cancel{j_\ell} \dots j_q}$.

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).

4

Exterior powers, the determinant, and orientation

This chapter focuses on *antisymmetric tensors*. We need to know about these because differential forms, when evaluated a point, are antisymmetric tensors. Antisymmetric tensors are also closely related to the determinant; we will define the determinant and explore this relationship in the second part of this chapter. We will also show how to use antisymmetric tensors- or, more specifically, elements of a *top exterior power*- to give *orientation* to a finite dimensional vector space. Lastly, we investigate *pushing forward* and *pulling back* elements of top exterior powers, as this concept will be necessary for discussing integration of differential forms in Chapter 8.

4.1 Exterior powers

Antisymmetric tensors

Definition 4.1. (Permutations on $\{1, ..., n\}$).

Let X be any set. A *permutation on* X is a bijection $X \to X$. Intuitively, a permutation on X is thought of as redistributing the names of the elements of X.

We use S_n to denote the *set of permutations on* $\{1, ..., n\}$. Formally, $S_n := \{\{1, ..., n\} \to \{1, ...n\}\}$.

Remark 4.2. (Sign of a permutation).

Without proof, we will use the fact that every permutation in S_n can be decomposed into a composition of "two-element swaps". (Formally, "two-element swaps" are called *transpositions*). The *sign* function on permutations in S_n is the function $\text{sgn} : S_n \to \{-1, 1\}$ defined by $\text{sgn}(\sigma) := (-1)^n$, where n is the number of "two-element swaps" that occur in *any* of the decompositions of σ into only two-element swaps. An equivalent definition of sgn is

$$\text{sgn}(\sigma) := \begin{cases} 1 & \text{there are an even number of "two-element" swaps in any decomposition of } \sigma \\ -1 & \text{there are an odd number of "two-element" swaps in any decomposition of } \sigma \end{cases}.$$

It may seem surprising that sgn is a well-defined function. That is, it may seem surprising that the parity (the "evenness" or "oddness") of the number of two-element swaps in a permutation's decomposition into only two-element swaps is always the same. This might seem relatively surprising, but it's true! (We do not prove this statement about permutations, either).

Definition 4.3. (Permuting an element of a tensor product space).

Let $V_1, ..., V_k$ be vector spaces over the same field. Given a permutation $\sigma \in S_k$ and a tensor $\mathbf{T} \in V_1 \otimes ... \otimes V_k$, we define the map $(\cdot)^\sigma : V_1 \otimes ... \otimes V_k \to V_1 \otimes ... \otimes V_k$ sending $\mathbf{T} \mapsto \mathbf{T}^\sigma$ by specifying its action on elementary tensors and extending linearly. We define

$$(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k)^\sigma := \mathbf{v}_{\sigma(1)} \otimes ... \otimes \mathbf{v}_{\sigma(k)}.$$

Definition 4.4. (Antisymmetric tensor).

Let $V_1, ..., V_k$ be vector spaces over the same field. We say a tensor in $V_1 \otimes ... \otimes V_k$ is *antisymmetric* iff $\mathbf{T}^\sigma = \text{sgn}(\sigma)\mathbf{T}$.

Remark 4.5. (We want to construct an algebra of antisymmetric tensors).

Consider a tuple $(A, K, +, \cdot, \star)$, where A is a set, K is a field, and $\cdot : K \times A \to A$, and $\star : A \times A \to A$ are binary functions. We say that A is an *algebra over* K iff

- $(A, K, +, \cdot)$ is a vector space
- $\star : A \times A \to A$ is bilinear

Let $V_1, ..., V_k$ be vector spaces over the same field. Observe that the tuple

$$(\{\text{antisymmetric tensors in } V_1 \otimes ... \otimes V_k\}, K, \cdot, \otimes)$$

is *not* an algebra, since the tensor product $\otimes$ of antisymmetric two tensors in $V_1 \otimes ... \otimes V_k$ is *not* necessarily another antisymmetric tensor. Consider, for example, the[1] antisymmetric $\binom{1}{0}$ tensors (i.e. vectors) $\mathbf{v}_1$ and $\mathbf{v}_2$; their tensor product $\mathbf{v}_1 \otimes \mathbf{v}_2$ does not satisfy $\mathbf{v}_1 \otimes \mathbf{v}_2 = -\mathbf{v}_2 \otimes \mathbf{v}_1$.

We will construct a function $\wedge : (V_1 \otimes ... \otimes V_k) \times (V_1 \otimes ... \otimes V_k) \to V_1 \otimes ... \otimes V_k$, called the *wedge product*, such that

$$(\{\text{antisymmetric tensors in } V_1 \otimes ... \otimes V_k\}, K, \cdot, \wedge)$$

is an algebra.

Constructing the wedge product

Definition 4.6. (Antisymmetrization of elements of tensor product spaces).

Let $V_1, ..., V_k$ be vector spaces over the same field. We define an *antisymmetrizing function* $\text{alt} : V_1 \otimes ... \otimes V_k \to V_1 \otimes ... \otimes V_k$ that converts any tensor $\mathbf{T} \in V_1 \otimes ... \otimes V_k$ into an antisymmetric tensor in $V_1 \otimes ... \otimes V_k$. We define alt on elementary tensors and extend linearly: for an elementary tensor $\mathbf{T} \in V_1 \otimes ... \otimes V_k$, we define

$$\text{alt}(\mathbf{T}) := \frac{1}{k!} \sum_{\sigma \in S_k} \text{sgn}(\sigma)\mathbf{T}^\sigma.$$

The division by $k!$ is used so that $\text{alt}(\mathbf{T}) = \mathbf{T}$ when $\mathbf{T}$ is antisymmetric.

How might you come up with this formula? Well, you first might start by noticing the case of $k = 2$, without the division by $2!$. That is, notice that we can send $\mathbf{v} \otimes \mathbf{w}$ to the antisymmetric tensor $\mathbf{v} \otimes \mathbf{w} - \mathbf{w} \otimes \mathbf{v}$.

Proof. We need to show that $\text{alt}(\mathbf{T})$ is antisymmetric, i.e., that $\text{alt}(\mathbf{T})^\pi = \text{sgn}(\pi)\text{alt}(\mathbf{T})$. We have

$$\text{alt}(\mathbf{T})^\pi = \Big(\sum_{\sigma \in S_k} \text{sgn}(\sigma)(\mathbf{T}^\sigma) \Big)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma)(\mathbf{T}^\sigma)^\pi = \sum_{\sigma \in S_k} \text{sgn}(\sigma)\mathbf{T}^{\pi \circ \sigma}.$$

Since S_k is closed under taking the inverse of a permutation (recall, a permutation in S_k is a bijection on $\{1, ..., k\}$), then for every $\tau \in S_k$ there is a $\sigma \in S_k$ such that $\tau = \pi \circ \sigma$; namely, $\sigma = \pi^{-1} \circ \tau$. So the sum becomes

$$\sum_{\tau \in S_k} \text{sgn}(\pi^{-1} \circ \tau)\mathbf{T}^\tau = \text{sgn}(\pi^{-1}) \sum_{\tau \in S_k} \text{sgn}(\tau)\mathbf{T}^\tau = -\text{sgn}(\pi)\text{alt}(\mathbf{T}).$$

We have shown $\text{alt}(\mathbf{T})^\pi = \text{sgn}(\pi)\text{alt}(\mathbf{T})$. □

[1] $\mathbf{v}_1$ is antisymmetric because its only permutation $\mathbf{v}_1^\sigma$ is equal to $\mathbf{v}_1^\sigma = \text{sgn}(\sigma)\mathbf{v}_1$, where $\sigma = i$ is the identity. For the same reason, $\mathbf{v}_2$ is antisymmetric.

Definition 4.7. (Wedge product).

Let $V_1, ..., V_k$ be vector spaces over the same field. We define the *wedge product* $\wedge : (V_1 \otimes ... \otimes V_k) \otimes (V_1 \otimes ... \otimes V_k) \to \text{alt}((V_1 \otimes ... \otimes V_k) \otimes (V_1 \otimes ... \otimes V_k))$ by $\mathbf{T} \wedge \mathbf{S} := \text{alt}(\mathbf{T} \otimes \mathbf{S})$.

Lemma 4.8. (Lemma for associativity of wedge product).

Let $V_1, ..., V_k$ be vector spaces over the same field, and consider $\mathbf{T}, \mathbf{S} \in V_1 \otimes ... \otimes V_k$. If $\text{alt}(\mathbf{T}) = \mathbf{0}$, then $\mathbf{T} \wedge \mathbf{S} = \mathbf{0} = \mathbf{S} \wedge \mathbf{T}$.

Proof. (This proof requires some abstract algebra. Understanding this proof is not really necessary to understand exterior powers; just take this theorem as an axiom).

Assume $\text{alt}(\mathbf{T}) = \mathbf{0}$. Let $T = \mathbf{v}_{i_1} \otimes ... \otimes \mathbf{v}_{i_k}$ 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_{2k} whose elements fix all of $j_1, ..., j_k$, and consider the right cosets $\{H\sigma \mid \sigma \in S_{2k}\}$ of H in S_{2k}. Since these right cosets partition S_{2k}, then

$$\begin{aligned}\text{alt}(\mathbf{T} \otimes \mathbf{S}) &= \sum_{[\pi]_\tau \sigma \in \{\text{right cosets}\}} (-1)^{[\pi]_\tau \sigma} (\mathbf{T} \otimes \mathbf{S})^{\pi_\tau \sigma} = \sum_{\sigma \in S_{k+m}} \sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau \sigma} (\mathbf{T} \otimes \mathbf{S})^{\pi_\tau \sigma} \\ &= \sum_{\sigma \in S_{k+m}} \Big(\sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} (\mathbf{T} \otimes \mathbf{S})^{\pi_\tau} \Big)^\sigma .\end{aligned}$$

Since $[\pi]_\tau \in H$, where H is the subgroup of S_{2k} whose elements fix all of $j_1, ..., j_\ell$, then $(\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} = \mathbf{T}^{[\pi]_\tau} \otimes \mathbf{S}$. With this, the innermost sum becomes

$$\sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} (\mathbf{T} \otimes \mathbf{S})^{[\pi]_\tau} = \sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} \mathbf{T}^{[\pi]_\tau} \otimes \mathbf{S} = \Big(\sum_{[\pi]_\tau \in H} (-1)^{[\pi]_\tau} \mathbf{T}^{[\pi]_\tau} \Big) \otimes \mathbf{S}.$$

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 4.9. (Wedge product is associative).

Let $V_1, ..., V_k$ be vector spaces over the same field. Then for all $\mathbf{T}, \mathbf{S}, \mathbf{R} \in V_1 \otimes ... \otimes V_k$, we have $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R})$, and are therefore justified in denoting both as $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R}) := \mathbf{T} \wedge \mathbf{S} \wedge \mathbf{R}$.

Proof. We will show $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$; a similar argument shows $\mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R}) = \text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$.

First, we have by definition of $\wedge$ that

$$(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \text{alt}((\mathbf{T} \wedge \mathbf{S}) \otimes \mathbf{R})$$

Subtracting $\text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R})$ from both sides and using linearity of alt, we get that

$$(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} - \text{alt}(\mathbf{T} \otimes \mathbf{S} \otimes \mathbf{R}) = \text{alt}((\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \otimes \mathbf{R}) = (\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R}.$$

If we show $(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R} = \mathbf{0}$, then our claim is true. The previous lemma says that if $\text{alt}(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) = \mathbf{0}$, then $(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) \wedge \mathbf{R} = \mathbf{0}$. And this is true, since $\text{alt}(\mathbf{T} \wedge \mathbf{S} - \mathbf{T} \otimes \mathbf{S}) = \text{alt}(\mathbf{T} \wedge \mathbf{S}) - \text{alt}(\mathbf{T} \otimes \mathbf{S}) = \mathbf{T} \wedge \mathbf{S} - \mathbf{T} \wedge \mathbf{S} = \mathbf{0}$, by linearity of alt and with use of the fact that $\mathbf{T} \wedge \mathbf{S}$ is antisymmetric. □

Wedge product spaces and exterior powers

Theorem 4.10. (Properties of the wedge product).

Let $V_1, ..., V_k$ be vector space over a field $K \neq \mathbb{Z}/2\mathbb{Z}$, and consider the tensor product space $V_1 \otimes ... \otimes V_k$. The wedge product $\wedge$ satisfies the following properties...

1. $\wedge$ looks as if it is bilinear, just as was the case with $\otimes$. That is, ...
    1.1. $(\mathbf{T} + \mathbf{S}) \wedge \mathbf{R} = \mathbf{T} \wedge \mathbf{R} + \mathbf{S} \wedge \mathbf{R}$ for all $\mathbf{T}, \mathbf{S}, \mathbf{R} \in V_1 \otimes ... \otimes V_k$.
    1.2. $\mathbf{T} \wedge (\mathbf{S} + \mathbf{R}) = \mathbf{T} \wedge \mathbf{S} + \mathbf{T} \wedge \mathbf{R}$ for all $\mathbf{T}, \mathbf{S}, \mathbf{R} \in V_1 \otimes ... \otimes V_k$.
    1.3. $(c\mathbf{T}) \wedge \mathbf{S} = c(\mathbf{T} \wedge \mathbf{S}) = \mathbf{T} \wedge (c\mathbf{S})$ for all $\mathbf{T}, \mathbf{S} \in V_1 \otimes ... \otimes V_k$ and $c \in K$.
2. $\wedge$ is associative, just as was the case with $\otimes$: $(\mathbf{T} \wedge \mathbf{S}) \wedge \mathbf{R} = \mathbf{T} \wedge (\mathbf{S} \wedge \mathbf{R})$ for all $\mathbf{T}, \mathbf{S} \in V_1 \otimes ... \otimes V_k$.
3. $\wedge$ looks as it it is an antisymmetric map: $\mathbf{T} \wedge \mathbf{S} = -\mathbf{S} \wedge \mathbf{T}$ for all $\mathbf{T}, \mathbf{S} \in V_1 \otimes ... \otimes V_k$.
4. $\wedge$ is *skew-commutative*: if $\mathbf{T} \in V_1 \otimes ... \otimes V_k$ and $\mathbf{S} \in V_1 \otimes ... \otimes V_\ell$, then $\mathbf{S} \wedge \mathbf{T} = (-1)^{k+\ell}(\mathbf{T} \wedge \mathbf{S})$.
5. $\mathbf{T} \wedge \mathbf{T} = \mathbf{0}$ for all $\mathbf{T} \in V_1 \otimes ... \otimes V_k$.

Proof. Property (1) follows by checking the definition $\mathbf{T} \wedge \mathbf{S} := \text{alt}(\mathbf{T} \otimes \mathbf{S})$. Property (2) was proved in Theorem 4.9. Property (3) follows from the definition of alt. Conditions (3) and (4) are logically equivalent, and conditions (3) and (5) are logically equivalent when $K \neq \mathbb{Z}/2\mathbb{Z}$. (Again, we need to require $K \neq \mathbb{Z}/2\mathbb{Z}$ here so that $2 \neq 0$, which allows division by 2). □

In Remark 4.5, we said that we wanted to find a function $\wedge : (V_1 \otimes ... \otimes V_k) \times (V_1 \otimes ... \otimes V_k) \to V_1 \otimes ... \otimes V_k$ for which the tuple $(\{\text{antisymmetric tensors in } V_1 \otimes ... \otimes V_k\}, K, \cdot, \wedge)$ is an algebra. We have done so by constructing the wedge product $\wedge$, and now formalize this result with the next definitions.

Definition 4.11. (Wedge product space).

Let $V_1, ..., V_k$ be vector spaces over a field K. We define the *wedge product space* $V_1 \wedge ... \wedge V_k$ to be the algebra

$$V_1 \wedge ... \wedge V_k := (\text{alt}(V_1 \otimes ... \otimes V_k), K, +, \cdot, \wedge).$$

Theorem 4.12. (Basis and dimension for wedge product spaces).

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

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

Proof. See the proof of Theorem 2.8. □

Theorem 4.13. (Exterior powers).

Let V be a vector space. We define the *kth exterior power* $\Lambda^k(V)$ of V to be the wedge product space $\Lambda^k(V) := V^{\wedge k}$.

Theorem 4.14. (Basis and dimension of exterior powers).

Let V be an n-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. Then $\Lambda^k(V)$ is a $\binom{n}{k}$ dimensional vector space with basis

$$\{\mathbf{e}_{i_1} \wedge ... \wedge \mathbf{e}_{i_k} \mid k \in \{1, ..., n\} \text{ and } i_1, ..., i_k \text{ is strictly increasing}\}.$$

Note, $i_1, ..., i_k$ must be strictly increasing because $\mathbf{v} \wedge \mathbf{v} = \mathbf{0}$.

Proof. To show that this set spans $\Lambda^k(V)$, use the seeming-multilinearity of $\wedge$ as was done for $\otimes$ in the proof of Theorem 2.8.

For linear independence, note that if the sequence $i_1, ..., i_k$ were not strictly increasing, then we would have a $\mathbf{0}$ in the claimed basis, which would make our claimed basis a linearly dependent set. Since there is no $\mathbf{0}$ in the claimed basis, we can follow the proof of Theorem 2.8. $\square$

Definition 4.15. (Alternating function).

Let $V_1, ..., V_k, W$ be vector spaces over the same field. We say a function $\mathbf{f} : V_1 \times ... \times V_k \to W$ is a *k-alternating function* iff for all $\sigma \in S_n$, we have $\mathbf{f}(\mathbf{v}_{\sigma(1)}, ..., \mathbf{v}_{\sigma(k)}) = \text{sgn}(\sigma)\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_k)$. Equivalently, $\mathbf{f}$ is k-alternating iff $\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_i, ..., \mathbf{v}_j, ..., \mathbf{v}_k) = -\mathbf{f}(\mathbf{v}_1, ..., \mathbf{v}_j, ..., \mathbf{v}_i, ..., \mathbf{v}_k)$ for all $i \in \{1, ..., k\}$.

When k is clear from the context, k-linear functions are called *alternating functions*.

Remark 4.16. ("Antisymmetric" vs. "alternating").

Many authors refer to the "antisymmetric tensors" of Definition 4.4 as "alternating tensors". Doing so is technically incorrect, because an "alternating tensor" is an element of a certain "quotient algebra". This quotient algebra is isomorphic to an algebra of antisymmetric tensors only when the "characteristic" of the field K is infinity. In other words, "antisymmetric tensors" are only the same as "alternating tensors" in certain special cases.

Given the previous paragraph, one might think that "alternating functions" should really be called "antisymmetric functions". This is actually not the case. We are correct to define the functions of the above to be "alternating functions" because the alternating functions of the above definition correspond to alternating tensors (which, remember, are elements of a certain quotient algebra) via *the universal property of the exterior algebra*. So, alternating functions share the same level of generality as alternating tensors (which we have not defined). Antisymmetric tensors, which we have defined, are what are "more specific".

We have not defined what the exterior algebra is (and won't), but it is good to know this context.

Definition 4.17. (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 4.18. (Universal property for exterior powers).

Let V_1, V_2, W be vector spaces over the same field, and let $\mathbf{f} : V_1 \times V_2 \to W$ be an alternating bilinear function. Then there exists a linear function $\mathbf{h} : V_1 \wedge V_2 \to W$ with $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ where $\mathbf{h}$ uniquely depends on $\mathbf{f}$, and where $\mathbf{g} : V_1 \times V_2 \to V_1 \otimes V_2$.

In Remark 4.16, we mentioned that there exists a "universal property of the exterior algebra". Note that this theorem is *not* the universal property of the exterior algebra, but a special case of it. (As mentioned before, we have not stated the universal property of the exterior algebra, and will not state it in this book).

Proof. The proof is similar to the proof of the universal property of tensor product spaces (Theorem 2.9). The only difference is that the maps we define in this proof are extended using antisymmetry and bilinearity, rather than just bilinearity. $\square$

Theorem 4.19. (Fundamental natural isomorphisms for exterior powers).

Theorem 2.33 stated that there are natural isomorphisms

$$\mathcal{L}(V_1 \times ... \times V_k \to W) \cong \mathcal{L}(V_1 \otimes ... \otimes V_k \to W)$$
$$\mathcal{L}(V \to W) \cong W \otimes V^*$$
$$(V \otimes W)^* \cong V^* \otimes W^*$$
$$T_q^p(V) \cong T_p^q(V^*)$$

Analogously, there are natural isomorphisms

$$\boxed{\begin{gathered}(\text{alt}\mathcal{L})(V_1 \times ... \times V_k \to W) \cong (\text{alt}\mathcal{L})(V_1 \wedge ... \wedge V_k \to W)\\ (\text{alt}\mathcal{L})(V \to W) \cong W \wedge V^*\\ (V \wedge W)^* \cong V^* \wedge W^*\\ \Lambda^k(V)^* \cong \Lambda^k(V^*)\end{gathered}}$$

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 2.10). 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 2.2, 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 2.2.

The third line in the box is proved similarly as was in Theorem 2.33; the only difference is that it is necessary to extend with antisymmetry and bilinearity rather than just bilinearity. The fourth line follows from the third line.

□

Pushforward and pullback

We now momentarily take a step backwards and introduce the *pushforward* and *pullback* operations for $\binom{k}{0}$ and $\binom{0}{k}$ tensors. This prepares us to understand pushforward and pullback for exterior powers.

Definition 4.20. (Pushforward on $T_0^k(V)$).

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Using $\mathbf{f}$, we can *push forward* $\binom{k}{0}$, $k \leq n$, tensors in $T_0^k(V) = V^{\otimes k}$ to $\binom{k}{0}$ tensors in $T_0^k(W) = W^{\otimes k}$ as follows:

$$\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k \overset{\otimes_0^k \mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \otimes ... \otimes \mathbf{f}(\mathbf{v}_k).$$

(As usual, we extend this map, which is stated in terms of elementary $\binom{k}{0}$ tensors, to act on arbitrary $\binom{k}{0}$ tensors with the seeming-multilinearity of $\otimes$). We have denoted this map by $\otimes_0^k \mathbf{f}$, so we can also state the above as

$$(\otimes_0^k \mathbf{f})(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k) = \mathbf{f}(\mathbf{v}_1) \otimes ... \otimes \mathbf{f}(\mathbf{v}_k).$$

The map $\otimes_0^k \mathbf{f} : T_0^k(V) \to T_0^k(W)$ is called the *pushforward*. Since we have extended it with the seeming-multilinearity of $\otimes$, the pushforward is a linear map.

Definition 4.21. (Pullback on $\widetilde{T}_k^0(W)$).

As in the previous definition, let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. We can use the dual $\mathbf{f} : W^* \to V^*$ of $\mathbf{f}$ to construct a map that *pulls back* elements of $T_k^0(W) = T_0^k(W^*)$, $k \leq n$, to elements of $T_k^0(V) = T_0^k(V^*)$, as follows:

$$\mathbf{w}_1 \otimes ... \otimes \mathbf{w}_k \overset{\otimes_k^0 \mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\mathbf{w}_1) \otimes ... \otimes \mathbf{f}^*(\mathbf{w}_k).$$

(As in the previous definition, this map is extended to act on arbitrary $\binom{0}{k}$ tensors with the seeming-multilinearity of $\otimes$). We have denoted this map by $\otimes_k^0 \mathbf{f}^*$, so we can also state the above as

$$(\otimes^0_k \mathbf{f}^*)(\mathbf{w}_1 \otimes ... \otimes \mathbf{w}_k) = \mathbf{f}^*(\mathbf{w}_1) \otimes ... \otimes \mathbf{f}^*(\mathbf{w}_k).$$

The map $\otimes^0_k \mathbf{f}^* : T^0_k(W) \to T^0_k(V)$ is called the *pullback*. As was the case with the pushforward, the pullback is a linear map because its definition was extended with the seeming-multilinearity of $\otimes$.

Theorem 4.22. [Lee, p. 320] (Properties of tensor pullbacks).

Having introduced pushforward and pullback for $\binom{k}{0}$ and $\binom{0}{k}$ tensors, we now are ready to look at pushforward and pullback for exterior powers.

Definition 4.23. (Pushforward on $\Lambda^k(V)$).
Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Using $\mathbf{f}$, we can *push forward* elements of $\Lambda^k(V)$, $k \leq n$, to elements of $\Lambda^k(W)$ by using the *pushforward* $\Lambda^k \mathbf{f}$ *on* $\Lambda^k(V)$ defined by

$$\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k \overset{\Lambda^k \mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_k).$$

We extend this map to act on arbitrary elements of $\Lambda^k(V)$ with the seeming-multilinearity and antisymmetry $\wedge$.

Definition 4.24. (Pullback on $\Lambda^k(W^*)$).
Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Using $\mathbf{f}$, we can *pull back* elements of $\Lambda^k(W^*)$ to elements of $\Lambda^k(V^*)$, $k \leq n$, by using the *pullback* $\Lambda^k \mathbf{f}^*$ *on* $\Lambda^k(W^*)$ defined by

$$\psi^1 \wedge ... \wedge \psi^k \overset{\Lambda^k \mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1) \wedge ... \wedge \mathbf{f}^*(\psi^k).$$

We extend this map to act on arbitrary elements of $\Lambda^k(W)$ with the seeming-multilinearity and antisymmetry $\wedge$.

Remark 4.25. (Star notation for pushforward and pullback).
The above notation of using $\otimes^0_k \mathbf{f} : V \to W$ or $\Lambda^k \mathbf{f}$ for the pushforward and using $\otimes^k_0 \mathbf{f}^* : W^* \to V^*$ or $\Lambda^k \mathbf{f}^*$ for the pullback is nonstandard. It is more common to denote a pullback by $\mathbf{f}^* : T^0_k(W) \to T^0_k(V)$ or $\mathbf{f}^* : \Lambda^k(W) \to \Lambda^k(V)$. Somewhat less commonly, a pushforward is denoted by $\mathbf{f}_* : T^k_0(V) \to T^k_0(W)$ or $\mathbf{f}_* : \Lambda^k(V) \to \Lambda^k(W)$.

Once the above definitions of pushforward and pullback are understood, this "star" notation can be useful. But it can be difficult to understand what pushfoward and pullback are if this notation is used from the start, due to the potential for confusing the dual $\mathbf{f}^* : W^* \to V^*$ with either of the pullbacks $\mathbf{f}^* : T^0_k(W) \to T^0_k(V), \mathbf{f}^* : \Lambda^k(W) \to \Lambda^k(V)$.

We will use the star notation after we define a pullback of a differential form in Chapter 8. Until then, we do not use this notation.

4.2 The determinant

Definition 4.26. (The determinant).

Let K be a field, and let $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ be the standard basis of K^n. We want to define a function $(K^n)^{\times n} \to K$ which, given $\mathbf{c}_1, ..., \mathbf{c}_n \in K^n$, returns the n-dimensional volume of the parallelapiped spanned by $\mathbf{c}_1, ..., \mathbf{c}_n$. We will denote this function by $\det : (K^n)^{\times n} \to K$. We require that det satisfy the following axioms:

1. $\det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$, since we want the unit n-cube to have an n-dimensional volume of 1.
2. det is multilinear, because...
   - The volume of a parallelapiped that is the disjoint union of two smaller parallelapipeds should be the sum of the volumes of the smaller parallelapipeds.
   - Scaling one of the sides of a parallelapiped by $c \in K$ should increase that paralellapiped's volume by a factor of c.
3. $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n) = 0$ when $\mathbf{c}_i = \mathbf{c}_j$ for all $\mathbf{c}_k \in K$, $k \in \{1, ..., n\}$. This should hold because when two sides of a parallelapiped coincide, its n-*dimensional* volume is zero.

When $K \neq \mathbb{Z}/2\mathbb{Z}$ so that $2 \neq 0$, which enables division by 2, then, due to the multilinearity of det, the third axiom is logically equivalent to det being an alternating function. (Proof left as exercise). This is the case when $K = \mathbb{R}$, for example. The fact that det is (almost always) alternating means that our intuitive assumptions about volume require that volume be *signed*, or *oriented*; the volume of the parallelapiped spanned by $\mathbf{c}_1, ..., \mathbf{c}_j, ..., \mathbf{c}_i, ..., \mathbf{c}_n$ is the negation of the volume of the parallelapiped spanned by $\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_j, ..., \mathbf{c}_n$.

The fact that the third axiom is logically equivalent to alternatingness also gives us a concise characterization of the determinant: $\det : (K^n)^{\times n} \to K$, when $K \neq \mathbb{Z}/2\mathbb{Z}$, is the unique multilinear alternating function satisfying $\det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) = 1$.

Definition 4.27. (Determinant of a square matrix).

We define the *determinant of a square matrix* to be the result of applying det to the column vectors of that matrix.

Theorem 4.28. (Consequent properties of the determinant).

5. det is invariant under linearly combining input vectors into a different input vector. That is, $\det(\mathbf{c}_1, ..., \mathbf{c}_i, ..., \mathbf{c}_n) = \det(\mathbf{c}_1, ..., \mathbf{c}_i + \sum_{j=1, j\neq i}^n d_j \mathbf{c}_j, ..., \mathbf{c}_n)$ 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 4.29. (Permutation formula for the determinant).

We now derive *permutation formula* for the det. The formula we obtain shows that the function det specified in Definition 4.26 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[2] set $(a^i_j) := ([\mathbf{c}_i]_{\hat{\mathfrak{E}}})^j$. Then...

$$
\begin{aligned}
\det((a^i_j)) = \det(\mathbf{c}_1, ..., \mathbf{c}_n) &= \det\Big(\sum_{i_1=1}^{n} a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., \sum_{i_n=1}^{n} a^n_{i_n}\hat{\mathfrak{e}}_{i_n}\Big) \\
&= \sum_{i_1=1}^{n} \det\Big(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., \sum_{i_n=1}^{n} a^n_{i_n}\hat{\mathfrak{e}}_{i_n}\Big) \\
&= \sum_{i_1=1}^{n} \det(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., \sum_{i_n=1}^{n} a^n_{i_n}\hat{\mathfrak{e}}_{i_n}\Big) \\
&\vdots \\
&= \sum_{i_1=1}^{n} ... \sum_{i_n=1}^{n} \det(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., a^n_{i_n}\hat{\mathfrak{e}}_{i_n}) \\
&= \sum_{i_1=1}^{n} ... \sum_{i_n=1}^{n} \det(a^1_{i_1}\hat{\mathfrak{e}}_{i_1}, ..., a^n_{i_n}\hat{\mathfrak{e}}_{i_n}), \text{ where } i_1, ..., i_n \text{ are distinct from each other} \\
&= \sum_{\sigma \in S_n} \det(a^1_{\sigma(1)}\hat{\mathfrak{e}}_{\sigma(1)}, ..., a^n_{\sigma(n)}\hat{\mathfrak{e}}_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)} \det(\hat{\mathfrak{e}}_{\sigma(1)}, ..., \hat{\mathfrak{e}}_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)}\operatorname{sgn}(\sigma) \det(\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n) \\
&= \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)}\operatorname{sgn}(\sigma)
\end{aligned}
$$

Therefore, we have

$$
\boxed{\det(\mathbf{c}_1, ..., \mathbf{c}_n) = \sum_{\sigma \in S_n} a^1_{\sigma(1)}...a^n_{\sigma(n)}\operatorname{sgn}(\sigma)}
$$

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

[2]There is no hidden meaning behind the upper and lower indices on a^i_j here; we only want to consider an arbitrary $n \times n$ matrix of scalars in K, and prefer to think of this matrix as storing the coordinates of a $\binom{1}{1}$ tensor rather those of a $\binom{2}{0}$ or $\binom{0}{2}$ tensor.

one nonzero entry in each column (see the line directly above the line labeled with "where $i_1, ..., i_n$ are distinct from each other"). The second step is to disregard all determinants in this previous sum whose matrix arguments have two or more columns that have their nonzero entries in the same row (i.e. whose matrix arguments are matrices of linearly dependent columns). This leaves us with a sum of determinants of diagonal matrices whose columns have been shuffled (this corresponds to the line labeled with "where $i_1, ..., i_n$ are distinct from each other" and the line directly below it). The third step, which corresponds to the third to last line, is to use multilinearity to pull out all the constants. The fourth step is to use the alternatingness of the determinant so that every determinant argument in the sum is the identity matrix; this results in multiplying each term in the sum by $\text{sgn}(\sigma)$.

Theorem 4.30. (Determinant of a matrix is transpose-invariant).

Let $\mathbf{A}$ be a square matrix with entries in K. Recall from the discussion after the statement of the permutation formula for the determinant that the determinant of a matrix is a sum of determinants of diagonal matrices whose columns have been shuffled. Each shuffled diagonal matrix in this sum can be momentarily converted to a diagonal matrix, transposed, and then re-shuffled (so that the columns of the shuffled-transposed-reshuffled matrix are in the order of the columns of the original shuffled diagonal matrix). Reversing the expansion that was accomplished with the multilinearity of det in the derivation of the permutation formula for the determinant, we see that the sum of the determinants of these shuffled-transposed-reshuffled matrices is equal to the determinant of $\mathbf{A}^\top$. Therefore

$$\det(\mathbf{A}) = \det(\mathbf{A}^\top).$$

Theorem 4.31. (Laplace expansion for the determinant).

Consider an $n \times n$ matrix $\mathbf{A} = (a^i_j)$, and let $\mathbf{A}^i_j$ denote the so-called ij *minor matrix* obtained by erasing the ith row and jth column of $\mathbf{A}$. We have

$$\det(\mathbf{A}) = \sum_{i=1}^n a^i_j \det(\mathbf{A}^i_j) \text{ for all } i \in \{1, ..., n\}$$
$$\det(\mathbf{A}) = \sum_{j=1}^n a^i_j \det(\mathbf{A}^i_j) \text{ for all } j \in \{1, ..., n\}$$

The first equation is called the *Laplace expansion for the determinant along the ith row*, and the second equation is called the *Laplace expansion for the determinant along the jth column.* Note that each equation implies the other because $\det(\mathbf{A}) = \det(\mathbf{A}^\top)$.

Proof. We prove the second equation of the theorem.

Consider all terms in the permutation formula's sum for $\det(\mathbf{A})$ that have the factor a^i_j. Let $\mathbf{B}$ denote the shuffled diagonal matrix that corresponds to one of these terms. We can view $\det(\mathbf{B})$ as $\det(\mathbf{B}) = \pm a^i_j \det(\mathbf{B}^i_j)$, where $\mathbf{B}^i_j$ is the determinant of the matrix obtained by removing the ith column and jth row from $\mathbf{B}$. The $\pm$ sign is a result of the fact that the matrices $\mathbf{B}$ and $\mathbf{B}^i_j$ may have different inversion counts.

The main effort of this proof is to determine the $\pm$ sign and specify how the inversion counts of $\mathbf{B}$ and $\mathbf{B}^i_j$ differ.

As a first step, note that the difference in the inversion count between $\mathbf{B}$ and $\mathbf{B}^i_j$ is the number of inversions that involve a^i_j. Thus, our problem reduces to determining an expression for the number of inversions that involve a^i_j. So, divide the matrix $\mathbf{B}$ into quadrants that are centered on a^i_j. Let k_1, k_2, k_3, k_4 be the number of inversions in the upper left, upper right, lower left, and bottom right corners of $\mathbf{A}$, respectively. The number of inversions involving a^i_j is $k_2 + k_3$. Since we know $k_1 + k_2 + 1 = i$ and $k_1 + k_3 + 1 = j$, we have $k_2 + k_3 = i + j - 2 - 2k_1 = i + j - 2(k_1 + 1)$. (We also know $k_1 + k_2 + k_3 + k_4 = n$, but this is not that helpful). Thus, if σ is the permutation corresponding to $\mathbf{B}$ and π is the permutation corresponding to $\mathbf{B}^i_j$, then $\text{sgn}(\sigma) = \text{sgn}(\pi)(-1)^{i+j-2(k_1+1)} = \text{sgn}(\pi)(-1)^{i+j}$. Thus $\text{sgn}(\sigma) = (-1)^{i+j}\text{sgn}(\pi) \iff \text{sgn}(\pi) = (-1)^{i+j}\text{sgn}(\sigma)$.

So,

$$
\begin{aligned}
a^i_j \det(\mathbf{B}^i_j) &= a^i_j \sum_{\pi \in S_n} a_1^{\pi(1)} ... \cancel{a_j^{\pi(i)}} ..., a_n^{\pi(n)} \text{sgn}(\pi) \\
&= a^i_j \sum_{\sigma \in S_n} a_1^{\pi(1)} ... \cancel{a_j^{\pi(i)}} ..., a_n^{\pi(n)} (-1)^{i+j} \text{sgn}(\sigma) \\
&= (-1)^{i+j} a^i_j \det(\mathbf{B})
\end{aligned}
$$

Thus $a^i_j \det(\mathbf{B}^i_j) = (-1)^{i+j} a^i_j \det(\mathbf{B}) \iff \det(\mathbf{B}) = (-1)^{i+j} a^i_j \det(\mathbf{B}^i_j)$. Now sum all of the $\mathbf{B}$'s (the diagonal shuffled matrices) to get $\det(\mathbf{A}) = \sum_{j=1}^n a^i_j \det(\mathbf{A}^i_j)$. □

Definition 4.32. (Determinant of a linear function).

Let V and W be finite-dimensional vector spaces *of the same dimension*, and let E and F be be bases for V and W. We define the *determinant of a linear function* $\mathbf{f} : V \to W$ to be the determinant of the matrix of $\mathbf{f}$ relative to E and F, $\det(\mathbf{f}) := \det([\mathbf{f}(E)]_F)$.

Remark 4.33. We have not yet shown that the determinant of a linear function $V \to V$ is well defined; we have not shown that it doesn't on the basis chosen for V. We will see that this is the case soon.

Theorem 4.34. (Determinant of a matrix is dual-invariant).

Let V and W be finite-dimensional vector spaces of the same dimension, and consider a linear function $\mathbf{f} : V \to W$. Consider also the dual $\mathbf{f} : W^* \to V^*$ (recall Definition 3.4). Then $\det(\mathbf{f}^*) = \det(\mathbf{f})$.

Proof. Recall from[3] condition (3) of Definition 3.21 that if $\mathbf{A}$ is the matrix of $\mathbf{f}$ relative to orthonormal bases $\widehat{U}_1$ and $\widehat{U}_2$, then the matrix of $\mathbf{f}^*$ relative to the induced dual bases $\widehat{U}_2^*$ and $\widehat{U}_1^*$ is $\mathbf{A}^\top$. Since the determinant of a matrix is transpose-invariant (recall Theorem 4.30,) we have $\det(\mathbf{f}) = \det(\mathbf{A}) = \det(\mathbf{A}^\top) = \det(\mathbf{f}^*)$. □

Theorem 4.35. (The pushforward on the top exterior power is multiplication by the determinant).

Let V and W be n-dimensional vector spaces over a field K, let $\mathbf{f} : V \to W$ be a linear function, and consider $\Lambda^n(V)$. We call $\Lambda^n(V)$ the *top exterior power of* V because n is the largest positive integer for which $\Lambda^k(V)$ is not a zero-dimensional vector space. Consequently, it is the exterior power of smallest dimension; $\dim(\Lambda^n(V)) = \binom{n}{n} = 1$.

Now consider the pushforward $\Lambda^n \mathbf{f} : \Lambda^n(V) \to \Lambda^n(W)$ on the top exterior power, which (recall Definition 4.23) is defined on elementary tensors by $\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n \mapsto \mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_n)$ and is extended with multilinearity. Defined this way, the pushforward $\Lambda^n \mathbf{f}$ is a multilinear alternating map. This pushforward is also a map of 1-dimensional vector spaces, so it must be multiplication by a constant. We will determine what this constant is.

Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be a basis of V, and consider the action of the pushforward on the basis vectors of E,

$$
\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n).
$$

Because $\Lambda^n(\mathbf{f})$ is multilinear and alternating, the wedge product on the left hand side closely resembles the determinant $\det(\mathbf{f}(\mathbf{e}_1), ..., \mathbf{f}(\mathbf{e}_n)) = \det([\mathbf{f}(E)]_F) = \det(\mathbf{f})$. So, set $(a^i_j) = [\mathbf{f}(E)]_F$, and then use essentially the same argument as was made to derive the permutation formula on the left hand side of the above. We obtain

[3] Technically, the equivalent conditions of the definition we reference only apply to linear functions $V \to V$. This is not an issue because V and W have the same dimension; if we want to be very formal, we can use the linear function $\widetilde{\mathbf{f}} : V \to V$ that is obtained from $\mathbf{f}$ by identifying $W \cong V$ with the identification that sends basis vectors of W to basis vectors of V.

$$\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n) = \sum_{\sigma \in S_n} \Big(a^1_{\sigma(1)} ... a^n_{\sigma(n)} \mathrm{sgn}(\sigma) \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n \Big) = \Big(\sum_{\sigma \in S_n} a^1_{\sigma(1)} ... a^n_{\sigma(n)} \mathrm{sgn}(\sigma) \Big) \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n$$
$$= \det([\mathbf{f}(E)]_F)(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n) = \det(\mathbf{f})(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n).$$

So, we have the statement on the basis E

$$\mathbf{f}(\mathbf{e}_1) \wedge ... \wedge \mathbf{f}(\mathbf{e}_n) = \det(\mathbf{f})(\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n).$$

Using the seeming-multilinearity of $\wedge$, we can extend this fact to apply to any list of vectors in V:

$$\boxed{\mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_n) = \det(\mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n) \text{ for all } \mathbf{v}_1, ..., \mathbf{v}_n \in V}$$

We can explicitly involve the pullback $\Lambda^n \mathbf{f}$ and write the above as

$$(\Lambda^n \mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n) = \det(\mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n) \text{ for all } \mathbf{v}_1, ..., \mathbf{v}_n \in V.$$

Thus, $\Lambda^n(\mathbf{f})$ is multiplication by $\det(\mathbf{f})$.

Theorem 4.36. (The pullback on the top exterior power is multiplicaiton by the determinant).
Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Consider additionally the dual $\mathbf{f}^* : W^* \to V^*$ (recall Definition 3.4). Then, applying the previous theorem and using that $\det(\mathbf{f}) = \det(\mathbf{f}^*)$, we see that the pullback $\Lambda^n \mathbf{f}^*$ on the top exterior power satisfies

$$(\Lambda^n \mathbf{f}^*)(\mathbf{w}_1 \wedge ... \wedge \mathbf{w}_n) = \det(\mathbf{f})(\mathbf{w}_1 \wedge ... \wedge \mathbf{w}_n) \text{ for all } \mathbf{w}_1, ..., \mathbf{w}_n \in V.$$

That is,

$$\boxed{\mathbf{f}^*(\mathbf{w}_1) \wedge ... \wedge \mathbf{f}^*(\mathbf{w}_n) = \det(\mathbf{f})(\mathbf{w}_1 \wedge ... \wedge \mathbf{w}_n)}$$

Theorem 4.37. (Product rule for determinants). Let V, W and Z be finite-dimensional vector spaces of the same dimension, and consider linear functions $\mathbf{f} : V \to W$ and $\mathbf{g} : W \to Z$. Then $\det(\mathbf{g} \circ \mathbf{f}) = \det(\mathbf{g}) \det(\mathbf{f})$. Thus, if $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times p$ matrix, then $\det(\mathbf{BA}) = \det(\mathbf{B}) \det(\mathbf{A})$.

Proof. Set $n := \dim(V) = \dim(W) = \dim(Z)$. By the previous theorem, $\det(\mathbf{g} \circ \mathbf{f})$ satisfies

$$(\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1) \wedge ... \wedge (\mathbf{g} \circ \mathbf{f})(\mathbf{v}_n) = \det(\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n)$$

for all $\mathbf{v}_1, ..., \mathbf{v}_n \in V$.
Notice that the left hand side is

$$\mathbf{g}(\mathbf{f}(\mathbf{v}_1)) \wedge ... \wedge \mathbf{g}(\mathbf{f}(\mathbf{v}_n)) = \det(\mathbf{g})(\mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_n)) = \det(\mathbf{g}) \det(\mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n).$$

Thus

$$\det(\mathbf{g} \circ \mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n) = \det(\mathbf{g}) \det(\mathbf{f})(\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n).$$

This is a statement on the basis vector $\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_n$ of $\Lambda^n(V)$. Extending this statement to a statement on any vector $\mathbf{T} \in \Lambda^n(V)$, we have $\det(\mathbf{g} \circ \mathbf{f})\mathbf{T} = \det(\mathbf{g}) \det(\mathbf{f})\mathbf{T}$. Thus $(\det(\mathbf{g} \circ \mathbf{f}) - \det(\mathbf{g}) \det(\mathbf{f}))\mathbf{T} = \mathbf{0}$ for all $\mathbf{T} \in \Lambda^n(V)$. Since we can choose $\mathbf{T} \neq \mathbf{0}$, this forces $\det(\mathbf{g} \circ \mathbf{f}) - \det(\mathbf{g}) \det(\mathbf{f}) = 0$. $\square$

Theorem 4.38. (Determinant of an inverse function).

Let V and W be finite-dimensional vector spaces of the same dimension, and consider a linear function $\mathbf{f} : V \to W$. Then $\det(\mathbf{f}^{-1}) = \frac{1}{\det(\mathbf{f})}$.

Proof. We have $\det(\mathbf{f} \circ \mathbf{f}^{-1}) = \det(\mathbf{I}) = 1$, and $\det(\mathbf{f} \circ \mathbf{f}^{-1}) = \det(\mathbf{f})\det(\mathbf{f}^{-1})$ by the previous theorem, so $\det(\mathbf{f})\det(\mathbf{f}^{-1}) = 1$. □

Theorem 4.39. (Determinant of a linear function is well-defined).

Let V and W be finite-dimensional vector spaces of the same dimension, and consider a linear function $\mathbf{f} : V \to W$. For simplicity, consider the linear function $\widetilde{\mathbf{f}} : V \to V$ induced by $\mathbf{f}$, where the "second" V is thought of as coming from the identification $W \cong V$ that identifies basis vectors of W with basis vectors of V. Let E and F be bases for V. Then, the matrix $[\widetilde{\mathbf{f}}(F)]_F$ of $\widetilde{\mathbf{f}}$ relative to F and F is related to the matrix $[\widetilde{\mathbf{f}}(E)]_E$ of $\widetilde{\mathbf{f}}$ relative to E and E by $[\widetilde{\mathbf{f}}(F)]_F = [\mathbf{F}]_F[\widetilde{\mathbf{f}}(E)]_E[\mathbf{F}]_F^{-1}$ (see Theorem 1.57). Therefore, by the previous theorem, $\det(\widetilde{\mathbf{f}})$ is well-defined, because $\det([\widetilde{\mathbf{f}}(F)]_F) = \det([\mathbf{F}]_F[\widetilde{\mathbf{f}}(E)]_E[\mathbf{F}]_F^{-1}) = \det([\mathbf{F}]_F)\det([\widetilde{\mathbf{f}}(E)]_E)\det([\mathbf{F}]_F^{-1}) = \det([\widetilde{\mathbf{f}}(E)]_E)$, so $\det([\widetilde{\mathbf{f}}(F)]_F) = \det([\widetilde{\mathbf{f}}(E)]_E)$.

4.3 Orientation of finite-dimensional vector spaces

Orientation is the mathematical formalization of the notions of "clockwise" and "counterclockwise"; it is the notion which distinguishes different "rotational configurations" from each other.

Our discussion of orientation will be as follows. First, we treat inner product spaces, so that we can speak of orthonormality; we will define an orientation on an inner product space to be a choice of an *ordered* orthonormal basis. Then, in order to complete the definition of orientation for inner product spaces, we introduce rotations in n-dimensions. After we finish the definition of orientation for inner product spaces, we end the subsection on oriented inner product spaces by presenting the fact that the determinant "tracks" orientation. This fact allows us to generalize the notion of orientation to finite dimensional vector spaces that may or may not have an inner product. Lastly, we show how the top exterior power of a finite-dimensional vector space can be used for the purposes of orientation.

First notions of orientation for inner product spaces

Definition 4.40. (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 4.41. (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 4.42. (Intuition for the antisymmetry of ordered bases).

Consider the plane $\mathbb{R}^2$, and consider also two permutations of the standard ordered basis $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ for $\mathbb{R}^2$: $\{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ and $\{\hat{\mathfrak{e}}_1, -\hat{\mathfrak{e}}_2\}$. (Draw these ordered bases out on paper). Notice that no matter how you rotate the entire second ordered basis (rotate each vector in the second ordered basis by the same amount), it is impossible to make all vectors from the second ordered basis simultaneously align with their counterparts from the first ordered basis. This is also impossible for the ordered bases $\{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ and $\{\hat{\mathfrak{e}}_2, \hat{\mathfrak{e}}_1\}$. Finally, consider the ordered bases $\{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2\}$ and $\{-\hat{\mathfrak{e}}_2, \hat{\mathfrak{e}}_1\}$ of $\mathbb{R}^2$. It *is* possible to make each vector from the first ordered basis with its counterpart from the second ordered basis by rotating either the entire first ordered basis or the entire second ordered basis.

What we have discovered is that *swapping adjacent vectors in an ordered basis of two vectors produces an ordered basis that is* equivalent under rotation *to the ordered basis obtained from the original by negating one of the vectors that have been swapped.* We refer to this fact as the *antisymmetry of ordered bases.* The notion of “equivalent under rotation” is formalized in the following definition.

We now work to formalize the antisymmetry of ordered bases by stating it in the context of a 2-dimensional inner product space. The notion of *rotational equivalence* is what facilitate this formalization. Before we define rotational equivalence, however, we must define what a 2-rotation is.

Definition 4.43. (2-rotation).
Let V be a 2-dimensional inner product space, and let $\widehat{U}$ be an orthonormal ordered basis for V. A *2-rotation on the 2-dimensional inner product space* V is a linear function $\mathbf{R}_\theta : V \to V$ whose matrix relative to $\widehat{U}$ and $\widehat{U}$ is

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

for some $\theta \in [0, 2\pi)$.
Now consider the case when V is an n-dimensional inner product space, $n > 2$. Suppose $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ is an orthonormal ordered basis for V. An *extension of a 2-rotation on the oriented subspace*[4] *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 $\widehat{U}$ and $\widehat{U}$ is, for some $i, j \in \{1, ..., n\}$,

$$\begin{array}{c} \begin{array}{cccccc} & & i\text{th column} & & j\text{th column} & \end{array} \\ \begin{pmatrix} 1 & \dots & \cos(\theta) & 0 & -\sin(\theta) & 0 \\ 0 & & 0 & \vdots & 0 & \vdots \\ 0 & & \vdots & 1 & \vdots & \vdots \\ \vdots & & 0 & \vdots & 0 & \vdots \\ 0 & \dots & \sin(\theta) & 0 & \cos(\theta) & 1 \end{pmatrix} \end{array}.$$

(The columns other than the ith and jth columns are the columns of the $n \times n$ identity matrix).
If $\mathbf{R}_\theta$ is a 2-rotation on a 2-dimensional inner product space or is an extension of a 2-dimensional rotation on an n-dimensional inner product space, it is simply called a *2-dimensional rotation.* 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 4.44. (Equivalence under rotation for 2-dimensional inner product spaces).
We define orthonormal ordered bases $\widehat{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ and $\widetilde{U} = \{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2\}$ of a 2-dimensional inner product space V to be *equivalent under rotation* iff there exists a 2-dimensional rotation $\mathbf{R}_\theta$ for which $\widetilde{U} = \mathbf{R}_\theta(\widehat{U})$. (Recall Definition 1.40 for the meaning of $\mathbf{R}_\theta \widehat{U}$)).

Theorem 4.45. (Antisymmetry of ordered bases for a 2-dimensional inner product space).
Now we see how the notion of rotational equivalence for 2-dimensional inner product spaces formalizes the antisymmetry of ordered bases. Let $\widehat{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ be an orthonormal ordered basis of a 2-dimensional inner product space V. When $\theta = -\frac{\pi}{2}$ or $\theta = \frac{\pi}{2}$, the matrix of $\mathbf{R}_\theta$ relative to $\widehat{U}$ and $\widehat{U}$ is

[4]When we say “oriented subspace”, we mean that the orientation of $\text{span}(\mathbf{u}_i, \mathbf{u}_j)$ is given by the ordered basis $\{\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j\}$.

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

Thus, as we noticed in the informal discussion (Derivation 4.42), 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 4.46. (Equivalence under rotation for n-dimensional inner product spaces).

We define orthonormal ordered bases $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ and $\widetilde{U} = \{\tilde{\mathbf{u}}_1, ..., \tilde{\mathbf{u}}_n\}$ of an n-dimensional inner product space V to be *equivalent under rotation*, and thus write $\widehat{U} \sim \widetilde{U}$, iff there is a composition of 2-dimensional rotations $\mathbf{R} = \mathbf{R}_k \circ ... \circ \mathbf{R}_1$ defined on some oriented subspaces $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ for which $\mathbf{R}(\widehat{U}) = \widetilde{U}$. This notion of "equivalence under rotation" is indeed an equivalence relation on orthonormal ordered bases of V.

Theorem 4.47. (Antisymmetry of ordered bases for an n-dimensional inner product space).

Similarly to what was done in the previous derivation, we use $\theta = -\frac{\pi}{2}$ and $\theta = \frac{\pi}{2}$ in the matrix relative to bases of a 2-dimensional rotation defined on the oriented subspace $\text{span}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)$ to obtain the formal statement of the antisymmetry of ordered bases: for any orthonormal ordered basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ of V, we have

$$\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}, ..., \hat{\mathbf{u}}_n\} \sim -\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_{i+1}, \hat{\mathbf{u}}_i, ..., \hat{\mathbf{u}}_n\}.$$

Equivalently, for any orthonormal ordered basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ of V and any permutation $\sigma \in S_n$,

$$\{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}, ..., \hat{\mathbf{u}}_n\} \sim \text{sgn}(\sigma)\{\hat{\mathbf{u}}_{\sigma(1)}, ..., \hat{\mathbf{u}}_{\sigma(n)}\}.$$

Definition 4.48. (Orientation of permuted ordered bases).

Let V be an n-dimensional inner product space, and fix an orthonormal ordered basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ for V. Suppose $\widehat{U}^\sigma = \{\hat{\mathbf{u}}_{\sigma(1)}, ..., \hat{\mathbf{u}}_{\sigma(n)}\}$, where $\sigma \in S_n$, is a permutation of $\widehat{U}$ that is not rotationally equivalent to $\widehat{U}$ (so choose any σ with $\text{sgn}(\sigma) = -1$). By the antisymmetry of ordered bases, any other permutation $\widehat{U}^\pi = \{\hat{\mathbf{u}}_{\pi(1)}, ..., \hat{\mathbf{u}}_{\pi(n)}\}$, $\pi \in S_n$, of $\widehat{U}$ is rotationally equivalent either to $\widehat{U}$ or to $\widehat{U}^\sigma$. In other words, there are only two equivalence classes[5] of "equivalence under rotation".

We can now begin to set up the notion of orientation. An *orientation for the n-dimensional inner product space V* is a choice of an orthonormal ordered basis $\widehat{U}$ for V. Iff V is given the orientation $\widehat{U}$, then the *orientation of a permutation of $\widehat{U}$ (relative to $\widehat{U}$)* is then defined to be *positive* iff that permutation of $\widehat{U}$ is rotationally equivalent to $\widehat{U}$, and is defined to be *negative* otherwise. Per the previous paragraph, every permutation of $\widehat{U}$ is either positively oriented or negatively oriented relative to $\widehat{U}$.

Remark 4.49. (The formalization of "counterclockwise" and "clockwise").

At the beginning of this section, we said that orientation would formalize the notions of "clockwise" and "counterclockwise". This formalization has been achieved by the previous definition.

[5] I find this relatively surprising. My intuition is that there would be something like 2^n or $n!$ equivalence classes of "equivalence under rotation" in n dimensions, but nope! There are 2 equivalence classes of "equivalence under rotation" for every n.

A *counterclockwise rotational configuration* is another name for the orientation given to $\mathbb{R}^3$ by the standard basis, *when we use the normal* human *convention* of drawing the ordered basis $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2, \hat{\mathfrak{e}}_3\}$ such that $\hat{\mathfrak{E}}$ can be rotated so that $\hat{\mathfrak{e}}_1$ points out of the page, $\hat{\mathfrak{e}}_2$ points to the right, and $\hat{\mathfrak{e}}_3$ points upwards on the page. In this visual convention, the direction of each basis vector corresponds to its position in $\hat{\mathfrak{E}}$.

A *clockwise rotational configuration* then corresponds to the ordered bases which are not rotationally equivalent to $\hat{\mathfrak{E}}$. One such ordered basis, $\{-\hat{\mathfrak{e}}_1, \hat{\mathfrak{e}}_2, \hat{\mathfrak{e}}_3\}$ can be depicted using the visual convention just established by drawing $\hat{\mathfrak{e}}_1$ as pointing into the page (i.e. $-\hat{\mathfrak{e}}_1$ points out of the page), $\hat{\mathfrak{e}}_2$ as pointing to the right, and $\hat{\mathfrak{e}}_3$ as pointing upwards on the page.

We could have easily picked a different visual convention (i.e. a different permutation of in/out, left/right, up/down) to represent the ordering of the basis that is considered to orient the space.

At this point, we need some definitions and facts about n-rotations before we complete our development of orientation.

Rotations in n-dimensions

Definition 4.50. (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 4.51. Every n-rotation is an orthogonal linear function.

Proof. Every n-rotation is a composition of 2-rotations, which are orthogonal linear functions. □

Lemma 4.52. (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 $\widetilde{U} = \{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2\}$ and $\widehat{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ are orthonormal ordered bases of $\widetilde{W}$ and W, respectively, then there exists a 3-rotation which takes $\widetilde{U}$ to $\widehat{U}^\sigma$ for some $\sigma \in S_2$.

Proof. By definition, a 3-rotation is of the form $\mathbf{R} = \mathbf{R}_\gamma \circ \mathbf{R}_\beta \circ \mathbf{R}_\alpha$, where $\mathbf{R}_\alpha, \mathbf{R}_\beta$, and $\mathbf{R}_\gamma$ are 2-rotations. (Sidenote: $\alpha, \beta, \gamma \in [0, 2\pi)$ are called *Euler angles*).

We must show that there exist α, β, γ for which $\mathbf{R}(\widetilde{U}) = \widehat{U}$. We choose α, β, γ such that $\mathbf{R}(\tilde{\mathbf{u}}_1) = \hat{\mathbf{u}}_1$. 2-rotations are orthogonal linear functions, so they preserve the orthonormality of bases. Thus, $\mathbf{R}(\tilde{\mathbf{u}}_2)$ must have length 1 and be orthogonal to $\mathbf{R}(\tilde{\mathbf{u}}_1) = \hat{\mathbf{u}}_1$, so $\mathbf{R}(\tilde{\mathbf{u}}_2)$ is either $\hat{\mathbf{u}}_2$ or $-\hat{\mathbf{u}}_2$. Therefore, $\mathbf{R}(\widetilde{U}) = \{\hat{\mathbf{u}}_1, \pm\hat{\mathbf{u}}_2\}$. More formally, $\mathbf{R}(\widetilde{U}) = \widehat{U}^\sigma$ for some $\sigma \in S_2$.

□

Remark 4.53. 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 4.54. (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 $\widetilde{U} = \{\tilde{\mathbf{u}}_1, ..., \tilde{\mathbf{u}}_n\}$ and $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ are orthonormal ordered bases of $\widetilde{W}$ and W, respectively, then there is an n-rotation taking $\widetilde{U}$ to some permutation $\widehat{U}^\sigma$ of $\widehat{U}$.

Proof. When $n \in \{2, 3\}$, the previous lemma yields the desired n-rotation as a composition of 3-rotations (for $n = 2$, just take the restriction of the 3-rotations). We need to show that the theorem holds when $n > 3$. We consider the cases when n is even and n is odd.

If n is even, then, for $i \in \{1, ..., n\}$, let $\mathbf{R}_i$ be the 3-rotation taking $\{\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}\}$ to $\{\tilde{\mathbf{u}}_{\sigma_i(i)}, \tilde{\mathbf{u}}_{\sigma_i(i+1)}\}$, where $\sigma_i \in S(\{i, i+1\})$ is some permutation. Then $\mathbf{R}_{\frac{n}{2}} \circ ... \circ \mathbf{R}_1$ is a composition of 3-rotations taking $\widetilde{U}$ to $\widehat{U}^\sigma$, where $\sigma \in S_n$ is the permutation defined by $\sigma(i) = \sigma_{j(i)}(i)$, where $j(i) = i$ when i is odd and $j(i) = i - 1$ when i is even.

If n is odd, then let $\mathbf{R}$ be the 3-rotation taking $\{\hat{\mathbf{u}}_{n-2}, \hat{\mathbf{u}}_{n-1}, \hat{\mathbf{u}}_n\}$ to $\{\tilde{\mathbf{u}}_{\sigma_{n-2}(n-2)}, \tilde{\mathbf{u}}_{\sigma_{n-2}(n-1)}, \tilde{\mathbf{u}}_{\sigma_{n-2}(n)}\}$, 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 $\widetilde{U} - \{\tilde{\mathbf{u}}_{n-2}, \tilde{\mathbf{u}}_{n-1}, \tilde{\mathbf{u}}_n\}$ to $\widehat{U} - \{\hat{\mathbf{u}}_{n-2}, \hat{\mathbf{u}}_{n-1}, \hat{\mathbf{u}}_n\})^\sigma$, for some permutation $\sigma \in S_n$. Thus $\mathbf{R} \circ \mathbf{R}_{\frac{n-3}{2}} \circ ... \circ \mathbf{R}_1$ is a composition of 3-rotations taking $\widetilde{U}$ to $\widehat{U}^\pi$, where $\pi(i) = \sigma(i)$ when $i \in \{1, ..., n-3\}$ and $\pi(i) = \sigma_{n-2}(i)$ when $i \in \{n-2, n-1, n\}$. □

Completing the definition of orientation for inner product spaces

Definition 4.55. (Orientation of arbitrary ordered bases).

Let V be an n-dimensional inner product space, and fix an orthonormal basis $\widehat{U} = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ for V. We know how to "orient" ordered bases for V that happen to be permutations of $\widehat{U}$. Now, we generalize the notion of orientation so that it applies to any ordered orthonormal basis of V.

We define the *orientation of an orthonormal ordered basis E of V* that is not necessarily a permutation of $\widehat{U}$ to be the orientation of the unique permutation $\widehat{U}^\sigma$, $\sigma \in S^n$, of $\widehat{U}$ for which there exists an n-rotation taking E to $\widehat{U}^\sigma$.

Then, we define the *orientation of an arbitrary orthonormal ordered basis E of V* to be the orientation of the unique orthonormal basis $\widehat{U}_E$ obtained from performing the Gram-Schmidt process on E (see Theorem 3.19).

Theorem 4.56. (The determinant tracks orientation).

Let V be an n-dimensional inner product space with an orientation given by an orthonormal ordered basis $\widehat{U}$. Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be any ordered basis (not necessarily orthonormal) of V. We have $\det([\mathbf{E}]\widehat{U}) > 0$ iff E is positively oriented relative to $\widehat{U}$, and $\det([\mathbf{E}]\widehat{U}) < 0$ iff E is negatively oriented relative to $\widehat{U}$.

Proof. This proof has two overarching steps. First, we pass the definition of orientation for arbitrary ordered bases of V to the definition of orthonormal ordered bases of V by obtaining an orthonormal ordered basis $\widehat{U}_E$ from E. Then we pass the definition of orientation for orthonormal ordered bases of V that are not permutations of $\widehat{U}$ to the definition of orientation for orthonormal ordered bases of V that are permutations of $\widehat{U}$.

To begin the first step, consider $\det([\mathbf{E}]_{\widehat{U}}) = \det([\mathbf{e}_1]_{\widehat{U}}, ..., [\mathbf{e}_n]_{\widehat{U}})$, and perform Gram-Schmidt on $\{[\mathbf{e}_1]_{\widehat{U}}, ..., [\mathbf{e}_n]_{\widehat{U}}\}$. In the ith step of Gram-Schmidt, a linear combination of the vectors $[\mathbf{e}_1]_{\widehat{U}}, ..., \cancel{[\mathbf{e}_i]_{\widehat{U}}}, ..., [\mathbf{e}_n]_{\widehat{U}}$ is added to $[\mathbf{e}_i]_{\widehat{U}}$. Recall from Theorem 4.28 that the determinant is invariant under linearly combining input vectors into a different input vector. Therefore, performing Gram-Schmidt does not change the determinant. That is, if $\widehat{U}_E = \{\tilde{\mathbf{u}}_1, ..., \tilde{\mathbf{u}}_n\}$ is the orthonormal basis obtained by performing Gram-Schmidt on E, then

$$\det([\mathbf{E}]_{\widehat{U}}) = \det([\mathbf{e}_1]_{\widehat{U}}, ..., [\mathbf{e}_n]_{\widehat{U}}) = \det([\tilde{\mathbf{u}}_1]_{\widehat{U}}, ..., [\tilde{\mathbf{u}}_n]_{\widehat{U}}) = \det([\widehat{U}_E]_{\widehat{U}}).$$

In performing this first step of the proof, the determinant has stayed the same as we've passed from E to $\widehat{U}_E$. We now show that the determinant continues to stay the same as we pass from $\widehat{U}_E$ to some permutation $\widehat{U}^\sigma$ of $\widehat{U}$.

Theorem 4.54 says that there is a n-rotation $\mathbf{R}$ taking $\widehat{U}_E$ to $\widehat{U}^\sigma$, for some $\sigma \in S_n$, and Theorem 4.51 guarantees that $\det(\mathbf{R}) = 1$. Thus, since $\widehat{U}^\sigma = \mathbf{R}(\widehat{U}_E)$, we have

$$\det([\widehat{U}_E]_{\widehat{U}}) = \det([\mathbf{R}(\widehat{U}_E)]_{\widehat{U}}) \det([\widehat{U}_E]_{\widehat{U}}) = \det([(\mathbf{R} \circ \mathbf{I})(\widehat{U}_E)]_U) = \det([\mathbf{R}(\widehat{U}_E)]_U) = \det([\widehat{U}^\sigma]_U).$$

To conclude the proof, we will show that $\det([\widehat{U}^\sigma]_{\widehat{U}}) = \operatorname{sgn}(\sigma)\det(\widehat{U}]_{\widehat{U}}) = \operatorname{sgn}(\sigma)$. Since any permutation is a composition of "swaps" (a "swap" is a permutation defined on a two-element set), then $\widehat{U}^\sigma$ can be obtained from $\widehat{U}$ by repeatedly swapping vectors in $\widehat{U}$. Whenever vectors are swapped in the determinant, the sign of the determinant is multiplied by -1. This accounts for the $\operatorname{sgn}(\sigma)$ factor in the equation $\det([\widehat{U}^\sigma]_{\widehat{U}}) = \operatorname{sgn}(\sigma)\det(\widehat{U}]_{\widehat{U}}) = \operatorname{sgn}(\sigma)$. $\square$

We state the next theorem only for completeness. (We will not use the forward implication of the next theorem, and our last use of the reverse implication, which was already shown, was in the proof of Theorem 4.56).

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 4.57. (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 4.58. (Antisymmetry of ordered bases).

Notice that we still have the previous antisymmetry of ordered bases due to the antisymmetry of the determinant.

We now show how the top exterior power of a vector space can be used to describe orientation.

Theorem 4.59. (Orientation with top degree wedges).

Let V be a finite-dimensional vector space with an orientation $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. All positively oriented bases of V are scalar multiples of E, and all negatively oriented bases of V are scalar multiples of $-E$, where $-E = E^\sigma$ for some σ with $\text{sgn}(\sigma) < 0$.

Notice, we can identify $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ with $\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n \in \Lambda^n(V)$, because the antisymmetry of ordered bases is manifested in elements of $\Lambda^n(V)$ due to the antisymmetry $\wedge$. Once one has noticed this, it is a natural next step to check that the union of the sets of positively oriented and negatively oriented bases, when considered under the operations of "basis addition" and multiplication by a scalar, is a vector space that is isomorphic to $\Lambda^n(V)$.

Therefore, another way to give an orientation to a finite-dimensional vector space is to choose an element of $\Lambda^n(V)$. Notice also that the fact that the pushforward on the top exterior power is multiplication by the determinant (recall Theorem 4.35) 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 $\mathbf{e}_i \mapsto \mathbf{f}_i$.

4.4 Exterior powers as vector spaces of actual functions

Let V be a finite-dimensional vector space over a field K. In the second bullet point of Remark 2.30, we mentioned that $\binom{p}{q}$ tensors are often defined to be elements of $\mathcal{L}((V^*)^{\times p} \times V^{\times q} \to K)$; that is, $\binom{p}{q}$ tensors often are taken to be actual multilinear functions. This new interpretation is a result of the "more fundamental" identification of $V_1^* \otimes ... \otimes V_k^*$, where $V_1, ..., V_k$ are finite-dimensional vector spaces, with $\mathcal{L}(V_1 \times ... \times V_k \to K)$. Similarly, wedge product spaces and exterior powers can be interpreted as spaces of actual functions. We will need this interpretation in the setting of differential forms. Our two goals in this subsection are to (1) present that a k-wedge of covectors in V^* can act on k vectors from V to produce a scalar and to (2) give a new presentation of the pullback of a k-wedge of covectors.

To achieve our goals, we first formalize some notation about the alternative interpretations of tensor product spaces, $\binom{p}{q}$ tensors, wedge product spaces, and exterior powers.

Derivation 4.60. (Tensor product as an actual function).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces. Observe that the involvement of $\otimes$ in $V_1^* \otimes ... \otimes V_k^*$ induces a binary operation $\widetilde{\otimes}$ on $\mathcal{L}(V_1 \times ... \times V_k \to K)$. We construct $\widetilde{\otimes}$ by explicitly[6] constructing a natural isomorphism $V_1^* \otimes ... \otimes V_k^* \cong \mathcal{L}(V_1 \times ... \times V_k \to K)$, and giving new notation (namely, $\widetilde{\otimes}$) notation to this isomorphism. The natural isomorphism $V_1^* \otimes ... \otimes V_k^* \cong \mathcal{L}(V_1 \times ... \times V_k \to K)$ is defined on the elementary tensor $\phi^1 \otimes ... \otimes \phi^k \in V_1^* \otimes ... \otimes V_k^*$, and extended with the seeming-multilinearity of $\otimes$ and the corresponding actual multilinearity of the newly-defined $\widetilde{\otimes}$. The isomorphism sends

$$\phi^1 \otimes ... \otimes \phi^k \mapsto \phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k,$$

where $\phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k : V^{\times k} \to K$ is the multilinear function defined by

$$(\phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k)(\mathbf{v}_1, ..., \mathbf{v}_k) := \phi^1(\mathbf{v}_1)...\phi^k(\mathbf{v}_k).$$

Note, we have essentially reused the idea of the natural isomorphism from the third bullet point of Theorem 2.33 (this isomorphism is discussed in the proof of the referenced theorem).

Definition 4.61. (Tensor product spaces of actual functions).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces over a field K. We define $V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$ to be the vector space spanned by elements of the form $\phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k$, where $\phi^i \in V_i^*$. Since $V_1 \widetilde{\otimes} ... \widetilde{\otimes} V_k \cong V_1^{**} \widetilde{\otimes} ... \widetilde{\otimes} V_k^{**}$ is a natural isomorphism, $V_1 \widetilde{\otimes} ... \widetilde{\otimes} V_k$ can be identified with the vector space spanned by elements of the form $\Phi_{\mathbf{v}_1} \otimes ... \Phi_{\mathbf{v}_n}$, where $\mathbf{v}_i \in V_i$, and where $\Phi_{\mathbf{v}_i} \in V_i^{**}$ is defined by $\Phi_{\mathbf{v}_i}(\phi_i) = \phi_i(\mathbf{v}_i)$.

Note that $V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^* = \mathcal{L}(V_1 \times ... \times V_k \to K)$ and $V_1 \widetilde{\otimes} ... \widetilde{\otimes} V_k = \mathcal{L}(V_1^* \times ... \times V_k^* \to K)$.

Definition 4.62. ($\binom{p}{q}$ tensors as actual functions).

We define $\widetilde{T}_q^p(V) := (V^*)^{\widetilde{\otimes} p} \widetilde{\otimes} V^{\widetilde{\otimes} q}$. Note that there is a natural isomorphism $T_q^p(V) \cong \widetilde{T}_q^p(V)$.

Derivation 4.63. (Alternization of elements of $V_1 \otimes ... \otimes V_k$ product as producing an actual function).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces, and consider the wedge product space $V_1^* \wedge ... \wedge V_k^*$. Due to the existence of the alt function on $V_1^* \wedge ... \wedge V_k^*$, the isomorphism $V_1^* \otimes ... \otimes V_k^* \cong V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$ induces a function $\widetilde{\text{alt}} : V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^* \to V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$:

$$\widetilde{\text{alt}}(\mathbf{T}) = \frac{1}{k!} \sum_{\sigma \in S_k} \text{sgn}(\sigma) \mathbf{T}^\sigma,$$

where $(\cdot)^\sigma$ is a permutation on elements of $V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$ induced by a permutation on elements of $V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$. To be extra clear, $(\cdot)^\sigma : V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^* \to V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*$ is defined on elementary "tensors", and extended with multilinearity, by:

[6] We know that a natural isomorphism $V_1^* \otimes ... \otimes V_k^* \cong \mathcal{L}(V_1 \times ... \times V_k \to K)$ exists due to other theorems about natural isomorphisms.

$$(\phi^1 \widetilde{\otimes} ... \widetilde{\otimes} \phi^k)^\sigma = \phi^{\sigma(1)} \widetilde{\otimes} ... \widetilde{\otimes} \phi^{\sigma(k)}.$$

Derivation 4.64. (Wedge product as an actual functions).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces. The wedge product $\wedge : (V_1^* \otimes ... \otimes V_k^*) \otimes (V_1^* \otimes ... \otimes V_k^*) \to \text{alt}((V_1^* \otimes ... \otimes V_k^*) \otimes (V_1^* \otimes ... \otimes V_k^*))$ induces a wedge product $\widetilde{\wedge} : (V_1 \widetilde{\otimes} ... \widetilde{\otimes} V_k^*) \widetilde{\otimes} (V_1 \widetilde{\otimes} ... \widetilde{\otimes} V_k^*) \to \widetilde{\text{alt}}((V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*) \widetilde{\otimes} (V_1^* \widetilde{\otimes} ... \widetilde{\otimes} V_k^*))$ defined by $\mathbf{T} \widetilde{\wedge} \mathbf{S} = \widetilde{\text{alt}}(\mathbf{T} \widetilde{\otimes} \mathbf{S})$. The wedge product $\widetilde{\wedge}$ is an alternating multilinear function because $\wedge$ is antisymmetric and appears multilinear.

Definition 4.65. (Wedge product spaces of actual functions).

Let $V_1, ..., V_k$ be finite-dimensional vector spaces over a field K. We define $V_1^* \widetilde{\wedge} ... \widetilde{\wedge} V_k^*$ to be the vector space spanned by elements of the form $\phi^1 \widetilde{\wedge} ... \widetilde{\wedge} \phi^k$, where $\phi^i \in V_i^*$. Since $V_1 \widetilde{\wedge} ... \widetilde{\wedge} V_k \cong V_1^{**} \widetilde{\wedge} ... \widetilde{\wedge} V_k^{**}$ is a natural isomorphism, $V_1 \widetilde{\wedge} ... \widetilde{\wedge} V_k$ can be identified with the vector space spanned by elements of the form $\Phi_{\mathbf{v}_1} \wedge ... \Phi_{\mathbf{v}_n}$, where $\mathbf{v}_i \in V_i$, and where $\Phi_{\mathbf{v}_i} \in V_i^{**}$ is defined by $\Phi_{\mathbf{v}_i}(\phi_i) = \phi_i(\mathbf{v}_i)$.

Note that $V_1^* \widetilde{\wedge} ... \widetilde{\wedge} V_k^* = (\text{alt}\mathcal{L})(V_1 \times ... \times V_k \to K)$ and $V_1 \widetilde{\wedge} ... \widetilde{\otimes} V_k = (\text{alt}\mathcal{L})(V_1^* \times ... \times V_k^* \to K)$.

Definition 4.66. (Exterior powers of actual functions).

Let V be a finite-dimensional vector space over a field K. We know $\Lambda^k(V^*) \cong \Lambda^k(V)^*$, and the previous definition shows $\Lambda^k(V^*) \cong (V^*)^{\widetilde{\wedge} k}$. Because of this, we define $\widetilde{\Lambda}^k(V^*) := (V^*)^{\widetilde{\wedge} k}$.

Note that $\widetilde{\Lambda}^k(V^*) \cong (\text{alt}\mathcal{L})(V^{\times k} \to K)$.

We have now laid out the landscape for the interpretation of tensor product spaces and wedge product spaces as spaces of actual functions. In doing so, we described explicitly how $\widetilde{\otimes}$ acts on multilinear functions to produce a multilinear function. To complete this section, we show how $\widetilde{\wedge}$ acts on alternating multilinear functions to produce an alternating multilinear function, and present the pullback of elements of exterior powers in the context that exterior powers are thought of as spaces of actual functions.

Lemma 4.67. (Pushforward on the dual is multiplication by $\det(\phi^i(\mathbf{e}_j))$).

Let V be an n-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^*. For all $\phi^1, ..., \phi^n \in V^*$, we have

$$\phi^1 \wedge ... \wedge \phi^n = \det(\mathbf{f}^*) \phi^{\mathbf{e}_1} \wedge ... \wedge \phi^{\mathbf{e}_n},$$

because the pushforward on a top exterior power is multiplication by the determinant (recall Theorem 4.35). Since $\det(\mathbf{f}^*)$ is the determinant of the matrix of $\mathbf{f}^*$ relative to E^* and E^*, which has ij entry $([\phi_i]_{E^*})_j = \phi^i(\mathbf{e}_j)$ (recall Theorem 3.26), then $\det(\mathbf{f}^*) = \det(\phi^i(\mathbf{e}_j))$. Here, $(\phi^i(\mathbf{e}_j))$ is the matrix with ij entry $\phi^i(\mathbf{e}_j)$. Thus

$$\phi^1 \wedge ... \wedge \phi^n = \det(\phi^i(\mathbf{e}_j)) \phi^{\mathbf{e}_1} \wedge ... \wedge \phi^{\mathbf{e}_n}.$$

For some geometric meaning regarding the above statement, consider the isomorphism $\mathbf{F} : V^* \to \mathbb{R}^n$ induced[7] by the isomorphism $\mathbf{G} : V \to \mathbb{R}^n$ that sends $\mathbf{e}_i \overset{\mathbf{G}}{\mapsto} \hat{\mathfrak{e}}_i$, where $\hat{\mathfrak{E}} = \{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ is the standard basis for $\mathbb{R}^n$. The determinant in the above computes the projected volume of the parallelapiped spanned by $\mathbf{F}(\phi^1), ..., \mathbf{F}(\phi^n) \in \mathbb{R}^n$ onto the parallelapiped spanned by $\mathbf{F}(\phi^{\mathbf{e}_1}), ..., \mathbf{F}(\phi^{\mathbf{e}_n}) \in \mathbb{R}^n$.

Remark 4.68. Interestingly, the fact $\det(\mathbf{f}^*) = \det(\mathbf{f})$ does *not* come in useful for the current line of argument (the "current line of argument" started with the previous lemma).

Lemma 4.69. (Action of dual n-wedge on dual basis).

Let V be an n-dimensional vector space with basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^*. Consider an element $\phi^1 \widetilde{\wedge} ... \widetilde{\wedge} \phi^n \in \widetilde{\Lambda}^n(V^*)$. We have

$$(\phi^{\mathbf{e}_1} \widetilde{\wedge} ... \widetilde{\wedge} \phi^{\mathbf{e}_n})(\mathbf{e}_1, ..., \mathbf{e}_n) = 1.$$

[7]Specifically, if $\mathbf{H} : V \to \mathbb{R}^n$ by $\mathbf{H}(\mathbf{e}_i) = \hat{\mathfrak{e}}_i$, then $\mathbf{F} = \mathbf{H} \circ \mathbf{G}$.

Proof. Using a similar argument to the one that showed the permutation formula for the determinant on the $\phi^{\mathbf{e}_i}$, we have

$$\phi^{\mathbf{e}_1}\widetilde{\wedge}...\widetilde{\wedge}\phi^{\mathbf{e}_n} = \widetilde{\text{alt}}(\phi^{\mathbf{e}_1}\widetilde{\wedge}...\widetilde{\otimes}\phi^{\mathbf{e}_n}) = \sum_{\sigma\in S_n} \text{sgn}(\sigma)\phi^{\mathbf{e}_{\sigma(1)}}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\mathbf{e}_{\sigma(n)}}.$$

Therefore

$$(\phi^{\mathbf{e}_1}\widetilde{\wedge}...\widetilde{\wedge}\phi^{\mathbf{e}_n})(\mathbf{e}_1,...,\mathbf{e}_n) = \Big(\sum_{\sigma\in S_n} \text{sgn}(\sigma)\phi^{\mathbf{e}_{\sigma(1)}}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\mathbf{e}_{\sigma(n)}}\Big)(\mathbf{e}_1,...,\mathbf{e}_n) = \sum_{\sigma\in S_n}\Big(\text{sgn}(\sigma)(\phi^{\mathbf{e}_{\sigma(1)}}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\mathbf{e}_{\sigma(n)}})(\mathbf{e}_1,...,\mathbf{e}_n)\Big)$$

Now we focus on the inner term, $(\phi^{\mathbf{e}_{\sigma(1)}}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\mathbf{e}_{\sigma(n)}})(\mathbf{e}_1,...,\mathbf{e}_n)$. By definition of $\widetilde{\otimes}$, we have

$$(\phi^{\mathbf{e}_{\sigma(1)}}\widetilde{\otimes}...\widetilde{\otimes}\phi^{\mathbf{e}_{\sigma(n)}})(\mathbf{e}_1,...,\mathbf{e}_n) = \phi^{\mathbf{e}_{\sigma(1)}}(\mathbf{e}_1)...\phi^{\mathbf{e}_{\sigma(n)}}(\mathbf{e}_n)$$

Since $\epsilon^{\sigma(i)}(\mathbf{e}_j) = \delta^{\sigma(i)}_j$, the only permutation $\sigma \in S_n$ for which the above expression is nonzero is the identity permutation i; when $\sigma = i$, the above is 1. Thus, we have

$$(\phi^{\mathbf{e}_1}\widetilde{\wedge}...\widetilde{\wedge}\phi^{\mathbf{e}_n})(\mathbf{e}_1,...,\mathbf{e}_n) = \text{sgn}(i)\cdot 1 = 1\cdot 1 = 1.$$

□

Theorem 4.70. (Action of dual k-wedge on vectors).

Let V be an n-dimensional vector space with basis $E = \{\mathbf{e}_1,...,\mathbf{e}_n\}$, and let $E^* = \{\phi^{\mathbf{e}_1},...,\phi^{\mathbf{e}_n}\}$ be the induced dual basis for V^*. Consider an element $\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n \in \widetilde{\Lambda}^n(V^*)$. We have

$$(\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n)(\mathbf{e}_1,...,\mathbf{e}_n) = \det(\phi^i(\mathbf{e}_j)).$$

By extending with antisymmetry and multilinearity, we have

$$\boxed{(\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n)(\mathbf{v}_1,...,\mathbf{v}_n) = \det(\phi^i(\mathbf{v}_j)) \text{ for all } \mathbf{v}_1,...,\mathbf{v}_n \in V}$$

The action of $\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n$ on $\mathbf{v}_1,...,\mathbf{v}_n$ can be interpreted as follows. Recall from Lemma 4.69 that if $\mathbf{F}: V^* \to \mathbb{R}^n$ is the isomorphism induced by the isomorphism $\mathbf{G}: V \to \mathbb{R}^n$ that sends $\phi^{\mathbf{e}_i} \overset{\mathbf{G}}{\mapsto} \hat{\mathfrak{e}}_i$, the determinant in the above computes the projected volume of the parallelapiped spanned by $\mathbf{F}(\phi^1),...,\mathbf{F}(\phi^n) \in \mathbb{R}^n$ onto the parallelapiped spanned by $\mathbf{F}(\mathbf{G}(\mathbf{v}_1),...,\mathbf{F}(\mathbf{G}(\mathbf{v}_n)) \in \mathbb{R}^n$.

Proof. The first of the previous two lemmas implies $\phi^1\widetilde{\wedge}...\widetilde{\wedge}\phi^n = \det(\mathbf{f}^*)\phi^{\mathbf{e}_1}\widetilde{\wedge}...\widetilde{\wedge}\phi^{\mathbf{e}_n}$. (We have just put $\sim$'s over the 's of the first lemma). Combine this with the second of the previous two lemmas. □

So, we have shown how a k-wedge of elements from V^*, when treated as an actual function, can act on vectors from V. Lastly, we present the pushforward and pullback of elements of exterior powers when exterior powers are interpreted to be spaces of actual functions. (We are interested in the pullback, but present the pushforward for completeness).

Pushforward and pullback on spaces of actual functions

Derivation 4.71. (Pushforward on $\widetilde{T}_0^k(V)$)

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Definition 4.20 defined the pushforward $\otimes_0^k \mathbf{f} : T_0^k(V) \to T_0^k(W)$ by

$$\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k \overset{\widetilde{\otimes}_0^k \mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \otimes ... \otimes \mathbf{f}(\mathbf{v}_k).$$

The natural isomorphisms $T_0^k(V) \cong \widetilde{T}_0^k(V)$ $T_0^k(W) \cong \widetilde{T}_0^k(W)$ induce a pushforward map $\widetilde{\otimes}_0^k \mathbf{f}$: $\widetilde{T}_0^k(V) \to \widetilde{T}_0^k(W)$ defined by

$$\mathbf{v}_1 \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{v}_k \overset{\widetilde{\otimes}_0^k \mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{f}(\mathbf{v}_k).$$

Notice that by definition of $\widetilde{\otimes}$, we have $\mathbf{f}(\mathbf{v}_1) \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{f}(\mathbf{v}_k) = (\mathbf{f}^{\widetilde{\otimes} k})(\mathbf{v}_1, ..., \mathbf{v}_i)$. Therefore

$$\mathbf{v}_1 \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{v}_k \overset{\widetilde{\otimes}_0^k \mathbf{f}}{\longmapsto} \mathbf{f}^{\widetilde{\otimes} k}(\mathbf{v}_1, ..., \mathbf{v}_k).$$

Derivation 4.72. (Pullback on $\widetilde{T}_k^0(W)$).

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Definition 4.21 defined the pullback $\otimes_0^k \mathbf{f}^* : T_k^0(W) \to T_k^0(V)$ by

$$\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k \overset{\Lambda^k \mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1) \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{f}^*(\psi^k).$$

The natural isomorphisms $T^{0k}(W) \cong \widetilde{T}_k^0(W)$ and $T^{0k}(V) \cong \widetilde{T}_k^0(V)$ induce a pushforward map $\widetilde{\otimes}_k^0 \mathbf{f}^* : \widetilde{T}_k^0(W) \to \widetilde{T}_k^0(V)$ defined by

$$\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k \overset{\widetilde{\otimes}_k^0 \mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1) \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{f}^*(\psi^k) = (\psi^1 \circ \mathbf{f}) \widetilde{\otimes} ... \widetilde{\otimes} (\psi^k \circ \mathbf{f})$$

It can be checked using the definition of $\widetilde{\otimes}$ that $(\psi^1 \circ \mathbf{f}) \widetilde{\otimes} ... \widetilde{\otimes} (\psi^k \circ \mathbf{f}) = (\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k) \circ \mathbf{f}$. Therefore

$$\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k \overset{\widetilde{\otimes}_k^0 \mathbf{f}^*}{\longmapsto} (\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k) \circ \mathbf{f}.$$

The above is a statement on the elementary tensor $\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k$. Extending this statement using the multilinearity of $\widetilde{\otimes}$, we have

$$\mathbf{T} \in \widetilde{T}_k^0(W) \overset{\widetilde{\otimes}_k^0 \mathbf{f}^*}{\longmapsto} \mathbf{T} \circ \mathbf{f} \in \widetilde{T}_k^0(V) \text{ for all } \mathbf{T} \in \widetilde{T}_k^0(W).$$

Since $\mathbf{T} \circ \mathbf{f} = \mathbf{f}^*(\mathbf{T})$, this is equivalent to

$$\mathbf{T} \in \widetilde{T}_k^0(W) \overset{\widetilde{\otimes}_k^0 \mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\mathbf{T}) \in \widetilde{T}_k^0(V) \text{ for all } \mathbf{T} \in \widetilde{T}_k^0(W).$$

Since elements of $\widetilde{T}_k^0(V)$ act on k vectors from V, we now ask, how does $\otimes_k^0 \mathbf{f}^*$ act on k vectors from V? Well,

$$\otimes_k^0 \mathbf{f}^*(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{f}^*(\mathbf{T})(\mathbf{v}_1, ..., \mathbf{v}_k) = (\mathbf{T} \circ \mathbf{f})(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k)).$$

Thus $\otimes_k^0 \mathbf{f}^*$ acts on k vectors from V by

$$\boxed{\otimes_k^0 \mathbf{f}^*(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k))}$$

Derivation 4.73. (Pushforward on $\widetilde{\Lambda}^k(V)$).

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Definition 4.23 defined the pushforward $\Lambda^k \mathbf{f} : \Lambda^k(V) \to \Lambda^k(W)$ by

$$\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k \overset{\Lambda^k \mathbf{f}}{\longmapsto} \mathbf{f}(\mathbf{v}_1) \wedge ... \wedge \mathbf{f}(\mathbf{v}_k).$$

Equivalently,

$$\mathrm{alt}(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k) \overset{\Lambda^k \mathbf{f}}{\longmapsto} \mathrm{alt}(\otimes^k_0 \mathbf{f}(\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_k)).$$

Since the pushforward $\otimes^k_0 : T^k_0(V) \to T^k_0(W)$ induces the pushforward $\widetilde{\otimes}^k_0 : \widetilde{T}^k_0(V) \to \widetilde{T}^k_0(W)$, then we obtain a pushforward $\widetilde{\Lambda}^k \mathbf{f} : \widetilde{\Lambda}^k(V) \to \widetilde{\Lambda}^k(W)$ defined by

$$\mathbf{v}_1 \widetilde{\wedge} ... \widetilde{\wedge} \mathbf{v}_k = \widetilde{\mathrm{alt}}(\mathbf{v}_1 \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{v}_k) \overset{\widetilde{\Lambda}^k \mathbf{f}}{\longmapsto} \widetilde{\mathrm{alt}}(\widetilde{\otimes}^k_0 \mathbf{f}(\mathbf{v}_1 \widetilde{\otimes} ... \widetilde{\otimes} \mathbf{v}_k)) = \mathbf{f}^{\widetilde{\wedge} k}(\mathbf{v}_1, ..., \mathbf{v}_k).$$

That is,

$$\mathbf{v}_1 \widetilde{\wedge} ... \widetilde{\wedge} \mathbf{v}_k \overset{\widetilde{\Lambda}^k \mathbf{f}}{\longmapsto} \mathbf{f}^{\widetilde{\wedge} k}(\mathbf{v}_1, ..., \mathbf{v}_k).$$

Derivation 4.74. (Pullback on $\widetilde{\Lambda}^k(W^*)$).

Let V and W be n-dimensional vector spaces, and consider a linear function $\mathbf{f} : V \to W$. Definition 4.23 defined the pullback $\Lambda^k \mathbf{f}^* : \Lambda^k(W^*) \to \Lambda^k(V^*)$ by

$$\psi^1 \wedge ... \wedge \psi^k \overset{\Lambda^k \mathbf{f}^*}{\longmapsto} \mathbf{f}^*(\psi^1) \wedge ... \wedge \mathbf{f}^*(\psi^k).$$

Equivalently,

$$\mathrm{alt}(\psi^1 \otimes ... \otimes \psi^k) \overset{\Lambda^k \mathbf{f}^*}{\longmapsto} \mathrm{alt}(\otimes^0_k \mathbf{f}^*(\psi^1 \otimes ... \otimes \psi^k)).$$

Since the pullback $\otimes^0_k \mathbf{f}^* : T^0_k(W) \to T^0_k(V)$ induces the pullback $\widetilde{\otimes}^0_k \mathbf{f}^* : \widetilde{T}^0_k(W) \to \widetilde{T}^0_k(V)$, then we obtain a pullback $\widetilde{\Lambda}^k \mathbf{f}^* : \widetilde{\Lambda}^k(W^*) \to \widetilde{\Lambda}^k(V^*)$ defined by

$$\begin{aligned} \psi^1 \widetilde{\wedge} ... \widetilde{\wedge} \psi^k = \widetilde{\mathrm{alt}}(\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k) &\overset{\widetilde{\Lambda}^k \mathbf{f}^*}{\longmapsto} \widetilde{\mathrm{alt}}(\widetilde{\otimes}^0_k \mathbf{f}^*(\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k)) \\ &= \widetilde{\mathrm{alt}}((\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k) \circ \mathbf{f}) = \widetilde{\mathrm{alt}}(\psi^1 \widetilde{\otimes} ... \widetilde{\otimes} \psi^k) \circ \mathbf{f} = (\psi^1 \widetilde{\wedge} ... \widetilde{\wedge} \psi^k) \circ \mathbf{f}. \end{aligned}$$

(The equality between the rightmost expression of the first line and the leftmost expression of the second line uses the fact that $\widetilde{\otimes} \mathbf{f}^*(\mathbf{T}) = \mathbf{f}^*(\mathbf{T})$, which is presented in Derivation 4.72. The validity of the second equals sign in the second line has not been proven, but it is quickly checked).

Overall, the above line reads

$$\psi^1 \widetilde{\wedge} ... \widetilde{\wedge} \psi^k \overset{\widetilde{\Lambda}^k \mathbf{f}^*}{\longmapsto} (\psi^1 \widetilde{\wedge} ... \widetilde{\wedge} \psi^k) \circ \mathbf{f}.$$

Extending with the multilinearity and alternatingness of $\widetilde{\wedge}$, we extend this statement to

$$\mathbf{T} \in \Lambda^k(W^*) \overset{\widetilde{\Lambda}^k \mathbf{f}^*}{\longmapsto} \mathbf{T} \circ \mathbf{f} = \mathbf{f}^*(\mathbf{T}) \in \Lambda^k(V^*).$$

Therefore, $\widetilde{\Lambda}^k\mathbf{f}^* : \widetilde{\Lambda}^k(W^*) \to \widetilde{\Lambda}^k(V^*)$ acts on k vectors from V by

$$\boxed{\widetilde{\Lambda}^k\mathbf{f}^*(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k))}$$

In appearance, it seems that $\widetilde{\Lambda}^k\mathbf{f}^*$ acts on $\mathbf{v}_1, ..., \mathbf{v}_k$ exactly as does $\widetilde{\otimes}_k^0\mathbf{f}^*$, since $\widetilde{\otimes}_k^0\mathbf{f}^*(\mathbf{v}_1, ..., \mathbf{v}_k) = \mathbf{T}(\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_k))$. The distinction between the two definitions is that $\widetilde{\Lambda}^k\mathbf{f}^*$ acts on alternating multilinear maps $\mathbf{T}$, while $\widetilde{\otimes}_k^0\mathbf{f}^*$ acts on multilinear (not necessarily alternating) maps $\mathbf{T}$.

Part II

Calculus and basic topology

5

Review of calculus

This chapter presents a review of calculus, particularly multivariable calculus.

Notation for covariance and contravariance is not used in this chapter. The use of both upper and lower indices to distinguish between "covariant" and "contravariant" will not be used in the following chapter of multivariable calculus review, even though these concepts have already been introduced. Only lower indices will be used.

5.1 Notational conventions in single-variable calculus

This section formalizes common notational conventions used in single-variable calculus. These conventions ripple up to multivariable calculus, so they are worth reviewing.

First, we must establish some unambigiuous notation.

Definition 5.1. (The derivative).

$$f'(t) := \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}.$$

We use the above "prime" notation as a starting point for explaining the various ways of notating derivatives.

Definition 5.2. (Common notation for derivatives).

Suppose $U \subseteq \mathbb{R}$ is an open set and $f : U \subseteq \mathbb{R} \to \mathbb{R}$ is a differentiable function. We define the *Leibniz* and *operator* notations for differentiation.

Leibniz notation

$$\frac{df}{dt} := f'$$

$$\frac{df}{dt}\Big|_{t_0} = f'(t_0)$$

Operator notation

$$\frac{d}{dt} f := f'$$

$$\frac{d}{dt} f(t) := \Big(\frac{d}{ds} f\Big)\Big|_{s=t} = f'(t)$$

$$\frac{df(t)}{dt} := \frac{d}{dt} f(t)$$

Definition 5.3. (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, so that $f \circ x$ is differentiable. We define $\frac{df}{dx} : U \to \mathbb{R}$ to be the function defined by

$$\frac{df}{dx}\Big|_{t_0} := \frac{df}{dt}\Big|_{t=x(t_0)}.$$

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:

$$\frac{d(f \circ x)}{dt} = \frac{df(u)}{du}\frac{du(t)}{dt}, \text{ where } u = x(t).$$

Remark 5.4. (Letters in the denominator).

The definition $\frac{df}{dt} := f'$ from Definition 5.2 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.

5.2 Multivariable calculus

Lemma 5.5. (Multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}$).

Let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ be sufficiently differentiable, and set $\mathbf{x}(t_0) = \mathbf{p}$. It can be proved that

$$\frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{p}} = \frac{\partial f(\mathbf{x})}{\partial x_1}\Big|_{\mathbf{p}} \frac{dx_1}{dt}\Big|_{t_0} + ... + \frac{\partial f(\mathbf{x})}{\partial x_n}\Big|_{\mathbf{p}} \frac{dx_n}{dt}\Big|_{t_0}.$$

In other words,

$$\frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{p}} = (\underset{\mathbf{x}}{\nabla} f)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}(t)}{dt}\Big|_{t_0},$$

where we have defined the *gradient of f with respect to the function* $\mathbf{x}$ to be

$$\boxed{\nabla_{\mathbf{x}} f := \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}}$$

Note that since $\mathbf{x}(\mathbf{p}) = (x_1(\mathbf{p}), ..., x_n(\mathbf{p}))^\top$, each x_i is a function. Thus, the derivative $\frac{\partial}{\partial x_i}$ is a derivative with respect to a function, in the sense of Definition 5.3. 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 5.6. (Directional derivative of a differentiable function $\mathbb{R}^n \to \mathbb{R}$).

Consider $f : \mathbb{R}^n \to \mathbb{R}$. Let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be the curve with $\mathbf{x}(t_0) = \mathbf{p}$ and $\frac{d\mathbf{x}}{dt}\Big|_{t_0} = \mathbf{v}$. Let $\mathbf{x}$ and f be sufficiently differentiable. We define the *directional derivative* $\frac{\partial f}{\partial \mathbf{v}}$ *of f in the direction of* $\mathbf{v}$ to be

$$\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} := \frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{p}} = (\underset{\mathbf{x}}{\nabla} f)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}(t)}{dt}\Big|_{t_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v}.$$

Therefore the directional derivative is expressed as

$$\boxed{\begin{aligned} \frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} &= (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} \\ \frac{\partial f}{\partial \mathbf{v}} &= \nabla f \cdot \mathbf{v} \end{aligned}}$$

In the second line, we interpret ∇ as the function sending $\mathbf{x} \mapsto \nabla_{\mathbf{x}}$.

Most authors denote $\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$ as $D_{\mathbf{p}} f(\mathbf{v})$ or as $Df[\mathbf{v}](\mathbf{p})$.

Remark 5.7. [Lee, p. 282, 283] (Infinitesimal velocity versus infinitesimal time).

We have defined the directional derivative as the ratio

$$\frac{\text{change in } f \text{ that results from moving away from } \mathbf{x} = \mathbf{p} \text{ with instantaneous velocity } \mathbf{v}}{\text{“infinitesimal” time interval over which this moving away occurs}}.$$

This interpretation of the directional derivative motivates why we write the directional derivative as $\frac{\partial f}{\partial \mathbf{v}}$. We will continue to favor this notation because it plays most nicely with the multivariable chain rule.

It should be noted, however, that the directional derivative can also be thought of as the "infinitesimal" change in f resulting from an "infinitesimally small velocity" $\mathbf{v}$. (So, to recap, the first interpretation relies on the concept of an "infinitesimal" time interval, while the second interpretation relies on the concept of an "infinitesimal" velocity). Thinking of the directional derivative in this way leads to notating it by $D_{\mathbf{p}} f(\mathbf{v})$, since this notation is reminiscent of the expression "df". Taylor's theorem justifies this interpretation as follows: if $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, then $(\Delta_{\mathbf{v}} f)|_{\mathbf{p}} = f(\mathbf{p} + \mathbf{v}) - f(\mathbf{p})$ is well approximated as $\Delta_{\mathbf{v}} f \approx \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} ([\mathbf{v}]_{\hat{\mathfrak{E}}})_i = (\nabla_{\mathbf{x}} f) \cdot \mathbf{v} = (D_{\mathbf{p}} f)(\mathbf{v})$ when $||\mathbf{v}||$ is small.

In Theorem 7.33, we discover that the *differential* of a function f is denoted df. Evidently, differential geometry favors the "infinitesimal velocity" interpretation over the "infinitesimal time" interpretation.

Theorem 5.8. (Directional derivative as a limit).

It is sometimes useful to express the directional derivative as a limit.

To do so, first consider the curve $\mathbf{x}$ from the definition of the directional derivative. Since $\mathbf{x}(t_0) = \mathbf{p}$ and $\frac{d\mathbf{x}}{dt}\Big|_{t_0} = \mathbf{v}$, then $\mathbf{x}(t) = \mathbf{p} + \mathbf{v}t$. Thus

$$\begin{aligned}
\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} &= \frac{d(f \circ \mathbf{x})(t)}{dt}\Big|_{t_0} = \frac{df(\mathbf{p} + \mathbf{v}t)}{dt}\Big|_{t_0} = \lim_{h \to 0} \Big(\frac{f(\mathbf{p} + \mathbf{v}(t+h)) - f(\mathbf{p} + vt)}{h} \Big)\Big|_{t_0} \\
&= \lim_{h \to 0} \Big(\Big(\frac{f(\mathbf{p} + \mathbf{v}(t+h)) - f(\mathbf{p} + vt)}{h} \Big)\Big|_{t_0} \Big) = \lim_{h \to 0} \frac{f(\mathbf{p} + h\mathbf{v}) - f(\mathbf{v})}{h}.
\end{aligned}$$

Therefore

$$\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} = \lim_{h \to 0} \frac{f(\mathbf{p} + h\mathbf{v}) - f(\mathbf{v})}{h}.$$

Many authors define the directional derivative using this formula.

Theorem 5.9. (Gradient is direction of greatest increase).

Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a sufficiently differentiable function. Then, at each $\mathbf{p} \in R^3$, the gradient $(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}$ is the direction of greatest increase in f at $\mathbf{p}$. When $||\mathbf{v}|| = 1$, then $||(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}||$ is the magnitude of this greatest increase.

Proof. The previous theorem shows that $\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v}$. We know $(\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} = ||(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}||\ ||\mathrm{proj}(\mathbf{v} \to (\nabla_{\mathbf{x}} f)|_{\mathbf{p}})||$. The dot product is maximized when the projection of $\mathbf{v}$ onto $(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}$ is $\mathbf{v}$ itself. Thus, when the directional derivative is maximized, $\mathbf{v} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}}$.

The magnitude of this maximal directional derivative is $(\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} = ||(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}||\ ||\mathrm{proj}(\mathbf{v} \to (\nabla_{\mathbf{x}} f)|_{\mathbf{p}})||$. When $||\mathbf{v}|| = 1$, this reduces to $||(\nabla_{\mathbf{x}} f)|_{\mathbf{p}}||$. □

Remark 5.10. (Directional derivative simplifies to partial derivative).

We have $\frac{\partial}{\partial \hat{\mathfrak{e}}_i} = \frac{\partial}{\partial x_i}$.

Lemma 5.11. (Multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}^m$).

Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be sufficiently differentiable, and set $\mathbf{p} = \mathbf{x}(t_0)$.

$$\frac{d(\mathbf{f} \circ \mathbf{x})(t)}{dt}\Big|_{\mathbf{p}} = \begin{pmatrix} \frac{d}{dt} f_1(\mathbf{x}(t)) \\ \vdots \\ \frac{d}{dt} f_m(\mathbf{x}(t)) \end{pmatrix} = \begin{pmatrix} (\nabla_{\mathbf{x}} f_1)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}}{dt}\Big|_{t_0} \\ \vdots \\ (\nabla_{\mathbf{x}} f_m)|_{\mathbf{p}} \cdot \frac{d\mathbf{x}}{dt}\Big|_{t_0} \end{pmatrix} = \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix}\Bigg|_{\mathbf{p}} \frac{d\mathbf{x}}{dt}\Big|_{t_0}.$$

In terms of functions, we have

$$\frac{d(\mathbf{f}\circ\mathbf{x})(t)}{dt} = \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix} \frac{d\mathbf{x}}{dt}$$

Recall from Derivation 1.42 and Theorem 1.52 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 5.12. (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 5.3), we define the *Jacobian matrix* $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ to be

$$\boxed{\frac{\partial \mathbf{f}}{\partial \mathbf{x}} := \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix} = \left(\frac{\partial f_i}{\partial x_j}\right)}$$

Using the Jacobian, the multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}^m$ is now succintly stated as

$$\boxed{\frac{d(\mathbf{f}\circ\mathbf{x})}{dt} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\frac{d\mathbf{x}}{dt}}$$

Definition 5.13. (Directional derivative of a differentiable function $\mathbb{R}^n \to \mathbb{R}^m$).

The directional derivative of a differentiable function $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is defined analogously to that of a differentiable function $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}$. Indeed, in the special case of $m = 1$, the two definitions are equivalent.

As was done previously, let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be the curve with $\mathbf{x}(t_0) = \mathbf{p}$ and $\frac{d\mathbf{x}}{dt}\Big|_{t_0} = \mathbf{v}$. We define the *directional derivative* $\frac{\partial f}{\partial \mathbf{v}}$ *of* f *in the direction of* $\mathbf{v}$ to be

$$\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\Big|_{\mathbf{p}} := \frac{d(\mathbf{f}\circ\mathbf{x})}{dt}\Big|_{\mathbf{p}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{p}}\frac{d\mathbf{x}}{dt}\Big|_{t_0}$$

So this most general definition of directional derivative is expressed as

$$\boxed{\begin{aligned} \frac{\partial \mathbf{f}}{\partial \mathbf{v}}\Big|_{\mathbf{p}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{p}}\frac{d\mathbf{x}}{dt}\Big|_{t_0} \\ \frac{\partial \mathbf{f}}{\partial \mathbf{v}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\frac{d\mathbf{x}}{dt} \end{aligned}}$$

Remark 5.14. (Linearity with respect to the $\mathbf{v}$ in the denominator).

The directional derivative $\frac{\partial \mathbf{f}}{\partial \mathbf{v}}$ is linear with respect to $\mathbf{v}$.

A technical theorem

Theorem 5.15. (The integral is "linear" with respect to the region of integration).

If $D = \cup_{i=1}^{k} D_i$, where $D_1, ..., D_k$ are domains of integration such that $D_i \cap C_j$ has measure zero[1] for all $i \neq j$, then

$$\int_D f = \sum_{i=1}^{k} \int_{D_i} f.$$

[1]Informally, a subset of $\mathbb{R}^n$ has *measure zero* iff its volume is zero.

6

Basic topology

In this chapter, we present the basic concepts of *topology*. Topology is useful to us because it generalizes ideas from calculus such as *convergence* and *continuity* to an abstract setting where no concept of distance is required; instead of "distance", topology relies on the concept of what are called *open sets*. Primarily, we require topology in order to use the topological concept of an *n-manifold*, which is a *topological space* that "looks like $\mathbb{R}^n$" at each point.

We just said that topology generalizes ideas of calculus. More accurately, topology generalizes the ideas of *real analysis*, which is the technical framework which calculus relies upon. As the name implies, real analysis is essentially the study of the real numbers $\mathbb{R}$. As the is not assumed to have a real analysis background or a topology background, we will introduce real analysis concepts and their generalizations in topology simultaneously. In my opinion, it is actually best to learn the introductory concepts of real analysis and topology at the same time, as the general apparatus of topology, not being bogged down by the nitty-gritty details of real analysis, helps one more quickly realize how concepts interact.

The standard topology on $\mathbb{R}$

Open sets, closed sets and their characterizations

We begin our investigation of the basic concepts of topology- *open sets* and *closed sets*- by starting in the context of the real numbers $\mathbb{R}$.

Definition 6.1. (Open set in $\mathbb{R}$).

An *open set in* $\mathbb{R}$ is an arbitrary union of open intervals in $\mathbb{R}$.

Recall, an open interval in $\mathbb{R}$ is a set of the form $(a, b) = \{x \mid a < x < b,\ a, b \in \mathbb{R}\}$. By "arbitrary union", we mean that there is no restriction on the cardinality of the sets in the union; sets in the union can be either finite, countably infinite, or uncountably infinite.

Definition 6.2. (Interior points and interior of $A \subseteq \mathbb{R}$).

Let $A \subseteq \mathbb{R}$ be any subset of $\mathbb{R}$. A point $x \in A$ is called an *interior point of* A iff there exists an open set contained in A that contains x. Symbolically, $x \in A$ is an interior point iff $\exists$open $U_x \ni x$ s.t. $U_x \subseteq A$.

We define the *interior* of A to be the set of all interior points of A,

$$\text{int}(A) := \{x \in A \mid x \text{ is an interior point of } A\}.$$

Theorem 6.3. (Interior point characterization of open sets in $\mathbb{R}$).

Let $A \subseteq \mathbb{R}$ be any subset of $\mathbb{R}$. Then A is open iff each $x \in A$ is an interior point. Symbolically, A is open iff $\forall x \in A$ $\exists$open $U_x \ni x$ s.t. $U_x \subseteq A$.

Equivalently, A is open iff $A = \text{int}(A)$. (Every set A satisfies $\text{int}(A) \subseteq A$. So $A = \text{int}(A)$ iff $A \subseteq \text{int}(A)$).

Proof. We show that (A is open) $\iff$ (each $x \in A$ is an interior point).

($\implies$). If A is open, then every $x \in A$ is contained in an open set $U_x \subseteq A$, namely, $U_x = A$ for all x.

($\impliedby$). If every $x \in A$ is contained in an open set $U_x \subseteq A$, then $A = \cup_{x \in A} U_x$. (We have $\cup_{x \in A} U_x \subseteq A$ since each $U_x \subseteq A$, and $A \subseteq \cup_{x \in A} U_x$ since each $x \in A$ is in U_x). We defined an open set in $\mathbb{R}$ to be an arbitrary union of open sets in $\mathbb{R}$, so A is open. □

Next, see what the interior point characterization of open sets implies for complements of open sets.

Derivation 6.4. (Limit points in $\mathbb{R}$).

Each of the lines in the following derivation are logically equivalent. Skim this derivation, but don't try to understand it on the first pass- some new notation that is used in the derivation is defined after the derivation!

$$
\begin{aligned}
&U \text{ is open} \\
U = \text{int}(U) \iff U \subseteq \text{int}(U) \iff &\forall x \in U \;\; x \in \text{int}(U) \\
&\forall x \in U \;\exists \text{open } U_x \ni x \text{ s.t. } U_x \subseteq U \\
&\forall x \in U \;\exists \text{open } U_x \ni x \text{ s.t. } U_x \cap U^c = \emptyset \\
&\forall x \; x \notin U^c \implies \text{not}(\forall \text{open } U_x \ni x \;\; U_x \cap U^c \neq \emptyset) \\
&\forall x \; x \notin U^c \implies \text{not}(\forall \text{open } U_x \ni x \;\; U_x \cap U^c - \{x\} \neq \emptyset) \\
&\forall x \; x \notin U^c \implies x \notin (U^c)' \\
&\forall x \; x \in (U^c)' \implies x \in U^c \\
&(U^c)' \subseteq U^c.
\end{aligned}
$$

Note that line 6 follows from line 5 because we can subtract x out of $U_x \cap U^c$ due to the hypothesis "$x \notin U^c$". Line 6 implies line 5 for the same reason.

In line 7, we define the notion of a limit point of a subset of $\mathbb{R}$. We say $x \in \mathbb{R}$ is a *limit point of* $A \subseteq \mathbb{R}$ iff $\forall$open U_x $\;U_x \cap U^c - \{x\} \neq \emptyset$. That is, $x \in \mathbb{R}$ is a *limit point of* $A \subseteq \mathbb{R}$ iff every open set containing x intersects A at a point other than x. The set of limit points of A is denoted A'.

The above shows that U is open iff $(U^c)' \subseteq U^c$. In words, a set is open iff its complement contains all of its limit points. For this reason, we define a *closed set in* $\mathbb{R}$ to be any set which is the complement of an open set, or, equivalently, any set which contains all of its limit points. We call the fact that C is closed iff $C' \subseteq C$ the *limit point characterization of closed sets*. We repeat these definitions below.

Definition 6.5. (Limit point in $\mathbb{R}$).

Let $A \subseteq \mathbb{R}$ be any subset of $\mathbb{R}$. We say $x \in \mathbb{R}$ is a *limit point of* $A \subseteq \mathbb{R}$ iff every open set containing x intersects A at a point other than x. Symbolically, $x \in \mathbb{R}$ is a *limit point of* $A \subseteq \mathbb{R}$ iff $\forall$open U_x $\;U_x \cap U^c - \{x\} \neq \emptyset$.

The set of limit points of A is denoted A'.

Definition 6.6. (Closed set in $\mathbb{R}$).

A subset $A \subseteq \mathbb{R}$ of $\mathbb{R}$ is said to be *closed* iff A is the complement of some open subset of $\mathbb{R}$. Equivalently, $A \subseteq \mathbb{R}$ is closed iff A contains all of its limit points.

Remark 6.7. (Open vs. closed).

Subsets of $\mathbb{R}$ can be open but not closed, closed but not open, both and closed, or neither open nor closed. (This is true for general topological spaces as well).

Theorem 6.8. (Limit point characterization of closed sets in $\mathbb{R}$).

A subset $A \subseteq \mathbb{R}$ is closed iff A contains all of its limit points.

Remark 6.9. (Characterizations of open sets and closed sets in $\mathbb{R}$).

We have seen that open sets in $\mathbb{R}$ are characterized by the "interior point characterization of open sets", while closed sets in $\mathbb{R}$ are characterized by the "limit point characterization of closed sets".

6.1 Topological spaces

To discover the general apparatus of topology, we investigate unions and intersections of open and closed sets in $\mathbb{R}$. It quickly follows from the definition of an open set as an arbitrary union of open intervals that an arbitrary union of open sets is an open set. DeMorgan's laws (recall [...]) then imply that an arbitrary intersection of closed sets is a closed set.

What about intersections of open sets- or, equivalently, by DeMorgan's laws- unions of closed sets in $\mathbb{R}$? Well, any infinite union of closed sets is not necessarily closed: consider $\cup_{i=1}^{\infty}\{\frac{1}{n}\} = \{\frac{1}{n} \mid n \in \mathbb{N}\}$, which does not contain its limit point of 0. Perhaps at least finite unions of closed sets must be closed? This is indeed the case for closed subsets of $\mathbb{R}$.

The following definition of a topological space is motivated by these properties of unions and intersections of open and closed sets in $\mathbb{R}$. Before this definition, we need some more technical language.

Definition 6.10. (Covers and generating covers).

Let X be any set. A set $\mathcal{C}$ of subsets of X is a *cover* of X iff $\cup_{C \in \mathcal{C}} C = X$.

Let τ be any set. A set $\mathcal{C}$ *generates* τ iff each $U \in \tau$ is an arbitrary union of the elements of $\mathcal{C}$; that is, iff each $U \in \tau$ is $U = \cup_{\alpha \in I} C_\alpha$, $\{C_\alpha\} \subseteq \mathcal{C}$.

We can now define topological spaces.

Definition 6.11. (Topological space).

Let (X, τ) be a tuple, where X is any set and τ is a set of subsets of X. We say that (X, τ) is a *topological space*, and that τ is a *topology* on X, iff...

1. There is a cover $\mathcal{B}$ of X that generates τ.
    - For the same reasons as earlier, when we had $X = \mathbb{R}$ and $\tau = \{\text{open sets} \subseteq \mathbb{R}\}$, (1) is equivalent to the interior point characterization of open sets, which is in turn equivalent to the limit point characterization of closed sets. Interior points and limit points will be defined analogously as to how they were before.
    - This requirement is equivalent to: arbitrary unions of open sets are open $\iff$ arbitrary intersections of closed sets are closed.
2. Finite unions of closed sets are closed $\iff$ finite intersections of open sets are open.

We interpret the elements of τ as being open sets. Formally, we say that $U \subseteq X$ is an *open set* iff $U \in \tau$.

Remark 6.12. ($\mathbb{R}^n$ with the standard topology).

In the section "Topology on $\mathbb{R}$", we were really doing topology on $(\mathbb{R}, \text{std})$, where std is the *standard topology on* $\mathbb{R}$, std := {subsets of $\mathbb{R}$ that are arbitrary unions of open intervals in $\mathbb{R}$}. In general, the *standard topology on* $\mathbb{R}^n$ is {subsets of $\mathbb{R}^n$ that are arbitrary unions of open intervals in $\mathbb{R}^n$}.

Definition 6.13. (Topological space).

The above definition of a topology can quickly be seen to be equivalent to the following most common definition, which makes no mention of a generating cover. The most common definition requires τ to satisfy the following:

3. Arbitrary unions of open sets are open.
    - As noted in (1), we have that (1) and (3) are equivalent.
4. Finite intersections of open sets are open.

Remark 6.14. ($\emptyset, X \in \tau$).

Most definitions of a topological space explicitly require that the empty set and X are open subsets of X. This requirement is implied by the above definition. X is generated by τ by definition. It's less easy to see that $\emptyset$ is also generated by τ, but this is true too. The empty set is the empty union (which employs an empty indexing set) of any collection of open sets.

Bases

Our first definition of a topological space (X, τ) involved a cover $\mathcal{B}$ that generated X. There was also another condition- we required finite intersections of open sets to be open. We would like to impose additional requirements on $\mathcal{B}$ so that a topological space (X, τ) can be said to be generated by $\mathcal{B}$, without any other requirements on (X, τ) being necessary. In other words, we want to find requirements on $\mathcal{B}$ such that finite intersections of open sets are guaranteed to be open just by the virtue of the requirements on $\mathcal{B}$. We would also like for these requirements on $\mathcal{B}$ to be not just sufficient for this finite intersection property, but also necessary. The following theorem gives us the answer we want and need.

Theorem 6.15. (Finite intersections of open sets are open iff the cover "refines with interior points").

If $\mathcal{B}$ is a cover that generates τ, then finite intersections of open sets are open iff $\mathcal{B}$ *refines with interior points*. We say that $\mathcal{B}$ *refines with interior points*, that is, iff
$\forall B_1, B_2 \in \mathcal{B} \;\; x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B}$ s.t. $x \in B_3$.

Proof.
$(\implies)$. Assume a finite union of closed sets is closed. Then by DeMorgan's laws, any finite intersection of open sets is open. We must show that $\mathcal{B}$ refines with interior points. Basis elements are by definition open, so if $B_1, B_2 \in \mathcal{B}$, then $B_1 \cap B_2$ is a finite intersection of open sets, and is thus open. By the interior point characterization of open sets, $x \in B_1 \cap B_2$ implies there is some open set $U \ni x$ such that $U \subseteq B_1 \cap B_2$. Because U is open, then it is a union of basis elements, so $x \in B_3$ for some $B_3 \in \mathcal{B}$.

$(\impliedby)$. Assume $\mathcal{B}$ is a basis; that is, $\mathcal{B}$ is a cover that refines with interior points. We want to show that finite unions of closed sets are closed. By DeMorgan's laws, we can instead show that finite intersections of open sets are open.

So, set $V = \cap_{i=1}^n U_i$, 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. □

Definition 6.16. (Basis for a topological space).

A cover $\mathcal{B}$ for X is called a *basis* iff $\mathcal{B}$ refines with interior points, that is, iff
$\forall B_1, B_2 \in \mathcal{B} \;\; x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B}$ s.t. $x \in B_3$.

Theorem 6.17. Every basis generates a topological space, and every topological space has a basis.

Proof. For the first part of the theorem, recall that the whole point of defining a basis was to find the types of generating covers that generate topological spaces.

For the second part, note that if (X, τ) is a topological space, then τ is a basis for X. □

Basic facts about topological spaces

We now state the generalizations of the topological results we found in the context of $\mathbb{R}$. We do not state any proofs in this section, as the arguments given in the "Topology in $\mathbb{R}$" section all generalize easily.

Definition 6.18. (Open set).

Let (X, τ) be a topological space. As was mentioned previously, an *open set* in X is an element of τ. We often refer to open sets in X as simply "open sets".

Open sets are also often called *neighborhoods*. A *neighborhood of a point* $x \in X$ is an open set which contains x.

Definition 6.19. (Interior points and interior).

Let (X, τ) be a topological space, and $A \subseteq X$. A point $x \in A$ is called an *interior point of* A iff there exists an open set contained in A that contains x. Symbolically, $x \in A$ is an interior point iff $\exists$open $U_x \ni x$ s.t. $U_x \subseteq A$.

We define the *interior* of A to be the set of all interior points of A,

$$\text{int}(A) := \{x \in A \mid x \text{ is an interior point of } A\}.$$

Definition 6.20. (Interior point characterization of open sets).

Let (X, τ) be a topological space, and let $A \subseteq X$. Then A is open iff each $x \in A$ is an interior point. Symbolically, A is open iff $\forall x \in A\ \exists \text{open } U_x \ni x$ s.t. $U_x \subseteq A$.

Equivalently, A is open iff $A = \text{int}(A)$. (Every set A satisfies $\text{int}(A) \subseteq A$. So $A = \text{int}(A)$ iff $A \subseteq \text{int}(A)$).

Definition 6.21. (Limit point).

Let (X, τ) be a topological space, and let $A \subseteq X$. We say $x \in A$ is a *limit point of* A iff every open set containing x intersects A at a point other than x. Symbolically, $x \in \mathbb{R}$ is a *limit point of* $A \subseteq \mathbb{R}$ iff $\forall \text{open } U_x \quad U_x \cap U^c - \{x\} \neq \emptyset$.

The set of limit points of A is denoted A'.

Definition 6.22. (Closed set).

Let (X, τ) be a topological space. A subset $A \subseteq X$ is said to be *closed* iff A is the complement of some open set. Equivalently, $A \subseteq X$ is closed iff A contains all of its limit points.

Theorem 6.23. (Limit point characterization of closed sets).

Let (X, τ) be a topological space. A subset $A \subseteq X$ is closed iff A contains all of its limit points.

The subspace topology

The following definition and theorem justify what we would might naturally assume the phrase "open set in $Y \subseteq X$"means.

Definition 6.24. (Subspace topology).

Let (X, τ_X) be a topological space, and let $Y \subseteq X$. The *subspace topology* τ_Y is defined to be the topology on Y whose open sets are $\tau_Y := \{U \cap Y \mid U \text{ is open in } X\}$.

If $A \subseteq Y$, then we say that A is *open in* Y iff $A \in \tau_Y$, and we say that A is *closed in* Y iff $X - A$ is open in Y.

Theorem 6.25. A set A is closed in Y iff $A = B \cap Y$, where B is closed in X.

Proof.

($\implies$). Suppose A is closed in Y. Then $Y - C$ is open in Y, so $Y - A = U \cap Y$ for some U open in X. Thus $A = Y - (Y - A) = Y - (Y \cap U) = Y - U = Y \cap (X - U)$. So, $A = (X - U) \cap Y$, where $X - U$ is closed in X.

($\impliedby$). Suppose $A = B \cap Y$, where B is closed in X. Now, we show that $Y - A$ is open in Y. We have $Y - A = Y - (B \cap Y) = Y \cap (X - B)$. Thus $Y - A = (X - B) \cap Y$, where $X - B$ is open in X, meaning $(X - B) \cap Y$ is open in Y. $\square$

Interior, closure, and boundary

For the purposes of differential forms, understanding this section in depth is not necessary. Some familiarity with this section on a intuitive level is required, though.

Theorem 6.26. (Equivalent definitions of interior).
Let (X, τ) be a topological space. The *interior* of $A \subseteq X$ is

- $\text{int}(A) := \{x \in A \mid x \text{ is an interior point of } A\}$, by definition.
- The "largest" open set contained in A; that is, the union of all open sets contained in A.

Proof. We need to show

$$\{x \in A \mid x \text{ is an interior point of } A\} = \bigcup_{U \subseteq X, U \in \tau} U.$$

The proof of this is quick: $x \in A$ is an interior point of A iff there exists an open set $U_x \subseteq A$ containing x iff x is in the union of open sets contained in X. □

Definition 6.27. (Closure).
Let (X, τ) be a topological space. The *closure* $\text{cl}(A)$ of $A \subseteq X$ is the "smallest" closed set that contains A; that is, it is the intersection of all closed sets containing A.

Theorem 6.28. (Condition for being in the closure).
Let (X, τ) be a topological space, and let $A \subseteq X$. We have

$$\begin{aligned} x \in \text{cl}(A) &\iff \forall \text{closed } C \supseteq A \; x \in C \\ &\iff \forall \text{open } U \; U \cap A = \emptyset \implies x \notin U \\ &\iff \forall \text{open } U \ni x \; U \cap A \neq \emptyset. \end{aligned}$$

In the second line, we've used $U = C^c$. The third line follows from the second because
$(P \text{ and not } Q) \iff (Q \implies P)$.
In all, we have

$$x \in \text{cl}(A) \iff \forall \text{open } U \ni x \; U \cap A \neq \emptyset.$$

Remark 6.29. (Being in the closure vs. being a limit point).
Notice that the condition for being in the closure is just a little bit weaker than the condition for being a limit point. Recall, $x \in A$ is a limit point of A iff $\forall$open $U \ni x \; U \cap A - \{x\} \neq \emptyset$.

Definition 6.30. (Boundary).
Let (X, τ) be a topological space. The *boundary* ∂A of $A \subseteq X$ is defined to be $\partial A := \text{cl}(A) - \int(A)$.

6.2 Continuous functions and homeomorphisms

Limits and continuity for functions $A \subseteq \mathbb{R} \to \mathbb{R}$

Definition 6.31. (Open balls).

We denote the *open ball in* $\mathbb{R}^n$ *of radius* r *centered at* $\mathbf{x} \in \mathbb{R}^n$ by $B(r, x)$: $B(r, \mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n \mid d(\mathbf{x}, \mathbf{y}) < r\}$, where $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|$, and where $|| \cdot ||$ is the norm on $\mathbb{R}^n$ induced by the dot product on $\mathbb{R}^n$.

Definition 6.32. (Limit of a function $A \subseteq \mathbb{R} \to \mathbb{R}$).

Let $A \subseteq \mathbb{R}$, and consider a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$. We write $\lim_{x \to x_0} f(x) = L$ iff $\forall \epsilon > 0\ \exists \delta > 0$ s.t. $x \in B(\delta, x_0) - \{x_0\} \implies f(x) \in B(\epsilon, L)$.

Definition 6.33. (Continuity of a function $A \subseteq \mathbb{R} \to \mathbb{R}$ at a point).

Let $A \subseteq \mathbb{R}$, and consider a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$. We say that f *is continuous at* $x_0 \in A$ iff $\forall \epsilon > 0\ \exists \delta > 0$ s.t. $x \in B(\delta, x_0) \implies f(x) \in B(\epsilon, f(x_0))$

Theorem 6.34. (Condition for continuity for functions $A \subseteq \mathbb{R} \to \mathbb{R}$).

Let $A \subseteq \mathbb{R}$. A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is continuous at $x_0 \in A$ iff $\lim_{x \to x_0} f(x) = f(x_0)$.

Proof.

($\implies$). If f is continuous at x_0, then we immediately know $\lim_{x \to x_0} f(x) = f(x_0)$ because $x \in B(\delta, x_0)$ is a weaker condition than $x \in B(\delta, x_0) - \{x_0\}$.

($\impliedby$). Suppose that $\forall \epsilon > 0\ \exists \delta > 0$ s.t. $x \in B(\delta, x) \cap A - \{x_0\} \implies f(x) \in B(\epsilon, f(x_0))$. We need to show that when $x = x_0$ we have $f(x) \in B(\epsilon, f(x_0))$. But this follows immediately because $f(x_0) \in B(\epsilon, f(x_0))$ for any ϵ, as $f(x_0) = f(x_0)$.

□

Limits and continuity for functions on subsets of topological spaces

Definition 6.35. (Limit of a function on a subset of a topological space).

Let X and Y be topological spaces, let $A \subseteq X$, and consider a function $f : A \subseteq X \to Y$. We write $\lim_{x \to x_0} f(x) = L$ iff $\forall$open $V \ni L, V \subseteq Y$ $\exists$open $U \ni x_0, U \subseteq X$ s.t. $f(U - \{x_0\}) \subseteq V$.

Definition 6.36. (Continuity of a function on a subset of a topological space).

Let X and Y be topological spaces, let $A \subseteq X$, and consider a function $f : A \subseteq X \to Y$. We say that f *is continuous at* $x_0 \in A$ iff $\forall$open $V \ni f(x_0), V \subseteq Y$ $\exists$open $U \ni x_0, U \subseteq X$ s.t. $f(U) \subseteq V$.

Remark 6.37. (Topological limits and continuity generalize real analytical notions).

Note, the earlier definitions of the limit of a function $A \subseteq \mathbb{R} \to \mathbb{R}$ and continuity at a point for a function $A \subseteq \mathbb{R} \to \mathbb{R}$ can both be viewed as a special case of the just-stated corresponding definitions for topological spaces. Specifically, these earlier definitions are the special case in which $(X, \tau_X) = (Y, \tau_Y) = (\mathbb{R}, \text{std})$, each V is $V = B(\epsilon, L)$, and each U is $U = B(\delta, x_0)$.

Theorem 6.38. (Condition for continuity for functions on subsets of topological spaces).

Let X and Y be topological spaces, and let $A \subseteq X$. A function $f : A \subseteq X \to Y$ is continuous at $x_0 \in A$ iff $\lim_{x \to x_0} f(x) = f(x_0)$.

Proof.

($\implies$). If f is continuous at x_0, then we immediately know $\lim_{x \to x_0} f(x) = f(x_0)$ because $f(U) \subseteq V$ is a weaker condition than $f(U - \{x_0\}) \subseteq V$.

($\impliedby$). Suppose that $\forall$open $V \ni f(x_0), V \subseteq Y$ $\exists$open $U \ni x_0, U \subseteq X$ s.t. $f(U - \{x_0\}) \subseteq V$. We need to show that when $x = x_0$ we have $f(x) \in V$ for any open $V \ni x_0$. But this is true by definition of V; we have $f(x_0) \in V$. □

We now present the classic topological interpretation of a continuous function.

Definition 6.39. (Continuous function).

Let X and Y be topological spaces, and let $A \subseteq X$. A function $f : A \subseteq X \to Y$ is called *continuous* iff it is continuous at every $x \in A$.

Theorem 6.40. (The inverse image of a continuous function preserves openness and closedness).

Let X and Y be topological spaces, and let $A \subseteq X$. A function $f : A \subseteq X \to Y$ is continuous iff the following equivalent conditions hold:

- For all open sets $V \subseteq Y$, the subset $f^{-1}(V) \subseteq X$ is open in X.
- For all closed sets $D \subseteq Y$, the subset $f^{-1}(D) \subseteq X$ is closed in X.

Proof. Left as an exercise. □

Homeomorphisms

Definition 6.41. (Homeomorphism).

Let (X, τ_X) and (Y, τ_Y) be topological spaces. A *homeomorphism* is an injective function $X \to Y$ that is continuous and has a continuous inverse.

Compact and Hausdorff topological spaces

This section briefly details the last two topological notions we require to define manifolds.

Compactness

Definition 6.42. (Compact topological space).

Let (X, τ) be a topological space. A cover $\mathcal{C}$ for X is said to be an *open cover* iff every set $A \in \mathcal{C}$ is open. We say that (X, τ) is *compact* iff for every open cover $\mathcal{O}$ of X, there exists a finite open cover $\mathcal{O}'$ of X with $\mathcal{O}' \subseteq \mathcal{O}$. More succintly, (X, τ) is *compact* iff every open cover of X admits a finite subcover.

Definition 6.43. (Compactness in $\mathbb{R}^n$ with the standard topology).

A set $A \subseteq \mathbb{R}^n$, where $\mathbb{R}^n$ has the standard topology, is compact iff A is closed and bounded[1].

Proof. See any textbook on topology. □

Hausdorff spaces

We need to know what Hausdorff spaces are because manifolds are special types of Hausdorff topological spaces.

Definition 6.44. (Hausdorff space).

Let (X, τ) be a topological space. We say that (X, τ) *is Hausdorff* iff $\forall x \in X \; \forall y \in X \; \exists$open U, V s.t. $U \cap V = \emptyset$. Intuitively, a topological space is Hausdorff iff every two points in that space can be "separated" by taking open sets about each point.

[1] A subset of $\mathbb{R}^n$ is *bounded* iff there exists some open ball that contains it.

Part III

Differential forms

7

Manifolds

This chapter follows various parts of [Lee] and sets the stage for differential forms. Differential forms will live on and be integrated over *smooth manifolds*, which can be thought of as "multidimensional surfaces". We begin the chapter by working simply with *manifolds*, then introduce the notions of *boundary* and *corners* on a manifold

We will discuss the *tangent space* to a point on a manifold, and use the machinery of *tangent vectors* to set up a coordinate-free method of differentiation on smooth manifolds, called the *differential*. In Theorems 7.38 and 7.39, we see a duality between coordiantes of tangent vectors and *tangent covectors* that is reminiscent of the duality between coordinates of vectors and covectors. Lastly, we discuss orientations of manifolds in preparation for the integration of the next chapter.

7.1 Introduction to manifolds

Definition 7.1. [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 7.2. [Lee, p. 3] (Topological invariance of dimension).

A nonempty n-manifold is homeomorphic to an m-dimensional manifold if and only if $n = m$. (See [Lee] for proof).

Definition 7.3. [Lee, p. 25] (Closed n-dimensional upper half-space). Consider $\mathbb{R}^n$ with the standard topology. We define the *closed n-dimensional half space* to be the topology

$$\mathbb{H}^n := \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^n \geq 0\},$$

where $\mathbb{H}^n$ has the subspace topology inherited from the standard topology of $\mathbb{R}^n$.

The point of defining $\mathbb{H}^n$ is to allow for a distinction between what will be called *interior points of M* and *boundary points of M*. To see how involving $\mathbb{H}^n$ facilitates this, note that the interior $\text{int}(\mathbb{H}^n)$ and boundary $\partial\mathbb{H}^n$ of $\mathbb{H}^n$, in the usual topological senses of "interior" and "boundary" (see Section 6.1 of Chapter 6), are

$$\text{int}(\mathbb{H}^n) = \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^n > 0\}$$
$$\partial\mathbb{H}^n = \{(x^1, ..., 0) \in \mathbb{R}^n\}.$$

Recall from Section 6.1 of Chapter 6 that $\mathbf{p} \in \mathbb{H}^n$ is an *interior point of* $\mathbb{H}^n$ iff $\mathbf{p} \in \text{int}(\mathbb{H}^n)$, and $\mathbf{p} \in \mathbb{H}^n$ is a *boundary point of* $\mathbb{H}^n$ iff $\mathbf{p} \in \partial H^n$.

Now let M be 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 $\mathbf{x}(\mathbf{p}) = (x^1(\mathbf{p}), ..., x^{n-1}(\mathbf{p}), 0)$ is in $\partial\mathbb{H}^n$. Then, in analogy to the notion of "locally Eulidean" introduced in the definition of a manifold, we can say that the open subset $U \subseteq M$ "looks like a piece of the boundary of $\mathbb{H}^n$", or that "M locally (near $\mathbf{p}$) looks like a piece of the boundary of $\mathbb{H}^n$". This motivates the following definition.

Definition 7.4. [Lee, p. 25] (Manifold with boundary). An *n-manifold with boundary* is a topological space M that is...

- Hausdorff, or "point-separable".
- *second-countable*; that is, M has a countable basis.
- *locally Euclidean of dimension n* in the sense that each point in M has a neighborhood that is homeomorphic to $\mathbb{R}^n$ (where $\mathbb{R}^n$ has the standard topology).
- such that "M has a (possibly empty) manifold boundary". That is, each point of M has a neighborhood that is either homeomorphic to an open subset of $\mathbb{R}^n$ (where $\mathbb{R}^n$ has the standard topology) or to an open subset $\mathbb{H}^n$ with the subspace topology (inherited from the standard topology on $\mathbb{R}^n$).

Remark 7.5. [Lee, p. 26] (Topological interior and boundary vs. manifold interior and boundary).
Let M be an n-manifold with boundary. We can obtain the *topological* interior and *topological* boundary of M by regarding M as a topological space and taking $\text{int}(M), \partial M$ in the usual topological senses of interior and boundary (see Section 6.1 of Chapter 6). In general, the topological interior and topological boundary are *not* the same as the manifold interior and manifold boundary.

To see this, we first need to remember that the topological notions of interior and boundary are only applicable when M is a subset of some other topological space X. For example, when $X = M$, the topological interior of M in the subspace topology of $X = M$ is M, and the topological boundary of M in the same subspace topology is $\text{cl}(M) - \text{int}(M) = M - M = \emptyset$. (int and cl denote topological interior and closure here). These facts conflict with the manifold interior and boundary of M: the manifold interior cannot be all of M when the manifold boundary of M is nonempty, and the manifold boundary is obviously not necessarily empty.

Definition 7.6. [Lee, p. 25] (Manifold interior and boundary).
The *(manifold) interior of M* is the set of interior points in M, and is denoted $\text{int}(M)$. The *(manifold) boundary of M* is set of all boundary points in M, and is denoted ∂M.

7.2 Coordinatizing manifolds

Definition 7.7. [Lee, p. 4] (Chart).

Let M be an n-manifold. A *(coordinate) chart* on M is a pair $(U, \mathbf{x})$, where $\mathbf{x} : U \to V \subseteq \mathbb{R}^n$ is a function from an open subset $U \subseteq M$, which is called the *domain* of the chart, to an open subset $V \subseteq \mathbb{R}^n$. Since $\mathbf{x}(\mathbf{p}) = \begin{pmatrix} x^1(\mathbf{p}) \\ \vdots \\ x^n(\mathbf{p}) \end{pmatrix}$, we often refer to the component functions $\{x^i\}_{i=1}^n$ as *(local) coordinates*. The component functions are local in the sense that their domain is U, rather than all of M.

A coordinate chart $(U, \mathbf{x})$ is said to be *about* $\mathbf{p} \in M$ iff $\mathbf{p} \in U$.

Definition 7.8. [Lee, p. 13] (Atlas).

Let M be an n-manifold with or without corners. An *atlas* for M is a collection of charts $\{(U_\alpha, \mathbf{x}_\alpha)\}$ whose domains cover M, $M = \cup_\alpha U_\alpha$.

Definition 7.9. [Lee, p. 33] (Coordinate representations of functions on manifolds).

Let M and N be manifolds, and consider a function $\mathbf{F} : M \to N$. Let $(U, \mathbf{x})$ be a chart on M about $\mathbf{p} \in M$, and let $(V, \mathbf{y})$ be a chart on N about $\mathbf{F}(\mathbf{p}) \in N$. The *coordinate representation of* $\mathbf{F} : U \subseteq M \to V \subseteq N$ *relative to the charts* $(U, \mathbf{x})$ *and* $(V, \mathbf{y})$ is the function $\widetilde{\mathbf{F}}_{(U,\mathbf{x}),(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^n \to \mathbb{R}^m$.

In the case that M or N is $\mathbb{R}^k$, we do not require a chart on whichever manifold is $\mathbb{R}^k$, and the coordinate representation of $\mathbf{F}$ is said to be "relative" to only a single chart. In these cases, we the coordinate representation is either $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^k \to N$ or $\widetilde{\mathbf{F}}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{F} : M \to \mathbb{R}^k$.

7.3 Smooth manifolds

Definition 7.10. [Lee, p. 11] (Differentiability classes, smooth functions, and diffeomorphisms).

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be of *differentiability class* C^k iff the partial derivatives 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 $f : M \to \mathbb{R}$ has *differentiability class* C^∞ iff $f \in C^k(M)$ for all $k \in \{0, 1, ...\}$.

A function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^k$ is said to be of *differentiability class* C^k if each component function F_i of $\mathbf{F}$ is of differentiability class C^k, in the previous sense of C^k for functions $\mathbb{R}^n \to \mathbb{R}$. Similarly, a function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ is said to be of *differentiability class* C^∞ if each component function F_i of $\mathbf{F}$ is of differentiability class C^∞, in the previous sense of C^∞ for functions $\mathbb{R}^n \to \mathbb{R}$.

We define $C^k(\mathbb{R}^n \to \mathbb{R}^k)$ to be the set of functions of differentiability class C^k, and define $C^\infty(\mathbb{R}^n \to \mathbb{R}) := \cup_{k=1}^\infty C^k(M)$. We will use the word "smooth" to mean C^∞. Following this convention, the set $C^\infty(\mathbb{R}^n \to \mathbb{R})$ is called the set of *smooth real-valued functions on* $\mathbb{R}^n$.

Lastly, we say that a function $\mathbb{R}^n \to \mathbb{R}$ is a *diffeomorphism* iff it is smooth, bijective, and has a smooth inverse.

Smooth manifolds with or without boundary

Derivation 7.11. [Lee, p. 27, 28] (Smooth manifold with or without boundary).

Let M be an n-manifold with or without boundary. We say that a function $\mathbf{F} : M \to \mathbb{R}^k$ is *smooth relative to a chart* $(U, \mathbf{x})$ of M iff the coordinate representation $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1} : \mathbf{x}(U \cap V) \subseteq \mathbb{R}^n \to \mathbb{R}^k$ is a smooth function.

Consider two charts $(U, \mathbf{x})$ and $(V, \mathbf{y})$. If $U \cap V \neq \emptyset$, then the *transition map* $\mathbf{y} \circ \mathbf{x}^{-1} : \mathbf{x}(U \cap V) \to \mathbf{y}(U \cap V)$ is a homeomorphism, 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 with or without boundary* to be such that, if M is a smooth manifold with or without boundary, then

$$
\begin{gathered}
\mathbf{F} : M \to \mathbb{R}^k \text{ is smooth relative to some chart } (U, \mathbf{x}) \\
\Longrightarrow \\
\mathbf{F} \text{ is smooth relative to all other charts } (V, \mathbf{y}) \text{ that overlap } (U, \mathbf{x}),\ U \cap V \neq \emptyset.
\end{gathered}
$$

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*.

Definition 7.12. (Smooth functions on manifolds).

Let M be an n-manifold, and let $(U, \mathbf{x})$ be a chart on M. A function $\mathbf{F} : U \subseteq M \to \mathbb{R}^n$ is said to be of *differentiability class* C^k *on* U iff the coordinate representation $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^n \to \mathbb{R}$ is of differentiability class C^k, and is said to be of *differentiability class* C^∞ *on* U, or to be *smooth*, iff the coordinate representation $\widetilde{\mathbf{F}}_{(U,\mathbf{x})} = \mathbf{F} \circ \mathbf{x}^{-1} : \mathbb{R}^n \to \mathbb{R}$ is of differentiability class C^∞. The sets of functions on U of differentiability class C^k and C^∞ are denoted $C^k(U \to \mathbb{R}^n)$ and $C^\infty(U \to \mathbb{R}^n)$, respectively.

Example 7.13. [Lee, p. 20] Graph of smooth functions into $\mathbb{R}^k$ are smooth manifolds.

[1]Note, the empty function is a diffeomorphism, so this definition covers the case in which $U \cap V = \emptyset$ and the transition map is the empty function.

Manifolds with or without boundary or corners

Definition 7.14. [Lee, p. 415] (Manifold with or without boundary or corners).

Just as we used $\mathbb{H}^n$ to construct a notion of "manifold boundary", we will use the topological closure of $\mathbb{H}^n$ to construct a notion of "corners on a manifold". Observe that the closure of $\mathbb{H}^n$ is

$$\text{cl}(\mathbb{H}^n) = \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^i \geq 0 \text{ for all } i\}.$$

We say that a topological space M is a *manifold WWBOC or corners* iff

- Hausdorff, or "point-separable".
- *second-countable*; that is, M has a countable basis.
- *locally Euclidean of dimension* n in the sense that each point in M has a neighborhood that is homeomorphic to $\mathbb{R}^n$ (where $\mathbb{R}^n$ has the standard topology).
- such that "M has a (possibly empty) manifold boundary and M has (possibly no) corners". That is, each point of M has a neighborhood that is either homeomorphic to an open subset of $\mathbb{R}^n$ (where $\mathbb{R}^n$ has the standard topology), to an open subset $\mathbb{H}^n$ with the subspace topology (inherited from the standard topology on $\mathbb{R}^n$), or to an open subset of $\text{cl}(\mathbb{H}^n)$ with the subspace topology (inherited from the standard topology on $\mathbb{R}^n$, and where cl denotes topological closure).

Remark 7.15. [Lee, p. 27] (Empty boundary and no corners).

Every smooth manifold is a smooth manifold with boundary, where the manifold boundary is empty. Every smooth manifold with boundary is a smooth manifold with corners, where the set of corner points is empty.

Theorem 7.16. [Lee, p. 26] [Lee, p. 416] (Topological invariance of interior, boundary, and corner points).

If $\mathbf{p} \in M$ is an interior, boundary, or corner point in some chart, then it is an interior, boundary, or corner point, respectively, in all charts. Furthermore, every interior point is neither a boundary point nor a corner point. (See [Lee] for the proofs).

Definition 7.17. [Lee, p. 4] [Lee, p. 415] (Classification of charts and points).

We classify a chart $(U, \mathbf{x})$ on M as follows. $(U, \mathbf{x})$ is a/an...

- *Interior chart* iff $\mathbf{x}(U)$ is an open subset of $\mathbb{R}^n$.
- *Boundary chart* iff $\mathbf{x}(U)$ contains a boundary point of $\mathbb{H}^n$, i.e., iff $\mathbf{x}(U)$ is an open subset of $\mathbb{H}^n$ that intersects the boundary of $\mathbb{H}^n$, $\mathbf{x}(U) \cap \partial\mathbb{H}^n \neq \emptyset$.
- *Chart with corners* iff $\mathbf{x}(U)$ contains a *corner point of cl*$(\mathbb{H}^n)$; a point $\mathbf{p} \in \text{cl}(\mathbb{H}^n)$ is a *corner point of cl*$(\mathbb{H}^n)$ iff more than one of the coordinate functions $x^i|_{\mathbf{p}}$ evaluated at $\mathbf{p}$ vanish.

We classify a point $\mathbf{p} \in M$ according to the type of chart it lies in. $\mathbf{p} \in M$ is a/an...

- *Interior point* iff there is an interior chart about $\mathbf{p}$.
- *Boundary point* iff there is a boundary chart about $\mathbf{p}$.
- *Corner point* iff there is a chart with corners about $\mathbf{p}$.

Definition 7.18. (Manifolds with/without boundary or corners).

We will frequently use the acronym "WWBOC" as shorthand to mean "with/without boundary or corners".

Remark 7.19. [Lee, p. 415] (Manifolds with corners are *topologically* the same as manifolds with boundary).

The title of this theorem is true because $\text{cl}(\mathbb{H}^n)$ is homeomorphic to $\mathbb{H}^n$. *Smooth* manifolds with corners are different than *smooth* manifolds with boundary because the "smoothly compatible" criterion for charts with corners is different from boundary charts.

7.4 Tangent vectors

In multivariable calculus, one can consider tangent vectors that reside in the tangent plane to a surface. We will present a coordinate-free generalization of this concept for manifolds. Surprisingly, we will see that tangent vectors anchored at a point can be identified with "directional derivatives" at that point.

Tangent vectors in $\mathbb{R}^n$

A good place to begin this generalization is to formalize the notion of thinking of vectors as being "anchored" at a point.

Definition 7.20. [Lee, p. 51] (Vectors "anchored" at a point).
Let M be a manifold WWBOC. Given $\mathbf{p} \in M$, we define $\mathbb{R}^n_\mathbf{p}$ to be the vector space $\mathbb{R}^n \times \{\mathbf{p}\}$. We denote a typical element $(\mathbf{v}, \mathbf{p}) \in \mathbb{R}^n_\mathbf{p}$ by $\mathbf{v}_\mathbf{p}$.

Now, we begin to investigate the vector space of "directional derivatives" at a point. Formally, these are called *derivations* at a point.

Definition 7.21. [Lee, p. 52] (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 7.24 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 $T_\mathbf{p}(\mathbb{R}^n)$. An element of $T_\mathbf{p}(\mathbb{R}^n)$ is called a *tangent vector to* $\mathbb{R}^n$ *(at* $\mathbf{p}$*)*.

Theorem 7.22. [Lee, p. 54] (A basis of $T_\mathbf{p}(\mathbb{R}^n)$).
The directional derivatives $\left\{ \frac{\partial}{\partial x^i}\Big|_\mathbf{p} \right\}_{i=1}^n$ form a basis for $T_\mathbf{p}(\mathbb{R}^n)$.

Proof. 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 7.23. [Lee, p. 53] (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 tangent vector to $\mathbb{R}^n$ 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 7.24. (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 tangent vectors to $\mathbb{R}^n$ at $\mathbf{p}$:

$$\Big\{ \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}} \mid \mathbf{v} \in \mathbb{R}^n \Big\} = T_{\mathbf{p}}(\mathbb{R}^n).$$

Proof.

($\subseteq$). Showing that every directional derivative at $\mathbf{p} \in \mathbb{R}^n$ is a derivation at $\mathbf{p} \in \mathbb{R}^n$ follows straightforwardly from the definition of "derivation at $\mathbf{p} \in \mathbb{R}^n$".

($\supseteq$). This direction of the proof is adapted from [Lee, p. 53]. Let $\mathbf{p} \in \mathbb{R}^n$ and let $v_{\mathbf{p}} : C^\infty(\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}$ be a derivation. We must show that $v_{\mathbf{p}}$ is a directinoal derivative at $\mathbf{p}$, and find a vector $\mathbf{v_p} \in \mathbb{R}^n_{\mathbf{p}}$ for which $v_{\mathbf{p}} = \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$. So, we will find a $\mathbf{v_p} \in \mathbb{R}^n_{\mathbf{p}}$ for which

$$v_{\mathbf{p}}(f) = \Big(\frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}\Big)(f) \text{ for all } f \in C^\infty(\mathbb{R}^n \to \mathbb{R}).$$

We use Taylor's theorem (see [Lee, p. 53]) to write $f \in C^\infty(\mathbb{R}^n \to \mathbb{R})$ as

$$f(\mathbf{x}) = f(\mathbf{p}) + \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}(x^i - a^i) + \sum_{i,j=1}^n (x^i - p^i)(x^j - p^j) \int_0^1 (1-t) \frac{\partial^2 f}{\partial x^i \partial x^j}\Big|_{\mathbf{p}+t(\mathbf{x}-\mathbf{p})} dt,$$

where $x^i = [\mathbf{x}]^i_E$, $p^i = [\mathbf{p}]^i_E$.

Now we produce the vector $\mathbf{v_p}$. Let $\{\hat{\mathfrak{e}}_1, ..., \hat{\mathfrak{e}}_n\}$ be the standard basis for $\mathbb{R}^n$ and let $\{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be the induced dual basis for $(\mathbb{R}^n)^*$. We set $\mathbf{v_p} = \sum_{i=1}^n v_{\mathbf{p}}(\phi^{\mathbf{e}_i})\mathbf{e}_i$. (Note that applying the derivation $v_{\mathbf{p}}$ to $\phi^{\mathbf{e}_i}$ makes sense because $\phi^{\mathbf{e}_i}$, being the "ith coordinate function on $\mathbb{R}^n$", is a smooth function $\mathbb{R}^n \to \mathbb{R}$). This choice of $\mathbf{v_p}$ makes more sense in hindsight after reading Theorem 7.39. (That theorem will logically depend on this result, though, so we must make this relatively unmotivated choice of $\mathbf{v_p}$!)

Apply $v_{\mathbf{p}}$ to f and use the second bullet point of the previous lemma to obtain

$$v_{\mathbf{p}}(f) = v_{\mathbf{p}}\Big(\sum_{i,j=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}(x^i - p^i)\Big) = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}(v_{\mathbf{p}}(x^i) - v_{\mathbf{p}}(p^i)) = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} v_{\mathbf{p}}(x^i).$$

When we apply $v_{\mathbf{p}}$ to x^i in the above, we are using a slight abuse of notation and interpreting x^i to be the coordinate function $\mathbf{x} \mapsto x^i$ evaluated at $\mathbf{x}$. Thinking of x^i in this way, we have $x^i = \phi^{\mathbf{e}_i}$, so the above further simplifies to

$$\sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} v_{\mathbf{p}}(\phi^{\mathbf{e}_i}) = (\nabla \mathbf{F})_{\mathbf{p}} \cdot \mathbf{v_p} = \frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}}.$$

Thus, $v_{\mathbf{p}} = \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$. □

Theorem 7.25. (Geometric tangent vectors are naturally isomorphic to derivations).

The map $\mathbf{v_p} \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$ 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 its definition is a basis-independent.

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 5.14). The map $\mathbf{v_p} \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$ 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 adapt the argument from [Lee, p. 53] (which is intended for a slightly different purpose in that book); we show the map $\mathbf{v_p} \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$ has a trivial kernel. So, assume $\mathbf{v_p}$ is sent to the zero function; we need to show $\mathbf{v_p} = \mathbf{0_p}$.

Let $E = \{\mathbf{e}^1, ..., \mathbf{e}^n\}$ be a basis for $\mathbb{R}^n$, and let $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ be its dual basis for $(\mathbb{R}^n)^*$. Note that since $\phi^{\mathbf{e}_i} : \mathbb{R}^n \to \mathbb{R}$ is defined by $\phi^{\mathbf{e}_i}(\mathbf{p}) = [\mathbf{p}]^i_E$, each $\phi^{\mathbf{e}_i}$ is a C^∞ function on $\mathbb{R}^n$.

At $\mathbf{p} \in \mathbb{R}^n$, the ith coordinate of the zero map is $\frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}(\phi^{\mathbf{e}_i})$ (see Theorem 3.26). All coordinates of the zero function relative to any basis must be zero, so, using the linearity of the directional derivative with respect to $\mathbf{v_p}$ (again, see Remark 5.14), we have

$$0 = \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}(\phi^{\mathbf{e}_i}) = \frac{\partial}{\partial(\sum_j [\mathbf{v}]^j_E \mathbf{e}^j)}\Big|_{\mathbf{p}}(\phi^{\mathbf{e}_i}) = \sum_j [\mathbf{v}]^j_E \Big(\frac{\partial}{\partial \mathbf{e}^j}\Big|_{\mathbf{p}}\Big)(\phi^{\mathbf{e}_i}).$$

Since $\frac{\partial}{\partial \mathbf{e}^j}\Big|_{\mathbf{p}} = \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$, the above becomes

$$0 = \sum_j [\mathbf{v}]^j_E \delta^i_j = [\mathbf{v}]^i_E.$$

We see $[\mathbf{v}]^i_E = 0$ for all i, so $\mathbf{v} = \mathbf{0}$, which means $\mathbf{v_p} = \mathbf{0_p}$. □

We have established all the facts about derivations we will need in the context of $\mathbb{R}^n$. We now extend these definitions to a manifold.

Tangent vectors on manifolds

Definition 7.26. [Lee, p. 54] (Derivation at $\mathbf{p} \in M$, tangent space to a manifold).

Let M be a smooth n-manifold WWBOC. A *derivation at* $\mathbf{p} \in M$ is a linear function $v_{\mathbf{p}} : C^\infty(M \to \mathbb{R}) \to \mathbb{R}$ that satisfies the product rule,

$$v_{\mathbf{p}}(fg) = v_{\mathbf{p}}(f)g(\mathbf{p}) + f(\mathbf{p})v_{\mathbf{p}}(g) \text{ for all } f, g \in C^\infty(M \to \mathbb{R}).$$

The set of derivations at $\mathbf{p} \in M$ is called the *tangent space to M at* $\mathbf{p}$, and is denoted $T_{\mathbf{p}}(M)$. An element of $T_{\mathbf{p}}(M)$ is called a *tangent vector to M (at* $\mathbf{p}$*)*.

Theorem 7.27. [Lee, p. 54] (Properties of derivations at $\mathbf{p} \in M$).

Let M be a smooth n-manifold WWBOC, let $\mathbf{p} \in \mathbb{R}^n$, $v_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^n)$ be a tangent vector at $\mathbf{p}$, and $f, g \in C^\infty(\mathbb{R}^n \to \mathbb{R})$. Then

- If f is a constant function, then $v_{\mathbf{p}}(f) = 0$.
- If $f(\mathbf{p}) = g(\mathbf{p}) = 0$, then $v_{\mathbf{p}}(fg) = 0$.

Proof. The proof is analogous to the proof of Lemma 7.23. □

Definition 7.28. [War, p. 14, 15] (Basis of $T_{\mathbf{p}}(M)$ and its abuse of notation).

Let M be a smooth n-manifold WWBOC, let $(U, \mathbf{x})$ be a smooth chart on M, and let $\mathbf{p} \in M$. In an abuse of notation, we define $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$, $i \in \{1, ..., n\}$, to be the derivations at $\mathbf{p} \in U \subseteq M$ for which

$$\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}(f) = \frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}(f \circ \mathbf{x}^{-1}),$$

where the $\frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial}{\partial \hat{\mathbf{i}}^i}\Big|_{\mathbf{x}(\mathbf{p})}$ 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, $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}_{i=1}^n$ is a basis of $T_{\mathbf{p}}(M)$. This follows because the $\frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$ on the right hand side are a basis for $T_{\mathbf{x}(\mathbf{p})}(\mathbb{R}^n)$ (see Theorem 7.22).

Remark 7.29. 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 condition of the above definition becomes

$$\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}(f) = \frac{\partial}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \widetilde{\mathbf{F}}_{(U,\mathbf{x})}.$$

Proof. This proof pertains to the above definition, not to the above remark. We need to show that the $\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}} : C^\infty(M \to \mathbb{R}) \to \mathbb{R}$ from the left side of the condition of the definition are derivations at $\mathbf{p} \in M$. In this proof, we put an "LHS" under $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ when we mean $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ to be from the left hand side of the condition in the above definition. All other occurrences of $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ in this proof are directional derivatives.

We need to show that $\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}$ are linear and follow the product rule. Linearity follows easily from the linearity of the directional derivative with respect to its argument from $C^\infty(\mathbb{R}^n \to \mathbb{R})$. We show the product rule holds:

$$\begin{aligned}
\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}(fg) &= \frac{\partial(fg \circ \mathbf{x}^{-1})}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\Big|_{\mathbf{x}^{-1}(\mathbf{x}(\mathbf{p}))} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} = \frac{\partial(fg)}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \\
&= \Big(\frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial g}{\partial x^i}\Big|_{\mathbf{p}}\Big) \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \\
&= \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial g}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial \mathbf{x}^{-1}}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})} \\
&= \frac{\partial(f \circ \mathbf{x})}{\partial x^i}\Big|_{\mathbf{p}} g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\frac{\partial(g \circ \mathbf{x})}{\partial x^i}\Big|_{\mathbf{p}} \\
&= \Big(\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}\Big)(f)\ g(\mathbf{x}(\mathbf{p})) + f(\mathbf{x}(\mathbf{p}))\ \Big(\underbrace{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}}_{\text{LHS}}\Big)(g).
\end{aligned}$$

Here, we used the chain and product rules and then reversed the product and chain rules. $\square$

Differentials of a smooth function on a manifold

Now that we have defined tangent vectors to a manifold, we can define a generalization of the derivative of a vector-valued function accepting a vector as input: the *differential*. Again, we start in $\mathbb{R}^n$, but then obtain a general framework that presents differentials as functions that send tangent vectors on one smooth manifold to tangent vectors on another smooth manifold.

Definition 7.30. (Differential of a smooth function $\mathbb{R}^n \to \mathbb{R}^m$).

Consider a smooth function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$. We define the *differential* $d\mathbf{F}_\mathbf{p} : T_\mathbf{p}(\mathbb{R}^n) \to T_{\mathbf{F}(\mathbf{p})}(\mathbb{R}^m)$ *of* $\mathbf{F}$ *at* $\mathbf{p}$ to be the 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

$$d\mathbf{F}_\mathbf{p}\Big(\frac{\partial}{\partial x^i}\Big) = \sum_{j=1}^m a_i^j \frac{\partial}{\partial y^j} = \sum_{j=1}^m \frac{\partial F^j}{\partial x^i}\Big|_\mathbf{p} \frac{\partial}{\partial y^j}\Big|_{\mathbf{F}(\mathbf{p})},$$

where $(a_j^i) = \Big[d\mathbf{F}_\mathbf{p}\Big(\Big\{\frac{\partial}{\partial x^i}\Big\}\Big)\Big]_{\{\frac{\partial}{\partial y^j}\}}$ 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 7.31. (Differential of a smooth function $\mathbf{F} : M \to N$).

Let M and N be smooth n- and m- dimensional manifolds WWBOC.

We define the *differential* $d\mathbf{F}_\mathbf{p} : T_\mathbf{p}(M) \to T_{\mathbf{F}(\mathbf{p})}(N)$ *of* $\mathbf{F}$ *at* $\mathbf{p}$ by

$$d\mathbf{F}_\mathbf{p}(v_\mathbf{p})(f) = v_\mathbf{p}(f \circ \mathbf{F}),$$

where $v_\mathbf{p} \in T_\mathbf{p}(M)$ and $f \in C^\infty(N)$.

Proof. We need to check that $d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})$ is a derivation at $\mathbf{F}(\mathbf{p}) \in N$. To do so, follow the proof of Theorem 7.28, 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 7.32. [Lee, p. 55] (Properties of differentials). Let M, N, and P be smooth manifolds WWBOC, let $\mathbf{F} : M \to N$ and $\mathbf{G} : N \to P$ be smooth functions, and let $\mathbf{p} \in M$.

- (Chain rule). $d(\mathbf{G} \circ \mathbf{F})_{\mathbf{p}} = d\mathbf{G}_{\mathbf{F}(\mathbf{p})} \circ d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{(\mathbf{G}\circ\mathbf{F})(\mathbf{p})}(P)$.
- (Differential of the identity). $d(\mathbf{I}_M)_{\mathbf{p}} = \mathbf{I}_M$, where $\mathbf{I}_M : T_{\mathbf{p}}(M) \to T_{\mathbf{p}}(M)$ is the identity on M.
- If $\mathbf{F}$ is a diffeomorphism, then $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{\mathbf{F}(\mathbf{p})}(N)$ is a linear isomorphism, and $(d\mathbf{F}_{\mathbf{p}})^{-1} = d(\mathbf{F}^{-1})_{\mathbf{F}(\mathbf{p})}$.

Proof. See [Lee] for the proof. □

Theorem 7.33. [Lee, p. 281] (Differential of a smooth function $M \to \mathbb{R}$).

What happens when we take the differential of a smooth function $f : M \to \mathbb{R}$? Well, by definition of the differential of a smooth function $M \to N$, we have

$$df_{\mathbf{p}}(v_{\mathbf{p}})(g) = v_{\mathbf{p}}(g \circ f).$$

We have $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \to T_{f(\mathbf{p})}(\mathbb{R})$. Notice that since $T_{f(\mathbf{p})}(\mathbb{R})$ is 1-dimensional, then $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$. There is therefore an induced function $\widetilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ for which the following diagram commutes:

$$\begin{tikzcd}
T_{\mathbf{p}}(M) \arrow[r, "v_{\mathbf{p}} \mapsto df_{\mathbf{p}}(v_{\mathbf{p}})"] \arrow[dr, "\widetilde{d}f_{\mathbf{p}}"'] & T_{f(\mathbf{p})}(\mathbb{R}) \arrow[d, "w \mapsto \sum_{i=1}^n w(x^i)\hat{\mathfrak{e}}_i"] \\
& \mathbb{R}
\end{tikzcd}$$

We think of $\widetilde{d}$ as the differential that is induced by the identification $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$.

From the diagram, we see that the map $\widetilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ acts on a tangent vector $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by

$$\widetilde{d}f_{\mathbf{p}}(v_{\mathbf{p}}) = \sum_{i=1}^{1} df_{\mathbf{p}}(v_{\mathbf{p}})(x^i)\hat{\mathfrak{e}}_i = df_{\mathbf{p}}(v_{\mathbf{p}})(x^1)\hat{\mathfrak{e}}_1 = v_{\mathbf{p}}(x^1 \circ f)\hat{\mathfrak{e}}_1.$$

Since $\mathbb{R}^1$ is one-dimensional, then $\mathbb{R}^1 \cong \mathbb{R}$. As we switch from $\mathbb{R}^1$ to $\mathbb{R}$, the coordinate function $x^1 : \mathbb{R}^1 \to \mathbb{R}^1$ becomes the identity on $\mathbb{R}$ and $\hat{\mathfrak{e}}_1$ becomes the scalar $1 \in \mathbb{R}$. These final identifications[2] give

$$\widetilde{d}f_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f).$$

In practice, we write $df_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$ to mean $\widetilde{d}f_{\mathbf{p}} : T_{\mathbf{p}}(M) \to \mathbb{R}$. (This includes "$dx^i$"; whenever x^i is a coordinate function, we write dx^i to mean $\widetilde{d}x^i$, where $\widetilde{d}$ is the differential obtained by identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$). So, the above characterizing condition of the differential of $f : M \to \mathbb{R}$ is restated as

$$\boxed{df_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)}$$

[2] To be very formal, we could write the above as $\widetilde{\widetilde{d}}f_{\mathbf{p}}(v_{\mathbf{p}})$ to indicate that $\widetilde{\widetilde{d}}$ is the differential obtained from $\widetilde{d}$ by identifying $\mathbb{R}^1 \cong \mathbb{R}$, but this identification is so trivial that we do not do this.

Theorem 7.34. [Lee, p. 281] (Differential of a smooth function $M \to \mathbb{R}$ in coordinates).

Let M be a smooth n-manifold WWBOC or corners, and let $(U, \mathbf{x})$ be a smooth chart on M, with x^i being the ith coordinate function of $\mathbf{x}$. Then the differential of a smooth function $f : M \to \mathbb{R}$ obtained by identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ (see the previous theorem) is given by

$$\boxed{df_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}}$$

Note, the d on the right hand side is the differential of the ith coordinate function x^i of $\mathbf{x}$ after identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ (see the previous theorem).

Proof. Notice that since $\mathbf{v} \mapsto df_{\mathbf{p}}(\mathbf{v})$ is a linear function, $df_{\mathbf{p}}$ is an element of $T_{\mathbf{p}}(M)^*$. (We will explore this fact more later, in Section 7.5).

Since $df_{\mathbf{p}} \in T_{\mathbf{p}}(M)^*$, we can decompose $df_{\mathbf{p}}$ by using the dual basis $\{\lambda^i|_{\mathbf{p}}\}_{i=1}^n$ for $T_{\mathbf{p}}(M)^*$ induced by the basis $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}_{i=1}^n$ for $T_{\mathbf{p}}(M)$. (The notation $\{\lambda^i|_{\mathbf{p}}\}_{i=1}^n$ is temporary; by the end of this proof, we will have a more meaningful notation for this dual basis). Thus, we can write

$$df_{\mathbf{p}} = \sum_{i=1}^{n} C_i(\mathbf{p}) \lambda^i|_{\mathbf{p}}$$

for some smooth functions $C_i : U \to \mathbb{R}$.

To determine the C_i, recall that Theorem 3.26 stated that if V is a finite-dimensional vector space, $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is a basis for V, and $E^* = \{\phi^{\mathbf{e}_1}, ..., \phi^{\mathbf{e}_n}\}$ is the basis for V^* induced by E, then for any $\phi \in V^*$, we have $([\phi]_{E^*})_i = \phi(\mathbf{e}_i)$. Applying this theorem and using $df_{\mathbf{p}}(v_{\mathbf{p}}) = v_{\mathbf{p}}(f)$, we have

$$C_i(\mathbf{p}) = df_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big) = \Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)(f) = \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}}.$$

Thus $df_{\mathbf{p}}$ is

$$df_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \lambda^i_{\mathbf{p}}.$$

It remains to determine the $\lambda^i|_{\mathbf{p}}$. We claim that $\lambda^i|_{\mathbf{p}} = dx^i|_{\mathbf{p}}$, where x^i is the ith coordinate function of $\mathbf{x}$, and where d of a smooth function $M \to \mathbb{R}$ obtained by identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ (so d is same differential that we've been using in this proof; it is the differential $\widetilde{d}$ of the previous theorem). To prove the claim, we compute

$$dx^i|_{\mathbf{p}} = \sum_{j=1}^{n} \frac{\partial x^i}{\partial x^j}\Big|_{\mathbf{p}} \lambda|^j_{\mathbf{p}}.$$

In the sum, $\frac{\partial x^i}{\partial x^j}$ is the partial derivative of the ith coordinate function $x^i : U \to \mathbb{R}$ relative to x^j. This partial derivative is $\frac{\partial x^i}{\partial x^j} = \delta^i_j$, so we obtain

$$dx^i|_{\mathbf{p}} = \sum_{j=1}^{n} \delta^i_j \lambda|^j_{\mathbf{p}} = \lambda|^i_{\mathbf{p}}.$$

□

Remark 7.35. (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

$$df_{\mathbf{p}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}.$$

Since $M = \mathbb{R}^n$, then $df_{\mathbf{p}} : T_{\mathbf{p}}(\mathbb{R}^n) \to \mathbb{R}$. Notice that we have $T_{f(\mathbf{p})}(\mathbb{R}^n) \cong \mathbb{R}^n$ because $T_{\mathbf{p}}(\mathbb{R}^n)$ and $\mathbb{R}^n_{\mathbf{p}}$ are both n-dimensional, so we get from $df_{\mathbf{p}} : T_{\mathbf{p}}(\mathbb{R}^n) \to \mathbb{R}$ the induced function $\widetilde{d}f_{\mathbf{p}} : \mathbb{R}^n_{\mathbf{p}} \to \mathbb{R}$, where $\widetilde{d}$ is thought of as the "induced differential". Above, we see that our original $df_{\mathbf{p}}$ is a linear combination of $dx^i|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. Thus, $\widetilde{d}f_{\mathbf{p}}$ is a linear combination of $\widetilde{d}x^i|_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^n)$ with the same weights:

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \widetilde{d}x^i$$

In the proof of the previous theorem, we saw that $dx^i|_{\mathbf{p}}$ is the ith coordinate function on $T_{\mathbf{p}}(M)$. Since $T_{\mathbf{p}}(\mathbb{R}^n) \cong \mathbb{R}^n_{\mathbf{p}}$, then $\widetilde{d}x^i$ must be the ith coordinate function on $\mathbb{R}^n_{\mathbf{p}}$. This implies that $\widetilde{d}f_{\mathbf{p}}$ acts on $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$ by

$$\widetilde{d}f_{\mathbf{p}}(\mathbf{v}_{\mathbf{p}}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} \widetilde{d}x^i(\mathbf{v}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} [\mathbf{v}_{\mathbf{p}}]^i_{\hat{\mathfrak{E}}} = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}} \cdot \mathbf{v} = \frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}},$$

where $\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$ is the directional derivative of f at $\mathbf{p}$ in the direction of $\mathbf{v}$. This is to be expected because we defined the differential of a smooth function of smooth manifolds (see Definition 7.30) so that its coordinate representation is represented by the Jacobian relative to the coordinate bases. (Here, the row-matrix of partial derivatives is the Jacobian matrix of f. Recall from Definition 5.12 that the Jacobian is used to express the directional derivative of a smooth function $\mathbb{R}^n \to \mathbb{R}^m$. In this case we have $m = 1$).

Remark 7.36. (Infinitesimal velocity versus infinitesimal time).

Recall from Remark 5.7 that there are two notations for the directional derivative of a function, where one favors the "infinitesimal time" interpretation and the other favors the "infinitesimal velocity" interpretation. Since the differential of a function $f : U \subseteq M \to \mathbb{R}$ is denoted df, we see that differential geometry favors the "infinitesimal velocity" interpretation.

7.5 Tangent vectors and tangent covectors with coordinates

Definition 7.37. [Lee, p. 275] (Cotangent space to a manifold).

Let M be an smooth n-manifold WWBOC, and let $\mathbf{p} \in M$. The *cotangent space* $T^*_{\mathbf{p}}(M)$ *to* M *at* $\mathbf{p}$ is the dual vector space to the tangent space at $\mathbf{p}$. That is, $T^*_{\mathbf{p}}(M) := T_{\mathbf{p}}(M)^*$.

An element $\phi_{\mathbf{p}} \in T^*_{\mathbf{p}}(M)$ of the cotangent space at $\mathbf{p}$ is called a *tangent covector*, or a *covector* for short.

Theorem 7.38. (Induced bases in a chart).

Let M be a smooth n-manifold WWBOC and let $(U, \mathbf{x})$ be a smooth chart on M about $\mathbf{p} \in M$. Consider the set $C^\infty(U \subseteq M \to \mathbb{R})$ of smooth real-valued functions defined on $U \subseteq M$ as a vector space over $\mathbb{R}$. It is a helpful *mnemonic* to pretend that $\{x^i|_{\mathbf{p}}\}_{i=1}^n$ is a basis for $C^\infty(U \subseteq M \to \mathbb{R})$, where x^i is the ith coordinate function of $\mathbf{x}$, and where we've denoted $x^i|_{\mathbf{p}} := x^i(\mathbf{p})$. (*When interpreted even with a little common sense, the previous statement is is clearly nonsensical*, because $C^\infty(U \subseteq M \to \mathbb{R})$ is an infinite-dimensional vector space). This is because, if we accept this *mnemonic*, then

1. The dual basis for $T_{\mathbf{p}}(M) = C^\infty(U \subseteq M \to \mathbb{R})^*$ induced by $\{x^i\}_{i=1}^n$ is $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}$.
2. The dual basis for $T^*_{\mathbf{p}}(M) = T_{\mathbf{p}}(M)^* = C^\infty(U \subseteq M \to \mathbb{R})^{**}$ induced by $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}$ is $dx^i|_{\mathbf{p}}$, where x^i is the ith coordinate function of $\mathbf{x}$, and where the d here is *not* obtained by identifying $T_{\mathbf{p}}(\mathbb{R}^n) \cong \mathbb{R}$, but by "leaving $T_{\mathbf{p}}(\mathbb{R}^n)$ alone" (recall Theorem 7.33 to see happens to d when we identify $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$).

Proof.

1. We have $\left(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right)(x^j) = \delta^j_i$.
2. This was proved as part of showing $df_{\mathbf{p}} = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}$ in Theorem 7.33.

□

If we continue to accept the mnemnoic of the previous theorem, we obtain the following theorem, which describes how to compute the coordinates of tangent vectors and cotangent vectors as a simple consequence of linear algebra.

Theorem 7.39. (Coordinates of tangent vectors and cotangent vectors).

Theorem 3.26 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

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

where $\Phi_{\mathbf{v}} \in V^{**}$ is the linear function $V^* \to K$ defined by $\Phi_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$.

We can apply this theorem to the pairs of bases and induced dual bases from the last theorem. Let M be a smooth n-manifold WWBOC and let $(U, \mathbf{x})$ be a smooth chart on M about $\mathbf{p} \in M$. Then the ith coordinate of a tangent vector $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ relative to $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}_{i=1}^n$ and the ith coordinate of a tangent covector $\phi_{\mathbf{p}} \in T^*_{\mathbf{p}}(M)$ are

$$\boxed{\begin{aligned}&([v_{\mathbf{p}}]_{\left\{\frac{\partial}{\partial x^j}\big|_{\mathbf{p}}\right\}_{j=1}^n})^i = v_{\mathbf{p}}(x^i) = \phi_{x^i}(v_{\mathbf{p}})\\ &([\phi_{\mathbf{p}}]_{\left\{dx^j|_{\mathbf{p}}\right\}_{j=1}^n})_i = \phi_{\mathbf{p}}\Big(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\Big)\end{aligned}}$$

In the second equation of the first line, ϕ_{x^i} is the element of $C^\infty(U \subseteq M \to \mathbb{R})^{**} = T^*_{\mathbf{p}}(M)$ that is identified with the ith coordinate function $x^i \in C^\infty(U \subseteq M \to \mathbb{R})$ of $\mathbf{x}$. Recall Theorem 2.23 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 7.40. (Interpretations of x^i).

Let M be a smooth n-manifold, and consider a smooth chart $(U, \mathbf{x})$ about $\mathbf{p} \in M$.

When considering the basis $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}$ for $T_{\mathbf{p}}(M)$, the x^i in the "denominator" is *not* a coordinate function of the smooth chart $\mathbf{x}$ that is involved in the definition $\left(\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right)(f) = \frac{\partial(f \circ \mathbf{x}^{-1})}{\partial x^i}\Big|_{\mathbf{x}(\mathbf{p})}$. (Recall Definition 7.28). 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 $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} := \frac{\partial}{\partial \hat{\mathfrak{e}}^i}\Big|_{\mathbf{p}}$ 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 7.41. (Change of coordinates for tangent vectors in terms of basis vectors of $T_{\mathbf{p}}(M)$).

Theorem 1.58 stated that if V is a finite-dimensional vector space with bases E and F, then

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

Let M be a smooth n-manifold, and consider smooth charts $(U, \mathbf{x})$ and $(V, \mathbf{y})$, where $\mathbf{p} \in U \cap V$. Let x^i and y^i denote the ith coordinate functions of $\mathbf{x}$ and $\mathbf{y}$, respectively. Applying the above theorem to the vector space $T_{\mathbf{p}}(M)$ and its bases $E = \left\{\frac{\partial}{\partial x^j}\Big|_{\mathbf{p}}\right\}_{j=1}^n$ and $F = \left\{\frac{\partial}{\partial y^j}\Big|_{\mathbf{p}}\right\}_{j=1}^n$, we have

$$[\mathbf{F}]_E = \left(\left[\frac{\partial}{\partial y^1}\Big|_{\mathbf{p}}\right]_F \quad \cdots \quad \left[\frac{\partial}{\partial y^n}\Big|_{\mathbf{p}}\right]_E\right) = \begin{pmatrix} \frac{\partial x^1}{\partial y^1}\Big|_{\mathbf{p}} & \cdots & \frac{\partial x^1}{\partial y^n}\Big|_{\mathbf{p}} \\ \vdots & & \vdots \\ \frac{\partial x^n}{\partial y^1}\Big|_{\mathbf{p}} & \cdots & \frac{\partial x^n}{\partial y^n}\Big|_{\mathbf{p}} \end{pmatrix} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}},$$

where $\frac{\partial x^i}{\partial y^j}\Big|_{\mathbf{p}} = \left(\frac{\partial}{\partial y^j}\Big|_{\mathbf{p}}\right)(x^i)$, and where x^i is the ith coordinate function of $\mathbf{x}$. The matrix $\frac{\partial \mathbf{x}}{\partial \mathbf{y}}$ is the Jacobian matrix described in Definition 5.12.

Applying the fact $\mathbf{f}_i = \sum_{j=1}^n ([\mathbf{F}]_E)_{ji} \mathbf{e}_j$ from above, we have

$$\boxed{\frac{\partial}{\partial y^i}\Big|_{\mathbf{p}} = \sum_{j=1}^n \frac{\partial x^j}{\partial y^i}\Big|_{\mathbf{p}} \frac{\partial}{\partial x^j}\Big|_{\mathbf{p}}.}$$

This change of basis equation strongly resembles the chain rule, and indeed simplifies to the chain rule when $M = \mathbb{R}^n$. 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.

7.6 Oriented manifolds and their oriented boundaries

In Section 4.3, we defined what it means to orient a finite-dimensional vector space. The key result we motivated was Definition 4.57, which said: if a basis E of a finite-dimensional vector space V is fixed, than another ordered basis F of V is said to be *positively oriented (relative to E)* iff $\det([\mathbf{F}]_E) > 0$ and *negatively oriented (relative to E)* otherwise. Given this definition, we showed in Theorem 4.59 that a choice of element of $\Lambda^{\dim(V)}(V)$, determines an orientation on V. We now extend the idea of orientation to manifolds.

Definition 7.42. [Lee, p. 380] (Oriented manifolds).
Let M be an n-manifold WWBOC. We define the notion of orientation on a manifold in the following steps:

- Since a choice of element of the top exterior power of a finite-dimensional vector space determines an orientation for that vector space, we define a *pointwise orientation form* for M to be any nonvanishing differential n-form on M that is an element of $\widetilde{\Omega}^n(M)$. Such a differential form is an element of $\widetilde{\Lambda}^k(T_{\mathbf{p}}(M)) \cong \Lambda^k(T_{\mathbf{p}}(M))$ at each $\mathbf{p} \in M$, and therefore determines an orientation on each tangent space.

- If ω is a pointwise orientation form on M and $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ is a local frame for the tangent bundle $T(M)$, then we say that $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ is *positively oriented* iff $\{\mathbf{E}_1|_{\mathbf{p}}, ..., \mathbf{E}_n|_{\mathbf{p}}\}$ is positively oriented relative to ω at each $\mathbf{p} \in M$, that $\{\mathbf{E}_1, ..., \mathbf{E}_n\}$ is *negatively oriented* iff $\{\mathbf{E}_1|_{\mathbf{p}}, ..., \mathbf{E}_n|_{\mathbf{p}}\}$ is negatively oriented relative to ω at each $\mathbf{p} \in M$.

- A pointwise orientation form ω for M is said to be *continuous* iff every point of M is in the domain of an oriented local frame, where the orientation of the oriented local frame is given by ω.

- An *orientation form* for M is a continuous pointwise orientation form for M. We say M is *orientable*, or that M is an *oriented manifold*, if there exists an orientation on M.

Definition 7.43. [Lee, p. 381, 382] (Orientation of a smooth chart on an oriented manifold).
A smooth chart $(U, \mathbf{x})$ on an oriented smooth n-manifold WWBOC is said to be *positively oriented* iff the coordiante frame $\{\frac{\partial}{\partial x^i}\}_{i=1}^n$ is positively oriented, and is said to be *negatively oriented* otherwise.

Boundary orientation

We now present the natural way in which the boundary of a manifold inherits orientation from the rest of the manifold.

Lemma 7.44. If M is an n-manifold, then the boundary ∂M is an $(n-1)$-manifold.

To describe the inherited orientation on a boundary, we need to define what it means for tangent vectors to be "inward-pointing" or "outward-pointing".

Definition 7.45. [Lee, p. 118] (Inward- and outward- pointing tangent vectors).
Let $(U, \mathbf{x})$ be a smooth chart on ∂M with $\mathbf{p} \in U$. We classify tangent vectors in $T_{\mathbf{p}}(M)$ as follows.

- A tangent vector is *inward-pointing (on ∂M)* iff it has positive-x^n component.

- A tangent vector is *tangent to ∂M* iff it has an x^n-component of zero.

- A tangent vector is *outward-pointing (on ∂M) iff it has negative-x^n component*, i.e., $\mathbf{v} \in T_{\mathbf{p}}(M)$ is outward pointing iff $-\mathbf{v}$ is inward pointing.

Lemma 7.46. [Lee, p. 200, problem 8-4] (Existence of inward- and outward-pointing vector fields on M).
There exists a global smooth vector field on M whose restriction to ∂M is everywhere inward-pointing, and one whose restriction to ∂M is everywhere outward-pointing.

Now we can describe the induced orientation on the boundary.

Theorem 7.47. [Lee, p. 385, 386] (Induced orientation form on the boundary).

Let ω_M be an orientation form on M. The previous lemma shows that there exists a vector field $\mathbf{N}$ that is nowhere tangent to ∂M, then there is an induced orientation form $\omega_{\partial M}$ on the boundary (due to interior multiplication)[3] defined by $\omega_{\partial M}(v_1|_{\mathbf{p}}, ..., v_{n-1}|_{\mathbf{p}}) = \omega_M(\mathbf{N}_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., v_{n-1}|_{\mathbf{p}})$. The orientation on ∂M induced by $\omega_{\partial M}$ does not depend on the vector field $\mathbf{N}$ that is nowhere tangent to ∂M.

Proof. We need to show (1) that $\omega_{\partial M}$ is indeed an orientation form on ∂M and (2) that the orientation induced by $\omega_{\partial M}$ is independent of the choice of the nowhere tangent vector field $\mathbf{N}$.

1. We need to show that $\omega_{\partial M}$ never vanishes. Note that if $\{e_1|_{\mathbf{p}}, ..., e_{n-1}|_{\mathbf{p}}\}$ is a basis for $T_{\mathbf{p}}(\partial M)$, then, since $\mathbf{N}$ is nowhere tangent to $T_{\mathbf{p}}(\partial M)$, the set $\{\mathbf{N}_{\mathbf{p}}, e_1|_{\mathbf{p}}, ..., e_{n-1}|_{\mathbf{p}}\}$ is a basis for $T_{\mathbf{p}}(M)$. Because of this, and as ω_M is nonvanishing on M, then $\omega_{\partial M}$ is also nonvanishing.

   Since ω_M is nonvanishing, $\omega_{\partial M}$ is also nonvanishing, which makes it an orientation form on ∂M. See [Lee, p. 385] for the more precise details.

2. Let $\mathbf{N}$ and $\mathbf{N}'$ be two vector fields that are both nowhere tangent to ∂M. We need to show that the ordered bases $E = \{\mathbf{N}_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., v_{n-1}|_{\mathbf{p}}\}$ and $F = \{\mathbf{N}'(\mathbf{p}), v_1|_{\mathbf{p}}, ..., v_{n-1}|_{\mathbf{p}}\}$ have the same orientation. To do so, we prove that the determinant of the change of basis matrix between the two ordered bases is positive.

   $\mathbf{N}$ and $\mathbf{N}'$ are both outward-pointing, so the nth component of $\mathbf{N}_{\mathbf{p}}$ relative to E and the nth component of $\mathbf{N}'(\mathbf{p})$ relative to F are both negative; denote these nth components by $(\mathbf{N}_{\mathbf{p}})_n$ and $(\mathbf{N}'(\mathbf{p}))_n$, respectively. Relative to the bases E, F, the change of basis matrix between E and F has a first column whose 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 $\frac{(\mathbf{N}'(\mathbf{p}))_n}{(\mathbf{N}_{\mathbf{p}})_n} > 0$.

□

Orientation of the boundary of a k-parallelapiped

As k-paralellapiped can be be given the structure of an oriented smooth submanifold of $\mathbb{R}^n$ with corners.

Definition 7.48. (Notation for n-parallelapipeds).

Let M be a smooth manifold WWBOC, and let $(U, \mathbf{x})$ be a smooth chart on M. Given $v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$, we define $P_{\mathbf{p}}(v_1, ..., v_n)$ to be the n-parallelapiped anchored at $\mathbf{x}(\mathbf{p}) \in \mathbb{R}^n_{\mathbf{p}}$ that is spanned by $\mathbf{x}(v_1|_{\mathbf{p}}), ..., \mathbf{x}(v_n|_{\mathbf{p}}) \in \mathbb{R}^n_{\mathbf{p}}$. Additionally, if $v \in T_{\mathbf{p}}(M)$ is identified with $\mathbf{v} \in \mathbb{R}^n_{\mathbf{p}}$ under the isomorphism $\frac{\partial}{\partial x^i}\mathbf{p} \mapsto \hat{\mathfrak{e}}_i$, we will use the slight abuse of notation $P_{\mathbf{p}+v}(v_1, ..., v_n)$ to denote the n-parallelapiped anchored at $\mathbf{x}(\mathbf{p}) + \mathbf{v} \in \mathbb{R}^n_{\mathbf{p}}$ that is spanned by $\mathbf{x}(v_1|_{\mathbf{p}}), ..., \mathbf{x}(v_n|_{\mathbf{p}}) \in \mathbb{R}^n_{\mathbf{p}}$.

Theorem 7.49. [HH, p. 542 - 544] (Oriented boundary of an n-parallelapiped).

Let M be an n-dimensional manifold, let $\mathbf{p} \in M$, and consider vectors $v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. Let $P_{\mathbf{p}}(v_1, ..., v_n)$ denote the n-parallelapiped anchored at $\mathbf{p}$ that is spanned by $v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}}$. The orientation of the boundary $\partial P_{\mathbf{p}}(v_1, ..., v_n)$ is given by

$$\boxed{\partial P_{\mathbf{p}}(v_1, ..., v_n) = \sum_{i=1}^{k} (-1)^{i-1} \Big(P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n) - P_{\mathbf{p}}(v_1, ..., \cancel{v_i}, ..., v_n) \Big)}$$

In the sum, $+$ and $-$ signs are used to indicate whether a n-parallelapiped is positively or negatively oriented relative to the orientation of $P_{\mathbf{p}}(v_1, ..., v_n)$.

[3] We could have put $\mathbf{N}_{\mathbf{p}}$ in any of ω_M's n argument slots, but we chose to use the first. This choice corresponds to the operation called *interior multiplication*, which you can read about in [Lee]. See p. 358 and Corollary 14.21 on p. 362.

Proof. Since $P_{\mathbf{p}}(v_1, ..., v_n)$ is an n-parallelapiped, then $\partial P_{\mathbf{p}}(v_1, ..., v_n)$ is a $2n$-parallelapiped. Each face of $\partial P_{\mathbf{p}}(v_1, ..., v_n)$ is of the form $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$ or $P_{\mathbf{p}}(v_1, ..., \cancel{v_i}, ..., v_n)$. We show that the faces of the form $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$ have the same orientation as does $P_{\mathbf{p}}(v_1, ..., v_n)$ and that the faces of the form $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$ have the opposite orientation as does $P_{\mathbf{p}}(v_1, ..., v_n)$.

Let the orientation on $P_{\mathbf{p}}(v_1, ..., v_n)$ be given by an orientation form ω_P. By Theorem 7.47, the induced orientation form $\omega_{\partial P}$ on the boundary is defined by $\omega_{\partial P} = \omega_P(v_i|_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., \cancel{v_i|_{\mathbf{p}}}, ..., v_n|_{\mathbf{p}})$. Since $v_i|_{\mathbf{p}}$ is outward pointing on the face $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$, the orientation of $P_{\mathbf{p}+\mathbf{v}_i|_{\mathbf{p}}}(v_1, ..., \cancel{v_i}, ..., v_n)$ is the same as the orientation of $P_{\mathbf{p}}(v_1, ..., v_n)$. Conversely, $v_i|_{\mathbf{p}}$ is inward pointing on the face $P_{\mathbf{p}}(v_1, ..., \cancel{v_i}, ..., v_n)$, so the orientation of $P_{\mathbf{p}}(v_1, ..., \cancel{v_i}, ..., v_n)$ is same as the orientation given by
$\omega_P(-v_i|_{\mathbf{p}}, v_1|_{\mathbf{p}}, ..., \cancel{v_i|_{\mathbf{p}}}, ..., v_n|_{\mathbf{p}}) = -\omega_{\partial P}$. $\square$

8

Differential forms on manifolds

At long last, we have reached the chapter where we define *differential forms* and how to integrate them over manifolds. In this chapter, we also define an operation on differential forms called the *exterior derivative*, which generalizes the differential of the previous chapter. We will see that the div, grad and curl of vector calculus can be expressed in terms of the exterior derivative and the *Hodge-dual* operator. The *generalized Stokes' theorem*, an elegant generalization of the fundamental theorem of calculus, is the key result of this chapter, and equates the integral of an exterior derivative over a manifold to the integral of the "exterior antiderivative" over the manifold's boundary.

8.1 Differential forms

Definition 8.1. [Lee, p. 360] (Differential k-form).

Let M be a smooth n-manifold WWBOC. We define $\Lambda^k(T^*(M)) := \bigsqcup_{\mathbf{p} \in M} \Lambda^k(T^*_{\mathbf{p}}(M))$, so that $\Lambda^k(T^*(M))$ is the the subset of $T^0_k(T(M))$ of antisymmetric tensors. The vector space $\Lambda^k(T^*(M))$ can be shown to be $\binom{n}{k}$ dimensional, just as is the case for $\Lambda^k(V)$ when V is an n-dimensional vector space.

A *differential k-form on M* is a continuous function $M \to \Lambda^k(T^*(M))$. So, you might say that a differential k-form is a "antisymmetric $\binom{0}{k}$ tensor field" (remember, all tensor fields are continuous maps). The vector space of differential k-forms on M is denoted $\Omega^k(M)$.

Theorem 8.2. [Lee, p. 360] (Differential k-form expressed relative to a coordinate chart).

Let M be a smooth n-manifold. Given any smooth chart $(U, \mathbf{x})$ on M, where x^i is the ith component function of $\mathbf{x}$, it follows by definition that a differential k-form ω on U can be expressed as

$$\omega = \sum_{1 \leq i_1 < \ldots < i_k \leq n} f_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k},$$

where each $f_{i_1 \ldots i_k} : U \to \mathbb{R}$.

Theorem 8.3. [Lee, p. 360] (Smoothness of a differential form).

When thought of as a tensor field, a differential k-form is smooth iff its component functions are smooth.

Remark 8.4. A differential 0-form on M is a continuous function $M \to \mathbb{R}$.

8.1.1 Differential forms as actual functions

We will occasionally need to think of differential k-forms as objects that, when evaluated at $\mathbf{p} \in M$, can act on k vectors in $T_{\mathbf{p}}(M)$.

Derivation 8.5. (Differential forms as actual functions).

Let M be a smooth manifold WWBOC. Since we are able to think of elements of $\Lambda^k(V^*)$ as "actual functions" by identifying them with elements of $\widetilde{\Lambda}^k(V^*)$, we can think of the evaluation at a point $\mathbf{p} \in M$ of a differential k-forms on M, which is an element of $\Lambda^k(T^*_{\mathbf{p}}(M))$, as an "actual function", by identifying it with an element of $\widetilde{\Lambda}^k(T^*_{\mathbf{p}}(M))$. Thus, we define $\widetilde{T}^*(M) := \bigsqcup_{\mathbf{p} \in M} \widetilde{\Lambda}^k(T^*_{\mathbf{p}}(M))$ so that we can define $\widetilde{\Omega}^k(M) := \{\text{continuous functions: } M \to \widetilde{T}^*(M)\}$ to be the set of differential k-forms that, when evaluated at a point, are actual functions- namely, multilinear alternating functions accepting k vectors from $T_{\mathbf{p}}(M)$.

8.2 Integration of differential forms on manifolds

In this section, we vaguely follow Chapter 16 of [Lee] (but take some cues from [HH], and fewer from [GP74]), and show how to integrate differential forms over manifolds. We will see that differential forms are the "natural" objects to integrate over manifolds because the pullback of a differential form interplays perfectly with the change of variables theorem from multivariable calculus.

On a more technical note, we will consider only compactly supported differential forms so that the integrals we consider are analogous to "proper" (as opposed to "improper") integrals of functions $f : \mathbb{R}^n \to \mathbb{R}$.

To begin, we define the pullback of a differential k-form by a diffeomorphism and show that it plays nicely with the change of variables theorem.

Pullback by a diffeomorphism and change of variables

Definition 8.6. [GP74, p. 163-165] (Pullback of a differential k-form by a diffeomorphism).

Let M and N be n-smooth manifolds, let $(U, \mathbf{x})$ and $(V, \mathbf{y})$ be smooth charts on M and N, respectively, and let $\mathbf{F} : U \subseteq M \to V \subseteq N$ be a smooth function. Consider a differential k-form $\omega \in \Omega^k(V)$, $k \leq n$, on $V \subseteq N$.

In Theorem 4.24, we showed how to pull back an element of $\Lambda^k(W^*)$ to obtain an element of $\Lambda^k(V^*)$, where V and W are vector spaces. Since the evaluation of ω at a point $\mathbf{q} \in V \subseteq N$ yields an element of $\Lambda^k T^*_{\mathbf{p}}(N)$, we can pull back differential k-forms on N to differential k-forms on M by using the pullback map on exterior powers.

We define the *pullback* $\Omega^k \mathbf{F}^* : \Omega^k(V) \to \Omega^k(U)$ *of the differential* k*-form* ω *on* V to be

$$\boxed{\Big((\Omega^k \mathbf{F}^*)(\omega)\Big)_{\mathbf{p}} := \begin{cases} \Big(\Lambda^k (d\mathbf{F}_{\mathbf{p}})^*\Big)(\omega_{\mathbf{F}(\mathbf{p})}) & \omega \text{ is a differential } k\text{-form on } V,\ k \geq 1 \\ (f \circ \mathbf{F})(\mathbf{p}) & f \text{ is a differential 0-form on } V \text{ (i.e. } f \text{ is a function } V \to \mathbb{R}) \end{cases}}$$

Here's a brief explanation of the notation. The map $(d\mathbf{F}_{\mathbf{p}})^* : T^*_{\mathbf{F}(\mathbf{p})}(V) \to T^*_{\mathbf{p}}(U)$ is the dual of the differential $d\mathbf{F}_{\mathbf{p}} : T_{\mathbf{p}}(U) \to T_{\mathbf{F}(\mathbf{p})}(V)$, which is a linear map, and $\Lambda^k(d\mathbf{F}_{\mathbf{p}})^*$ is the pullback from the kth exterior power $\Lambda^k(T^*_{\mathbf{F}(\mathbf{p})}(V))$ to the kth exterior power $\Lambda^k(T^*_{\mathbf{p}}(U))$. (For the exact meaning of $\Lambda^k(d\mathbf{F}_{\mathbf{p}})^*$, recall Definition 4.24).

Theorem 8.7. (Basic properties of pullbacks of differential forms).

Let M and N be smooth n-manifolds, let $U \subseteq M$ be open, and consider a smooth function $\mathbf{F} : U \subseteq M \to N$.

1. If ω is a differential 1-form on N, then $\Omega^k \mathbf{F}^*(\omega)_{\mathbf{p}} = (d\mathbf{F}_{\mathbf{p}})^*(\omega)$, where $(d\mathbf{F}_{\mathbf{p}})^*$ is the dual of the linear map $d\mathbf{F}_{\mathbf{p}}$.
2. If ω and η are differential k-forms, $k \leq n$, on N, then

$$\Omega^k \mathbf{F}^*(\omega \wedge \eta) = \Omega^k \mathbf{F}^*(\omega) \wedge \Omega^k \mathbf{F}^*(\eta).$$

Proof. All items follow straightforwardly from the definition of $\Omega^k \mathbf{F}^*$. □

Theorem 8.8. (The differential commutes with the pullback of a diffeomorphism).

Let M and N be smooth manifolds WWBOC, let $U \subseteq M$ be open, and consider a smooth function $\mathbf{F} : U \subseteq M \to N$. Additionally, let $\mathbf{p} \in U \subseteq M$, and let $d_{\mathbf{p}}$ denote the map $f \overset{d_{\mathbf{p}}}{\mapsto} df_{\mathbf{p}}$. Then for any smooth function $f : \mathbb{R}^n \to U \subseteq M$, we have

$$\Omega^k \mathbf{F}^* \circ d_{\mathbf{p}} = d_{\mathbf{p}} \circ \Omega^k \mathbf{F}^* \iff \Omega^k \mathbf{F}^*(d_{\mathbf{p}} f) = d_{\mathbf{p}}(\Omega^k \mathbf{F}^*(f)).$$

Here, the differential d can be interpreted in both of the ways we have mentioned earlier. (That is, d can be considered to be the differential which results from identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ or the differential in which this identification is not performed, and $T_{f(\mathbf{p})}(\mathbb{R})$ is "left alone" (recall Theorem 7.33 and Definition 7.31, respectively).

Proof. For this proof, we will assume that d is the differential in which $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ is "left alone", so that $df_{\mathbf{p}}$ acts on a smooth function $u_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by $df_{\mathbf{p}}(u_{\mathbf{p}}) = u_{\mathbf{p}}(f)$ (recall Theorem 7.33). (After we have proven the theorem for this interpretation of d, the theorem holds for when d is obtained by identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ precisely because we have an isomorphism $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$).

Let $v_{\mathbf{p}} \in T_{\mathbf{p}}(M)$. We will show $\Omega^k\mathbf{F}^*(df_{\mathbf{p}})(v_{\mathbf{p}}) = d(\Omega^k\mathbf{F}^*(f))_{\mathbf{p}}$.

By property (1) of the previous theorem, $\Omega^k\mathbf{F}^*(df_{\mathbf{p}})(v_{\mathbf{p}}) = (d\mathbf{F}_{\mathbf{p}})^*(df_{\mathbf{p}})(v_{\mathbf{p}})$. Then $(d\mathbf{F}_{\mathbf{p}})^*(df_{\mathbf{p}})(v_{\mathbf{p}}) = (df_{\mathbf{p}} \circ d\mathbf{F}_{\mathbf{p}})(v_{\mathbf{p}}) = df_{\mathbf{p}}(d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}}))$. Since $df_{\mathbf{p}}(u_{\mathbf{p}}) = u_{\mathbf{p}}(f)$, we use $u_{\mathbf{p}} = d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})$ to obtain $df_{\mathbf{p}}(d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}})) = (d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}}))(f)$. By the definition of the differential $d\mathbf{F}_{\mathbf{p}}$ (recall Definition 7.31), $(d\mathbf{F}_{\mathbf{p}}(v_{\mathbf{p}}))(f) = v_{\mathbf{p}}(f \circ \mathbf{F})$. Since f is a 0-form, $f \circ \mathbf{F} = \Omega^k\mathbf{F}^*(f)$ by definition of $\Omega^k\mathbf{F}^*$. So $v_{\mathbf{p}}(f \circ \mathbf{F}) = v_{\mathbf{p}}(\Omega^k\mathbf{F}^*(f))$. Recall that because d is the differential in which $T_{f(\mathbf{p})}(\mathbb{R})$ is "left alone" (see the remarks at the beginning of this proof), we have $v_{\mathbf{p}}(g) = dg_{\mathbf{p}}(v_{\mathbf{p}})$. We set $g = \Omega^k\mathbf{F}^*(f)$ to obtain $v_{\mathbf{p}}(\Omega^k\mathbf{F}^*(f)) = d(\Omega^k\mathbf{F}^*(f))_{\mathbf{p}}(v_{\mathbf{p}})$. $\square$

Lemma 8.9. (Technical lemma: $\Omega^k\mathbf{F}^*(dy^i) = \widetilde{F}^i_{(V,\mathbf{y})}$).

Let M and N be smooth n-manifolds WWBOC, let $(U, \mathbf{x})$ be a smooth chart on M, and let $(V, \mathbf{y})$ be a smooth chart on N. Let $\mathbf{F} : U \subseteq M \to V \subseteq N$ be a diffeomorphism. Recall from Definition 7.9 that the coordinate representation of $\mathbf{F}$ relative to the chart $(V, \mathbf{y})$ is $\widetilde{\mathbf{F}}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{F}$, so $y^i \circ \mathbf{F}$ is $y^i \circ \mathbf{F} = \widetilde{F}^i$, the ith coordinate function of $\widetilde{\mathbf{F}}_{(V,\mathbf{y})} = (\widetilde{F}^1_{(V,\mathbf{y})}, ..., \widetilde{F}^n_{(V,\mathbf{y})})^\top$. But $y^i \circ \mathbf{F} = \Omega^k\mathbf{F}^*(y^i)$, so

$$\widetilde{F}^i_{(V,\mathbf{y})} = \Omega^k\mathbf{F}^*(y^i).$$

Since the differential commutes with the pullback of a diffeomorphism, we have $\Omega^k\mathbf{F}^*(dy^i) = d(\Omega^k\mathbf{F}^*)(y^i)$. By the previous lemma, $\Omega^k\mathbf{F}^*(y^i) = \widetilde{F}^i_{(V,\mathbf{y})}$. Therefore

$$\Omega^k\mathbf{F}^*(dy^i) = \widetilde{F}^i_{(V,\mathbf{y})}.$$

Theorem 8.10. [Lee, p. 361] (Pullback of a top degree differential form).

Let M and N be smooth n-manifolds WWBOC, let $(U, \mathbf{x})$ be a smooth chart on M, and let $(V, \mathbf{y})$ be a smooth chart on N. Let $\mathbf{F} : U \subseteq M \to V \subseteq N$ be a diffeomorphism, and consider a differential n-form $\omega \in \Omega^n(V)$ on $V \subseteq N$. (We say that ω is *top degree* differential form on N, since its degree (the "degree" of a differential k-form is k) is the same as the dimension of N). If $f \; dy^1 \wedge ... \wedge dy^n$ is a differential n-form on V, the pullback of $f \; dy^1 \wedge ... \wedge dy^n$ is

$$\boxed{\Omega^k\mathbf{F}^*(f \; dy^1 \wedge ... \wedge dy^n) = (f \circ \mathbf{F}) \det\Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j}\Big) dx^1 \wedge ... \wedge dx^n}$$

Here $\widetilde{\mathbf{F}}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{F}$ is the coordinate representation of $\mathbf{F}$ taken relative to the chart $(V, \mathbf{y})$, and $\widetilde{F}^i_{(V,\mathbf{y})}$ is the *ith* component function of $\widetilde{\mathbf{F}}_{(V,\mathbf{y})}$.

Proof. By Theorem 8.2, we can write ω in the chart $(V, \mathbf{y})$ as

$$\omega = f \; dy^1 \wedge ... \wedge dy^n,$$

where f is a continuous function $N \to \mathbb{R}$, and where y^i is the ith coordinate function of $\mathbf{y}$.

First, we simply rewrite the above after applying the fact $\Omega^k\mathbf{F}^*(f) = f \circ \mathbf{F}$:

$$\Big(\Omega^k\mathbf{F}^*(\omega)\Big)_{\mathbf{p}} = (f \circ \mathbf{F})(\mathbf{p}) \; \Big(\Omega^k\mathbf{F}^*(dy^1)\Big)_{\mathbf{p}} \wedge ... \wedge \Big(\Omega^k\mathbf{F}^*(dy^n)\Big)_{\mathbf{p}}.$$

Since $\Omega^k \mathbf{F}^*(dy^i) = \widetilde{F}^i_{(V,\mathbf{y})}$ (see the previous lemma), we have

$$\Big(\Omega^k \mathbf{F}^*(f \ dy^1 \wedge ... \wedge dy^n)\Big)_{\mathbf{p}} = (f \circ \mathbf{F})(\mathbf{p}) \ d\widetilde{F}^1_{(V,\mathbf{y})}|_{\mathbf{p}} \wedge ... \wedge d\widetilde{F}^n_{(V,\mathbf{y})}|_{\mathbf{p}}.$$

Applying Theorem 4.70, we have

$$d\widetilde{F}^1_{(V,\mathbf{y})}|_{\mathbf{p}} \wedge ... \wedge d\widetilde{F}^n_{(V,\mathbf{y})}|_{\mathbf{p}} = \det\Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j}\Big) dx^1 \wedge ... \wedge dx^n.$$

Plug this expression for $d\widetilde{F}^1_{(V,\mathbf{y})}|_{\mathbf{p}} \wedge ... \wedge d\widetilde{F}^n_{(V,\mathbf{y})}|_{\mathbf{p}}$ into the previous equation to obtain the theorem. □

Remark 8.11. (Star notation for the pullback of a top degree differential form).

In Remark 4.25, we introduced the practice of using $\mathbf{F}^*$ to denote a pullback. From now on, we denote the pullback $\Omega^k \mathbf{F}^* : T_{\mathbf{F}(\mathbf{p})}(V \subseteq N) \to T_{\mathbf{p}}(U \subseteq M)$ by $\mathbf{F}^* : T_{\mathbf{F}(\mathbf{p})}(V \subseteq N) \to T_{\mathbf{p}}(U \subseteq M)$.

Integrating differential forms

We now use the theorem of the previous section to start constructing the notion of integration of a differential form over manifolds. We need the following two technical definitions before doing so, however.

Definition 8.12. [Lee, p. 653] (Domain of integration in $\mathbb{R}^n$).

A *domain of integration in* $\mathbb{R}^n$ is a bounded subset of $\mathbb{R}^n$ whose boundary has an n-dimensional measure of zero.

Definition 8.13. [Lee, p. 43] (Support of a differential form on a manifold).

Let M be a (not necessarily smooth) manifold and let ω be a differential form on M. The *support* of ω is defined to be the closure of the set of points where ω is nonzero, $\text{supp}(\omega) := \text{cl}(M - \omega^{-1}(\mathbf{0}))$. Iff $\text{supp}(\omega) \subseteq A$, then we say ω *is supported in* A. We say ω is *compactly supported* iff $\text{supp}(\omega)$ is compact.

Now, we use the formula for the pullback of a diffeomorphism derived at the end of the previous section to restate the change of variables theorem.

Derivation 8.14. [GP74, p. 166] [Lee, p. 403] (Change of variables theorem in light of the pullback, part 1).

The change of variables theorem says that if $(D, \mathbf{x})$ and $(E, \mathbf{y})$ are charts on $\mathbb{R}^n$ or $\mathbb{H}^n$, where D and E are open domains of integration, then, for every continuous function $f : \text{cl}(E) \to \mathbb{R}$, we have

$$\int_D f = \int_E \Big|(f \circ \mathbf{F}) \det\Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j}\Big)\Big|.$$

Notice that, when $\Big|\det\Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j}\Big)\Big| = \det\Big(\frac{\partial \widetilde{F}^i_{(V,\mathbf{y})}}{\partial x^j}\Big)$, i.e., when $\det(d\mathbf{F}) > 0$, the integrand f of the left hand side is *almost* the pullback of $f \ dy^1 \wedge ... \wedge dy^k$. The integrand f of the left hand side *would* be this pullback if it were instead $f \ dx^1 \wedge ... \wedge dx^n$. This observation motivates the next definition.

Definition 8.15. [Lee, p. 402] (Integral of a top degree differential form on a domain of integration in $\mathbb{R}^n$ or $\mathbb{H}^n$).

Let D be a domain of integration in $\mathbb{R}^n$ or $\mathbb{H}^n$. Due our discovery in the previous derivation, we define the *integral of a compactly supported differential n-form $f \ dx^1 \wedge ... \wedge dx^n$ over D*, to be

$$\boxed{\int_D f \ dx^1 \wedge ... \wedge dx^n := \int_D f}$$

Remark 8.16. (The meaning of $dx_1...dx_n$).

We should note that our recent definition of the integral of a differential form gives meaning to the $dx_1...dx_n$ that is used as a placeholder in an integral. If we use the convention of writing the placeholder $dx_1...dx_n$ after the integrand, so that

$$\int_D f = \int_D f \; dx_1...dx_n,$$

then the definition of the integral of a differential n-form on an open domain of integration becomes

$$\int_D f \; dx_1 \wedge ... \wedge dx_n := \int_D f \; dx^1...dx^n.$$

In some sense, the placeholder $dx_1...dx_n$ is "secretly" $dx^1 \wedge ... \wedge dx^n$. So, while the definition technically defines the left hand side in terms of the right hand side, you might think of it as giving algebraic meaning to the old placeholder notation of the right hand side.

Remark 8.17. One could discover the definition of differential forms and the wedge product by starting with the change of variables theorem and trying to formalize the notion of "pulling back" by treating the notation $dy^1...dy^n$ as the formal symbol $dy^1 \wedge ... \wedge dy^n$. When approached this way, the involvement of the determinant implies the seeming-multilinearity and seeming-antisymmetry of the wedge product.

Theorem 8.18. [Lee, p. 403] (Change of variables theorem in light of the pullback, part 2).

Consider the hypotheses of the previous derivation, Derivation 8.14. *In the case that* **F** *is orientation-preserving or orientation-reversing*, the change of variables theorem can be restated as

$$\int_E f \; dy^1 \wedge ... \wedge dy^n = \begin{cases} \int_D \mathbf{F}^*(f \; dy^1 \wedge ... \wedge dy^n) & \mathbf{F} \text{ is orientation-preserving} \\ -\int_D \mathbf{F}^*(f \; dy^1 \wedge ... \wedge dy^n) & \mathbf{F} \text{ is orientation-reversing} \end{cases}.$$

To reemphasize what we discovered in the previous derivation, we again state the pullback of $f \; dy^1 \wedge ... \wedge dy^n$:

$$\mathbf{F}^*(f \; dy^1 \wedge ... \wedge dy^n) = \det(d\mathbf{F})(f \circ \mathbf{F})dx^1 \wedge ... \wedge dx^n.$$

(It is possible for **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 **F** over U).

This definition us to concisely state the change of variables theorem by involving the integrals and pullbacks of differential forms.

Theorem 8.19. [Lee, p. 404] (Change of variables theorem for top degree differential forms on open domains of integration).

Let D and E be open domains of integration in $\mathbb{R}^n$ or $\mathbb{H}^n$, and let ω be a compactly supported differential n-form on $\text{cl}(E)$. Given the above definition, the restatement of the change of variables theorem presented at the end of the previous derivation is further restated as

$$\boxed{\int_E \omega = \begin{cases} \int_D \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-preserving} \\ -\int_D \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-reversing} \end{cases}}$$

This most recent boxed equation is not merely a restatement of the change of variables theorem, but a generalization. Previously, we only had a change of variables theorem for real-valued functions defined on open domains of integration in $\mathbb{R}^n$ or $\mathbb{H}^n$; now, we have a change of variables theorem for differential forms defined on open domains of integration in $\mathbb{R}^n$ or $\mathbb{H}^n$.

At this point, we know how to integrate differential forms over open domains of integration in $\mathbb{R}^n$ or $\mathbb{H}^n$. As a stepping stone to defining integration on manifolds, we now define the notion of integrating differential forms over open subsets of $\mathbb{R}^n$ or $\mathbb{H}^n$.

Definition 8.20. [Lee, p. 404] (Integral of a top degree differential form over an open subset of $\mathbb{R}^n$ or $\mathbb{H}^n$).

Let U be an open subset of $\mathbb{R}^n$ or $\mathbb{H}^n$. We define the *integral of a differential n-form ω that is compactly supported on U* to be

$$\int_U \omega := \int_D \omega,$$

where D is any domain of integration in $\mathbb{R}^n$ or $\mathbb{H}^n$ containing $\operatorname{supp}(\omega)$. This definition does not depend on the choice of D; see [Lee, p. 403] for the details. The right hand side is interpreted with the definition of the previous derivation.

We now state a slight generalization of the change of variables theorem that relies on the previous definition.

Theorem 8.21. [Lee, p. 404] (Change of variables theorem for top degree differential forms on open subsets of $\mathbb{R}^n$ or $\mathbb{H}^n$).

Let U and V be open subsets of $\mathbb{R}^n$ or $\mathbb{H}^n$. If $\mathbf{F} : U \to V$ is a diffeomorphism and $\omega = f\ dx^1 \wedge ... \wedge dx^n$ is a compactly supported differential n-form on V, then

$$\boxed{\int_V \omega = \begin{cases} \int_U \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-preserving} \\ -\int_U \mathbf{F}^*(\omega) & \mathbf{F} \text{ is orientation-reversing} \end{cases}}$$

Proof. The proof is a matter of using technical properties of diffeomorphisms. See [Lee, p. 404]. □

As promised, we are now ready to define integration on smooth manifolds.

Definition 8.22. [Lee, p. 404] (Integral of a top degree differential form that is compactly supported in a single chart over a smooth manifold).

Let M be an oriented smooth n-manifold with or without boundary, let $(U, \mathbf{x})$ be a smooth oriented chart on M, and let ω be a differential n-form on M with compact support in U. We define the *integral of ω over M* to be

$$\int_M \omega := \begin{cases} \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega) & (U, \mathbf{x}) \text{ is positively oriented} \\ -\int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega) & (U, \mathbf{x}) \text{ is negatively oriented} \end{cases}.$$

(Recall from Definition 7.43 that a smooth chart is positively oriented iff ... and ...).

Here, have used the pullback $(\mathbf{x}^{-1})^*$ to produce a differential n-form $(\mathbf{x}^{-1})^*(\omega)$ that is compactly supported on the open subset $\mathbf{x}(U)$ of $\mathbb{R}^n$ or $\mathbb{H}^n$. The previous definition allows us to evaluate the integral of such a differential n-form, which the right hand side.

Note, this definition does not depend on the smooth chart $(U, \mathbf{x})$ for which $\operatorname{supp}(\omega) \subseteq U$; see [Lee, p. 404] for the proof of this.

The previous theorem has sound algebraic intuition, but we have not yet seen any geometric intuition. To find geometric insight, we consider following special case.

Integrating differential forms treated as actual functions

Derivation 8.23. (Pullback of a differential k-form, interpreted as a pointwise actual function, by a diffeomorphism).

The pullback of a "regular" differential form defined in Definition 8.6 induces a pullback on the differential forms that, when evaluated a point, are actual functions. We construct this induced pullback now.

Let M and N be n-smooth manifolds, let $(U, \mathbf{x})$ and $(V, \mathbf{y})$ be smooth charts on M and N, respectively, and let $\mathbf{F} : U \subseteq M \to V \subseteq N$ be a smooth function. Consider a differential k-form $\widetilde{\omega} \in \widetilde{\Omega}^k(V)$, $k \leq n$, on $V \subseteq N$. Note that $\widetilde{\omega}_{\mathbf{p}}$ is an alternating multilinear function for all $\mathbf{p} \in M$.

Referring back to Definition 8.6, we see the induced pullback $\widetilde{\Omega}^k \mathbf{F}^*$ is defined by:

$$\Big((\widetilde{\Omega}^k \mathbf{F}^*)(\widetilde{\omega})\Big)_{\mathbf{p}} := \begin{cases} \Big(\widetilde{\Lambda}^k (d\mathbf{F}_{\mathbf{p}})^*\Big)(\widetilde{\omega}_{\mathbf{F}(\mathbf{p})}) & \widetilde{\omega} \text{ is a differential } k\text{-form on } V,\ k \geq 1 \\ (f \circ \mathbf{F})(\mathbf{p}) & f \text{ is a differential 0-form on } V \text{ (i.e. } f \text{ is a function } V \to \mathbb{R}) \end{cases}$$

Since $\mathbf{T} \in \Lambda^k(W^*) \overset{\widetilde{\Lambda}^k \mathbf{f}^*}{\longmapsto} \mathbf{T} \circ \mathbf{f} = \mathbf{f}^*(\mathbf{T}) \in \Lambda^k(V^*)$ (see the equation in Theorem 4.74 right above the boxed equation), the above becomes

$$\begin{cases} (d\mathbf{F}_{\mathbf{p}})^*(\widetilde{\omega}_{\mathbf{F}(\mathbf{p})}) = \widetilde{\omega}_{\mathbf{F}(\mathbf{p})} \circ d\mathbf{F}_{\mathbf{p}} & \widetilde{\omega} \text{ is a differential } k\text{-form on } V,\ k \geq 1 \\ (f \circ \mathbf{F})(\mathbf{p}) & f \text{ is a differential 0-form on } V \text{ (i.e. } f \text{ is a function } V \to \mathbb{R}) \end{cases}.$$

(In the first case, $(d\mathbf{F}_{\mathbf{p}})^*$ denotes the dual of the linear map $d\mathbf{F}_{\mathbf{p}}$). Therefore, $\Omega^k \mathbf{F}^*(\widetilde{\omega})$ acts on k vectors $v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by

$$\boxed{\Omega^k \mathbf{F}^*(\widetilde{\omega})(v_1|_{\mathbf{p}}, ..., v_n|_{\mathbf{p}})) = \widetilde{\omega}_{\mathbf{F}(\mathbf{p})}(d\mathbf{F}_{\mathbf{p}}(v_1|_{\mathbf{p}})..., d\mathbf{F}_{\mathbf{p}}(v_n|_{\mathbf{p}}) \quad \widetilde{\omega} \text{ is a differential } k\text{-form on } V,\ k \geq 1}$$

When f is a 0-form, we still have $\Omega^k \mathbf{F}^*(f) = f \circ \mathbf{F}$.

(Note, instead of using the intermediary step of the equation before the above boxed equation, we could have directly applied the boxed equation of Theorem 4.74 to the case $k \geq 1$).

The next theorem describes a way to interpret the integral of a "regular" differential form $\omega \in \Omega^k(M)$ over $U \subseteq M$ as an integral of a corresponding differential form $\widetilde{\omega} \in \widetilde{\Omega}^k(M)$ evaluated on a basis of $T_{\mathbf{p}}(M)$.

Theorem 8.24. [HH, p. 515] (Integral of a top degree differential form, interpreted as a pointwise actual function, that is compactly supported in a single chart over a smooth manifold).

Let M be an oriented smooth n-manifold with or without boundary, let $(V, \mathbf{y})$ be a smooth oriented chart on M, let $\omega = f\ dy^1 \wedge ... \wedge dy^n \in \Omega^k(M)$ be a "regular" differential n-form on M with compact support in V, and let $\widetilde{\omega} = f\ dy^1 \widetilde{\wedge} ... \widetilde{\wedge} dy^n \in \widetilde{\Omega}^k(M)$ be the corresponding differential form that is pointwise a multilinear alternating function. Define a chart $(U, \mathbf{x})$ on $\mathbf{x}(U)$ by $U = \mathbf{y}(V)$ and $\mathbf{x} = \mathbf{y}^{-1}$, and let $\widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}$ denotes the ith coordinate function of $\mathbf{x}$ relative to the chart $(V, \mathbf{y})$. Then

$$\boxed{\int_U \omega = \pm \int_U \widetilde{\omega}\Big(\frac{\partial \mathbf{x}}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial \mathbf{x}}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}}\Big)}$$

Proof. By Definition 8.22, the integral of ω over M is

$$\int_M \omega = \int_M f\ dy^1 \wedge ... \wedge dy^n = \pm \int_{\mathbf{y}(V)} (\mathbf{y}^{-1})^*(f\ dy^1 \wedge ... \wedge dy^n) = \pm \int_U \mathbf{x}^*(f\ dy^1 \wedge ... \wedge dy^n),$$

where the $\pm$ sign depends on the orientation of the chart $(V, \mathbf{y})$. We are again using $(\mathbf{y}^{-1})^* = \mathbf{x}^*$ to denote the pullback $\Omega^k(\mathbf{y}^{-1})^* = \Omega^k \mathbf{x}^*$.

Since $\mathbf{x}^*(dy^i) = \mathbf{x}^i_{(V,\mathbf{y})}$ (recall Lemma 8.9), the above becomes

$$\pm \int_U \mathbf{x}^*(f\ dy^1 \wedge ... \wedge dy^n) = \pm \int_U (f \circ \mathbf{x})\ d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}.$$

Here, $\widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}$ denotes the ith coordinate function of $\mathbf{x}$ relative to the chart $(V, \mathbf{y})$. (Sidenote: $\mathbf{x}$ relative to $(V, \mathbf{y})$ is $\mathbf{x}_{(V,\mathbf{y})} = \mathbf{y} \circ \mathbf{x} = \mathbf{x}^{-1} \circ \mathbf{x} = \mathbf{I}_U$).

Definition 8.15 says that the integral of a differential form over a subset of $\mathbb{R}^n$ or $\mathbb{H}^n$ is computed by "erasing" the $d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}$. Thus, we have

$$\pm \int_U (f \circ \mathbf{x})\ d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})} = \pm \int_V f \circ \mathbf{x}.$$

We now consider the differential form $\mathbf{x}^*(\widetilde{\omega}) = (f \circ \mathbf{x})\ d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})} \in \widetilde{\Omega}^k(U)$ that corresponds to the "regular" differential form $\mathbf{x}^*(\omega) = (f \circ \mathbf{x})\ d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \wedge ... \wedge d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})} \in \Omega^k(U)$. Applying Lemma 4.69 at each point in U, we have

$$d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})} \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}} \Big) = 1.$$

(Here, $\frac{\partial}{\partial \widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}}$ denotes the map $\mathbf{p} \mapsto \frac{\partial}{\partial \widetilde{\mathbf{x}}^i_{(V,\mathbf{y})}}\Big|_{\mathbf{p}}$). Thus

$$\pm \int_U f \circ \mathbf{x} = \pm \int_U \underbrace{(f \circ \mathbf{x}) d\widetilde{\mathbf{x}}^1_{(V,\mathbf{y})} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}}_{\mathbf{x}^*(\widetilde{\omega})} \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}} \Big) = \pm \int_U \mathbf{x}^*(\widetilde{\omega}) \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}} \Big).$$

We have translated our integral into one that is expressed in terms of a differential form acting on tangent vectors. Now, we apply the pullback from the previous theorem that acts on such differential forms to see

$$\pm \int_U \mathbf{x}^*(\widetilde{\omega}) \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}} \Big) = \pm \int_U \widetilde{\omega} \Big(d\mathbf{x} \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}} \Big), ..., d\mathbf{x} \Big(\frac{\partial}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}} \Big) \Big) = \pm \int_U \widetilde{\omega} \Big(\frac{\partial \mathbf{x}}{\partial \widetilde{\mathbf{x}}^1_{(V,\mathbf{y})}}, ..., \frac{\partial \mathbf{x}}{\partial \widetilde{\mathbf{x}}^n_{(V,\mathbf{y})}} \Big).$$

($d\mathbf{x}$ is shorthand for the map $d\mathbf{x} \mapsto d\mathbf{x}_{\mathbf{p}}$). $\square$

Theorem 8.25. (Integral of a differential top degree differential form over a smooth chart is "linear" with respect to the region of integration).

Let M be a smooth n-manifold, let $(U, \mathbf{x})$ be a smooth oriented chart on M, and consider a smooth differential n-form ω that is compactly supported in $\bigcup_{i=1}^k U_i$, where each $U_i \subseteq U$, and where $U_i \cap U_j$ has measure zero[1] for all $i \neq j$. Then

$$\int_U \omega = \sum_{i=1}^k \int_{U_i} \omega$$

Proof. The proof of this theorem adapts the proof of Proposition 16.8 in [Lee] (which was intended for a slightly different purpose in that book). By definition,

$$\int_U \omega = \pm \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega),$$

where the $\pm$ sign depends on the orientation of $(U, \mathbf{x})$. Since $\text{supp}(\omega) \subseteq \bigcup_{i=1}^k U_i$, then

Since $\text{supp}(\omega) \subseteq \bigcup_{i=1}^k U_i$, then $\text{supp}((\mathbf{x}^{-1})^*) \subseteq \bigcup_{i=1}^k \mathbf{x}(U_i)$. Additionally, since $\mathbf{x}$ is a smooth map, then $U_i \cap U_j$ having measure zero for all $i \neq j$ implies[2] that $\mathbf{x}(U_i) \cap \mathbf{x}(U_j)$ has measure zero for all $i \neq j$.

[1]Informally, a subset of $\mathbb{R}^n$ has *measure zero* iff its volume is zero.

[2]*Sard's theorem* is what is used to formally prove this implication.

Thus, using the standard calculus theorem pertaining to "breaking up integrals" (Theorem 5.15), we see $\mathbf{x}(U_i), ..., \mathbf{x}(U_k)$ satisfy the conditions that are required to "break up" an integral. Applying this same theorem, we have

$$\pm \int_{\mathbf{x}(U)} (\mathbf{x}^{-1})^*(\omega) = \sum_{i=1}^{k} \pm \int_{\mathbf{x}(U_i)} (\mathbf{x}^{-1})^*(\omega) = \sum_{i=1}^{k} \int_{U_i} \omega,$$

as claimed. □

The previous theorem is generalized by the following definition, which finalizes the definition of integration of differential forms.

Definition 8.26. [Lee, p. 408] (Integral of a top degree differential form on a smooth manifold).

Let M be an oriented smooth n-manifold WWBOC and let ω be a compactly supported n-form on M. Since $\text{supp}(\omega)$ is compact, there is a finite collection $\{(U_i, \mathbf{x}_i)\}_{i=1}^k$ of charts on M for which $\{U_i\}_{i=1}^k$ is an open cover of $\text{supp}(\omega)$, where the pairwise intersections have measure zero.

We define *the integral of ω over M to be*

$$\boxed{\int_M \omega := \sum_{i=1}^{k} \int_{U_i} \omega}$$

Note that each integral in the sum on the right hand side is interpreted with the previous definition.

It is necessary to show that this definition doesn't depend on the choice of open cover. See [**?**] for this detail.

Theorem 8.27. (Integral of a differential form on a smooth manifold is linear).

Let M be an oriented smooth n-manifold WWBOC, and let $(U, \mathbf{x})$ be a smooth chart on M. Then for all constant functions $c_1, c_2 : U \subseteq M \to \mathbb{R}$ and all compactly supported differential k-forms ω, η on M, we have

$$\int_M c_1\omega + c_2\eta = c_1 \int_M \omega + c_2 \int_M \eta.$$

Proof. It suffices to prove that this theorem holds when ω and η are supported in the same single smooth chart $(U, \mathbf{x})$. After applying Definition 8.22- pull back ω and η and integrate each over $\mathbf{x}(U) \subseteq \mathbb{R}^n$- we see the theorem holds in this special case due to the linearity of the integral of a function $\mathbb{R}^n \to \mathbb{R}$. □

8.3 The exterior derivative

To define a notion of differentiating differential forms, we will interpret differential forms to be objects that, when evaluated a point of a manifold, are multilinear alternating functions accepting tangent vectors as input.

Definition 8.28. [HH, p. 545] (The exterior derivative).

If V is a compact region in $\mathbb{R}^3$ with volume $|V|$, then the divergence div of a smooth vector field $\mathbf{V} : V \subseteq \mathbb{R}^3 \to \mathbb{R}^3$ is defined as flux per "infinitesimal" volume:

$$\text{div}(\mathbf{V}) := \lim_{|V| \to 0} \frac{1}{|V|} \int_{\partial V} \mathbf{V} \cdot \hat{\mathbf{n}} dS.$$

Similarly, if C is a simple closed curve in $\mathbb{R}^3$, $A(C)$ is the oriented area enclosed by C, and $\hat{\mathbf{n}}$ is the positively oriented unit vector that is normal to $A(C)$, then $\text{curl}(\mathbf{V})$ is defined as work per "infinitesimal" surface area:

$$\text{curl}(\mathbf{V}) \cdot \hat{\mathbf{n}} := \lim_{A(C) \to 0} \frac{1}{A(C)} \int_C \mathbf{V} \cdot d\mathbf{r} = \lim_{A(C) \to 0} \frac{1}{A(C)} \int_C \mathbf{V} \cdot \frac{d\mathbf{r}}{dt} dt.$$

We present a notion of derivative on differential forms that is defined analagously to divergence and curl.

Let M be a smooth n-manifold, let $(U, \mathbf{x})$ be a smooth chart on M, and let ω be a smooth differential k-form (thought of as an element of $\widetilde{\Omega}^{k+1}(M)$) with compact support in U. Let $P_{\mathbf{p}}(\mathbf{v}_1, ..., \mathbf{v}_{k+1})$ denote the subset of $T_{\mathbf{p}}(U)$ spanned by $\mathbf{v}_1|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}} \in T_{\mathbf{p}}(U)$. We say $P_{\mathbf{p}}(\mathbf{v}_1, ..., \mathbf{v}_{k+1})$ is a $(k+1)$-*parallelapiped.*

We define the *exterior derivative* $d\omega$ of ω to be the differential $(k+1)$-form defined at $\mathbf{p} \in M$ by

$$\boxed{d\omega_{\mathbf{p}}(\mathbf{v}_1, ..., \mathbf{v}_{k+1}) := \lim_{h \to 0} \frac{1}{h^{k+1}} \int_{\partial P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})} \omega.}$$

In words, the exterior derivative $d\omega$ is evaluated at $\mathbf{p}$ on $k+1$ tangent vectors $\mathbf{v}_1|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}} \in T_{\mathbf{p}}(M)$ by considering the "infinitesimally small" $(k+1)$-parallelapiped in $T_{\mathbf{p}}(M)$ spanned by these vectors, and then integrating ω over the faces of the $(k+1)$-parallelpiped, each of which is an "infinitesimally small" k-parallelapiped.

As is noted in [HH, p. 545], it is not immediately obvious why the limit in the definition of the exterior derivative exists. Since the integral in the limit can be broken up into a sum of integrals over the faces of the $(k+1)$-paralellapiped, and as the ith edge of the $(k+1)$-parallelapiped is $h\mathbf{v}_i|_{\mathbf{p}}$, it seems that each integral in the sum is dominated by $h^{(k+1)-1} = h^k$. (The faces of the paralellapiped are $(k+1) - 1 = k$ dimensional). The apparent problem comes when we move $\frac{1}{h^{k+1}}$ inside the limit, so that each integral is now dominated by $\frac{h^k}{h^{k+1}} = \frac{1}{h}$. This seems problematic because $\lim_{h \to 0} \frac{1}{h}$ does not exist. We will see in the proof below that the limit *does* exist.

Theorem 8.29. [HH, p. 652 - 655] (Computing the exterior derivative).

Let M be a smooth n-dimensional manifold with or without boundary or corners, and let $(U, \mathbf{x})$ be a smooth chart on M, where x^i is the ith component function of $\mathbf{x}$.

For any smooth function $f : U \to \mathbb{R}$,

1. The exterior derivative (denoted df) is equal to the differential (also denoted df), where the differential of f is the differential after identifying $T_{f(\mathbf{p})}(\mathbb{R}) \cong \mathbb{R}$ described by Theorem 7.33).
2. $d(f \ dx^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dx^{i_k}) = df \widetilde{\wedge} dx^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dx^{i_k}$.

Notice that (1) implies that the exterior derivative of any constant differential form is the zero differential form 0.

Proof.

1. $df(P_{\mathbf{p}}(\mathbf{v})) = \lim_{h\to 0}\frac{1}{h}\Big(f(\mathbf{x}+h\mathbf{v}) - f(\mathbf{x})\Big)$. Recalling Theorem 5.8, we see that this expression is exactly the directional derivative of f at $\mathbf{p}$ in the direction of $\mathbf{v}$.

2. The $(k+1)$-parallelapiped $P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})$ has $2(k+1)$ faces. For each $h > 0$, we construct charts $\{(V_{\mathrm{fi}}^h, \mathbf{y}_{\mathrm{fi}})\}_{i=1}^{k+1}$ that parameterize the $k+1$ faces on the "front" of this $(k+1)$-parallelapiped and charts $\{(V_{\mathrm{b}i}^h, \mathbf{y}_{\mathrm{b}i})\}_{i=1}^{k+1}$ that parameterize the remaining $k+1$ faces on its "back".

   Specifically, let $\mathbf{v}_i|_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$ be the image of $v_i|_{\mathbf{p}}$ under the isomorphism $T_{\mathbf{p}}(P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})) \cong \mathbb{R}^n_{\mathbf{p}}$ that sends $\frac{\partial}{\partial y^i} \mapsto \hat{\mathfrak{e}}_i$.We define, for $\mathbf{q} \in [0,h]^k \subseteq \mathbb{R}^k$,

$$\mathbf{y}_{\mathrm{fi}}^{-1}(\mathbf{q}) := q^1\mathbf{v}_1|_{\mathbf{p}} + ... + q^{i-1}\mathbf{v}_{i-1}|_{\mathbf{p}} + \cancel{q^i\mathbf{v}_i|_{\mathbf{p}}} + q^{i+1}\mathbf{v}_{i+1}|_{\mathbf{p}} + ... + q^{k+1}\mathbf{v}_{k+1}|_{\mathbf{p}}$$
$$\mathbf{y}_{\mathrm{b}i}^{-1}(\mathbf{q}) := q^1\mathbf{v}_1|_{\mathbf{p}} + ... + q^{i-1}\mathbf{v}_{i-1}|_{\mathbf{p}} + h\mathbf{v}_i|_{\mathbf{p}} + q^{i+1}\mathbf{v}_{i+1}|_{\mathbf{p}} + ... + q^{k+1}\mathbf{v}_{k+1}|_{\mathbf{p}}.$$

   Following the conventions of Theorem 8.24, we define $U_{\mathrm{fi}}^h = \mathbf{y}(V_{\mathrm{fi}}^h)$, $\mathbf{x}_{\mathrm{fi}} = \mathbf{y}^{-1}$ and $U_{\mathrm{b}i}^h = \mathbf{y}(V_{\mathrm{b}i}^h)$, $\mathbf{x}_{\mathrm{b}i} = \mathbf{y}^{-1}$ so that $\{(U_{\mathrm{fi}}^h, \mathbf{x}_{\mathrm{fi}})\}_{i=1}^{k+1}$ and $\{(U_{\mathrm{b}i}^h, \mathbf{x}_{\mathrm{b}i})\}_{i=1}^{k+1}$ are charts in $\mathbb{R}^n$ or $\mathbb{H}^n$. Note that $\mathbf{x}_{\mathrm{b}i} = \mathbf{y}_{\mathrm{b}i}^{-1} = \mathbf{y}_{\mathrm{fi}}^{-1} + h\mathbf{v}_i = \mathbf{x}_{\mathrm{fi}} + h\mathbf{v}^i$. Additionally, since we required $\mathbf{q} \in [0,h]^k \subseteq \mathbb{R}^k$, then $U_{\mathrm{fi}}^h \sqcup U_{\mathrm{b}i}^h$ is a cube of side length h in $\mathbb{R}^k$.

   We need to compute

$$\lim_{h\to 0}\frac{1}{h^{k+1}}\int_{\partial P_{\mathbf{p}}(hv_1|_{\mathbf{p}},...,hv_{k+1}|_{\mathbf{p}})} f\ dy^{i_1}\widetilde{\wedge}...\widetilde{\wedge}dy^{i_k}.$$

   We use Theorem 8.24 to treat the integral as the integral of a differential form evaluated on tangent vectors. Also using Theorem 8.25 to break up the domain of integration and Theorem 7.49 to account for the orientation of the boundary $\partial P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})$, the integral inside the limit becomes

$$\begin{aligned}\int_{\partial P_{\mathbf{p}}(\mathbf{v}_1,...,\mathbf{v}_{k+1})} f\ d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k} = &\underbrace{-\sum_{i=1}^{k+1}(-1)^{i-1}\int_{U_{\mathrm{fi}}^h}(f\circ\mathbf{x}_{\mathrm{fi}})\ d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k}\Big(\frac{\partial\mathbf{x}_{\mathrm{fi}}}{\partial q^1}, ..., \cancel{\frac{\partial\mathbf{x}_{\mathrm{fi}}}{\partial q^i}}, ..., \frac{\partial\mathbf{x}_{\mathrm{fi}}}{\partial q^{k+1}}\Big)}_{\text{sum of integrals over front faces}}\\ &\underbrace{+\sum_{i=1}^{k+1}(-1)^{i-1}\int_{U_{\mathrm{b}i}^h}(f\circ\mathbf{x}_{\mathrm{b}i})\ d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k}\Big(\frac{\partial\mathbf{x}_{\mathrm{b}i}}{\partial q^1}, ..., \cancel{\frac{\partial\mathbf{x}_{\mathrm{b}i}}{\partial q^i}}, ..., \frac{\partial\mathbf{x}_{\mathrm{b}i}}{\partial q^{k+1}}\Big)}_{\text{sum of integrals over back faces}}.\end{aligned}$$

   For $j \neq i$, we have $\frac{\partial\mathbf{x}_{\mathrm{fi}}}{\partial q^j} = \frac{\partial\mathbf{x}_{\mathrm{b}i}}{\partial q^j} = \mathbf{v}_j|_{\mathbf{p}}$. Since the argument of $d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k}$ in each integrand is the same, the above is

$$\sum_{i=1}^{k+1}(-1)^{i-1}\int_{U_{\mathrm{fi}}^h\sqcup U_{\mathrm{b}i}^h}(f\circ\mathbf{x}_{\mathrm{b}i} - f\circ\mathbf{x}_{\mathrm{fi}})\ d\widetilde{\mathbf{x}}^{i_1}\widetilde{\wedge}...\widetilde{\wedge}d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).$$

   Recall from the hypotheses of the theorem that f is assumed to be a continuous function $U \to \mathbb{R}$, where $(\mathbf{x}, U)$ is a smooth chart on M. (Don't confuse $\mathbf{x} : U \to \mathbb{R}^n$ with $\mathbf{x}_{\mathrm{fi}} : U_{\mathrm{fi}}^h \subseteq \mathbb{R}^n \to U$ and $\mathbf{x}_{\mathrm{b}i} : U_{\mathrm{b}i}^h \subseteq \mathbb{R}^n \to U$!). We now write $f : U \to \mathbb{R}$ as a Taylor polynomial centered at $\mathbf{p}_0 \in U$ with remainder:

$$T_f(\mathbf{p}) = T_f^0(\mathbf{p}) + T_f^1(\mathbf{p}) + R_f(\mathbf{p}) = f(\mathbf{p}_0) + df_{\mathbf{p}_0}(\mathbf{p}) + R_f(\mathbf{p}),$$

where $|R_f(\mathbf{p})| \leq C_1||\mathbf{x}(\mathbf{p})||^2$ for some $C_1 > 0$.

Plugging this Taylor expansion into the term $f \circ \mathbf{x}_{\text{f}i} - f \circ \mathbf{x}_{\text{b}i}$ from the above integrand, we see

$$\begin{aligned}
T_f \circ \mathbf{x}_{\text{b}i} - T_f \circ \mathbf{x}_{\text{f}i} &= (T_f^0 - T_f^1 + R_f) \circ \mathbf{x}_{\text{b}i} - (T_f^0 - T_f^1 + R_f) \circ \mathbf{x}_{\text{f}i} \\
&= (T_f^0 \circ \mathbf{x}_{\text{b}i} - T_f^0 \circ \mathbf{x}_{\text{f}i}) + (T_f^1 \circ \mathbf{x}_{\text{b}i} - T_f^1 \circ \mathbf{x}_{\text{f}i}) + (R_f \circ \mathbf{x}_{\text{b}i} - R_f \circ \mathbf{x}_{\text{f}i}).
\end{aligned}$$

Since T_f^0 is a constant function, the first term is zero: $T_f^0 \circ \mathbf{x}_{\text{f}i} - T_f^0 \circ \mathbf{x}_{\text{b}i} = 0$. We now plug this term into the integrand to complete the proof. The integral from the above sum of integrals (where the sum of integrals is inside the limit) becomes

$$\begin{aligned}
&\int_{U_{\text{f}i}^h \sqcup U_{\text{b}i}^h} (T_f^1 \circ \mathbf{x}_{\text{b}i} - T_f^1 \circ \mathbf{x}_{\text{f}i}) d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i}|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}) \\
&+ \int_{U_{\text{f}i}^h \sqcup U_{\text{b}i}^h} (R_f \circ \mathbf{x}_{\text{b}i} - R_f \circ \mathbf{x}_{\text{f}i}) d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i}|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).
\end{aligned}$$

Observe that $|R_f(\mathbf{p})| \leq C_1|| \max_{\mathbf{x}(\mathbf{p})} \mathbf{x}(\mathbf{p})||^2 = C_1 h^2$, which implies $||R_f \circ \mathbf{x}_{\text{f}i} - R_f \circ \mathbf{x}_{\text{b}i}||^2 \leq C_2 h^2$ for some $C_2 > 0$. Since the $\mathbf{v}_i|_{\mathbf{p}}$'s in $d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i}|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}})$ are constant with respect to the limit, which takes $h \to 0$, then $|d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i}|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}})|$ is bounded above with respect to h. Thus, the integrals containing the terms $R_f \circ \mathbf{x}_{\text{b}i} - R_f \circ \mathbf{x}_{\text{f}i}$ disappear in the limit. This means that the limit becomes

$$\begin{aligned}
&\lim_{h \to 0} \frac{1}{h^{k+1}} \int_{\partial P_{\mathbf{p}}(hv_1|_{\mathbf{p}}, ..., hv_{k+1}|_{\mathbf{p}})} f \ dy^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dy^{i_k} \\
&= \lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} \int_{U_{\text{f}i}^h \sqcup U_{\text{b}i}^h} (T_f^1 \circ \mathbf{x}_{\text{b}i} - T_f^1 \circ \mathbf{x}_{\text{f}i}) d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i}|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}})
\end{aligned}$$

Observe that the term $T_f^1 \circ \mathbf{x}_{\text{b}i} - T_f^1 \circ \mathbf{x}_{\text{f}i}$ in the integrand is $T_f^1 \circ \mathbf{x}_{\text{b}i} - T_f^1 \circ \mathbf{x}_{\text{f}i} = hdf_{\mathbf{p}_0}(\mathbf{v}_i)$:

$$\begin{aligned}
(T_f^1 \circ \mathbf{x}_{\text{b}i} - T_f^1 \circ \mathbf{x}_{\text{f}i})(\mathbf{p}) &= df_{\mathbf{p}_0}(\mathbf{x}_{\text{b}i}(\mathbf{p})) - df_{\mathbf{p}_0}(\mathbf{x}_{\text{f}i}(\mathbf{p})) = df_{\mathbf{p}_0}(\mathbf{x}_{\text{b}i}(\mathbf{p})) - df_{\mathbf{p}_0}(\mathbf{x}_{\text{f}i}(\mathbf{p})) \\
&= df_{\mathbf{p}_0}(\mathbf{x}_{\text{f}i}(\mathbf{p}) + h\mathbf{v}_i) - d_{\mathbf{p}_0}(\mathbf{x}_{\text{f}i}(\mathbf{p})) = hdf_{\mathbf{p}_0}(\mathbf{v}_i).
\end{aligned}$$

Thus, the limit is

$$\lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} \int_{U_{\text{f}i}^h \sqcup U_{\text{b}i}^h} hdf_{\mathbf{p}_0}(\mathbf{v}_i) \ d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i}|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).$$

(Here, $df_{\mathbf{p}_0}(\mathbf{v}_i) \in U_{\text{f}i}^h \sqcup U_{\text{b}i}^h$ denotes the constant map $\mathbf{q} \in\mapsto df_{\mathbf{p}_0}(\mathbf{v}_i)$). Since the integral is taken over points in $U_{\text{f}i}^h \sqcup U_{\text{b}i}^h$, then $hdf_{\mathbf{p}_0}(\mathbf{v}_i)$ is constant with respect to the integral, and the above is

$$\lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} hdf_{\mathbf{p}_0}(\mathbf{v}_i) \int_{U_{\text{f}i}^h \sqcup U_{\text{b}i}^h} d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i}|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).$$

$U^h_{\text{fi}} \sqcup U^h_{\text{bi}}$ is a cube in $\mathbb{R}^k$ with side length h, so the limit at last disappears, since $\frac{1}{h^{k+1}}$ is canceled by the h and h^k inside the sum:

$$\begin{aligned}
&\lim_{h \to 0} \frac{1}{h^{k+1}} \sum_{i=1}^{k+1} (-1)^{i-1} h df_{\mathbf{p}_0}(\mathbf{v}_i) h^k \ d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}) \\
&= \sum_{i=1}^{k+1} (-1)^{i-1} df_{\mathbf{p}_0}(\mathbf{v}_i) \ d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \cancel{\mathbf{v}_i|_{\mathbf{p}}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}) \\
&= \text{alt}(df \widetilde{\otimes} (d\widetilde{\mathbf{x}}^{i_1} \widetilde{\otimes} ... \widetilde{\otimes} d\widetilde{\mathbf{x}}^{i_k}))(\mathbf{v}_1|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}) \\
&= df \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} d\widetilde{\mathbf{x}}^{i_k}(\mathbf{v}_1|_{\mathbf{p}}, ..., \mathbf{v}_{k+1}|_{\mathbf{p}}).
\end{aligned}$$

□

Theorem 8.30. [Lee, p. 366] (The exterior derivative commutes with the pullback of a smooth function $M \to N$).

Let M and N be smooth manifolds WWBOC, let $U \subseteq M$ be open, and consider a smooth function $\mathbf{F} : U \subseteq M \to N$. Additionally, let $\mathbf{p} \in U \subseteq M$, and let $d_{\mathbf{p}}$ denote the map $f \overset{d_{\mathbf{p}}}{\mapsto} df_{\mathbf{p}}$. Then for any smooth differential k-form $\omega \in \widetilde{\Omega}^k(N)$ on N, we have

$$\mathbf{F}^* \circ d_{\mathbf{p}} = d_{\mathbf{p}} \circ \mathbf{F}^* \iff \mathbf{F}^*(d_{\mathbf{p}}\omega) = d_{\mathbf{p}}(\mathbf{F}^*(\omega)).$$

Proof. The proof of this relies on extrapolating the result of Theorem 8.8. □

Theorem 8.31. [Lee, p. 364] ($d \circ d = \mathbf{0}$).

Let M be a smooth manifold WWBOC, and let $(U, \mathbf{x})$ be a smooth chart on M. For any smooth differential form ω on U, performing the exterior derivative twice on ω produces the zero differential form. That is, $d \circ d$ is the zero map, $d \circ d = \mathbf{0}$.

Proof. According to [Ebe19], this property follows from the fact that the exterior derivative commutes with the pullback of a diffeomorphism. Here is the rough idea.

A differential form ω on $\mathbb{R}^n$ is said to be *invariant* with respect to a diffeomorphism $\mathbf{F}$ iff $\mathbf{F}^*(\omega) = \omega$. A differential form ω on $\mathbb{R}^n$ that is invariant with respect to all translation diffeomorphisms is said to be *translation-invariant*. Using the fact that the exterior derivative commutes with the pullback of a diffeomorphism, we can show that the exterior derivative of a translation-invariant differential form must also be translation-invariant.

Consider the diffeomorphism $\mathbf{F}_x : \mathbb{R}^n \to \mathbb{R}^n$, where $x \in \mathbb{R}$, defined by $\mathbf{F}_x(\mathbf{q}) = x\mathbf{q}$. Supposedly, it is easy to check that each x acts on translation-invariant differential k-forms with multiplication by x^k. Using this fact, we show that all translation-invariant differential forms ω satisfy $d\omega = 0$. Since any differential form on $\mathbb{R}^n$ can be written as a linear combination of translation-invariant differential forms on $\mathbb{R}^n$, it follows that $d^2\omega = 0$ for any differential form ω on $\mathbb{R}^n$. This result is quickly extended to differential forms on arbitrary smooth manifolds.

We now prove the claim that all translation-invariant differential forms ω satisfy $d\omega = 0$. Suppose that ω is a translation invariant differential form that is acted on by x in the manner described above. Then $x^k\omega = \mathbf{F}_x^*(\omega)$. Taking the exterior derivative, we have $d(x^k\omega) = d(\mathbf{F}_x^*(\omega))$. We take out the constant on the left hand side and use the fact that the exterior derivative commutes with the pullback of a diffeomorphism on the right hand side to get $x^k d\omega = \mathbf{F}_x^*(d\omega)$. Then, since $d\omega$ is a differential $(k+1)$-form, $\mathbf{F}_x^*(d\omega) = x^{k+1}d\omega$. In all, we have $x^k d\omega = x^{k+1}d\omega$, so $d\omega = xd\omega$. Somehow, this implies that $d\omega = 0$. Thus, we have shown that all translation invariant differential forms ω satisfy $d\omega = 0$. □

Theorem 8.32. [HH, p. 652 - 655] (Linearity of the exterior derivative).

Let M be a smooth manifold WWBOC, and let $(U, \mathbf{x})$ be a smooth chart on M.

- $d(f_1\omega + f_2\eta) = f_1 d\omega + f_2 d\eta$ for all $\omega, \eta \in \widetilde{\Omega}^k(U)$
- $d(c\omega) = cd\omega$ for all $\omega \in \widetilde{\Omega}^k(U)$ when $c : U \to \mathbb{R}$ is a constant function.

Proof. By part (1) of Theorem 8.29, the exterior derivative on 0-forms is the differential, which is linear. So the exterior derivative is linear on 0-forms. Extend this result using part (2) of Theorem 8.29, which is $d(f\ dx^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dx^{i_k}) = df \widetilde{\wedge} dx^{i_1} \widetilde{\wedge} ... \widetilde{\wedge} dx^{i_k}$. □

Remark 8.33. [Lee, p. 364] (Exterior derivative on $\Omega^k(M)$).

Let $(U, \mathbf{x})$ be a smooth chart on M. The exterior derivative we have presented operates on differential forms from $\widetilde{\Omega}^k(M)$. There is a corresponding induced exterior derivative operation d on $\Omega^k(M)$ ("regular" differential forms) satisfying

1. $d(f\ dx^{i_1} \wedge ... \wedge dx^{i_k}) = df \wedge dx^{i_1} \wedge ... \wedge dx^{i_k}$ for all smooth functions $f : U \to \mathbb{R}$.
2. The induced exterior derivative commutes with the pullback of a smooth function $M \to N$.
3. d is linear.
4. d satisfies a "product rule": $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta$, for all $\omega \in \Omega^k(U)$ and $\eta \in \Omega^\ell(U)$.

It is shown in [Lee, p. 364] that these axioms indeed determine a unique operation.

The exterior derivative and vector calculus

We now present how the exterior derivative, when coupled with the *Hodge-dual* operator (soon to be introduced), generalizes div, grad, and curl from multivariable calculus.

Definition 8.34. (Hodge-dual).

Let V be an n-dimensional vector space. The Hodge star $*$ is the linear map $\Lambda^k(V) \to \Lambda^k(V)$ that, for any ordered basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ of V, satisfies

$$(\mathbf{e}_{i_1} \wedge ... \wedge \mathbf{e}_{i_k}) \wedge *(\mathbf{e}_{i_1} \wedge ... \wedge \mathbf{e}_{i_k}) = \mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n.$$

That is, for any permutation $\sigma \in S_n$,

$$*(\mathbf{e}_{\sigma(1)} \wedge ... \wedge \mathbf{e}_{\sigma(k)}) = \mathrm{sgn}(\sigma)(\mathbf{e}_{\sigma(k+1)} \wedge ... \wedge \mathbf{e}_{\sigma(n)}).$$

Theorem 8.35. (Div, grad, and curl via the exterior derivative).

Consider $\mathbb{R}^3$ as a smooth manifold, and let $\sharp$ and $\flat$ be the musical isomorphisms induced by the choice of basis $\left\{\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}\right\}_{i=1}^n$ for $T_{\mathbf{p}}(\mathbb{R}^3)$. This means that

$$\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \in T_{\mathbf{p}}(M) \overset{\flat}{\mapsto} dx^i|_{\mathbf{p}} \in T^*_{\mathbf{p}}(M)$$
$$dx^i|_{\mathbf{p}} \in T^*_{\mathbf{p}}(M) \overset{\sharp}{\mapsto} \frac{\partial}{\partial x^i}\mathbf{p} \in T_{\mathbf{p}}(M).$$

(It may be helpful to recall Theorems 7.38 and 7.39).

Additionally, let $\mathbf{F}$ be the isomorphism $T_{\mathbf{p}}(\mathbb{R}^3) \cong \mathbb{R}^3$ sending $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}} \mapsto \hat{\mathfrak{e}}_i \in \mathbb{R}^3$. Then, if $f : \mathbb{R}^3 \to \mathbb{R}$ is a smooth function and $\mathbf{V} : \mathbb{R}^3 \to \mathbb{R}^3$ is a smooth vector field, we can express div, grad, and curl as:

$$\boxed{\begin{aligned} \nabla f &= \mathbf{F}((df)^\sharp) \\ \mathrm{curl}(\mathbf{V}) &= \mathbf{F}((*d(\mathbf{V}^\flat))^\sharp) \\ \mathrm{div}(\mathbf{V}) &= *d*(\mathbf{V}^\flat) \end{aligned}}$$

Proof.

- From Theorem 7.34, we have $df_{\mathbf{p}} = \sum_{i=1}^{3} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} dx^i|_{\mathbf{p}}$, so $(df)^{\sharp} = \sum_{i=1}^{3} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$. Then $\mathbf{F}((df)^{\sharp}) = \sum_{i=1}^{3} \frac{\partial f}{\partial x^i}\Big|_{\mathbf{p}} \hat{\mathfrak{e}}_i = (\nabla_{\mathbf{x}} f)|_{\mathbf{p}}$.
- Here's an outline of what happens in the computation. $\mathbf{V}^{\flat}$ is a differential 1-form corresponding to the vector field $\mathbf{V}$. Then $d(\mathbf{V}^{\flat})$ is a differential 2-form. The 2-wedges in the linear combination for $d(\mathbf{V}^{\flat})$ get sent to the "1-wedges" $dx^1|_{\mathbf{p}}, dx^2|_{\mathbf{p}}, dx^3|_{\mathbf{p}}$ (which are really "no wedges") by $*$. Then each $dx^i|_{\mathbf{p}}$ is sent to $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ by $\sharp$. Lastly, each $\frac{\partial}{\partial x^i}\Big|_{\mathbf{p}}$ is sent to $\hat{\mathfrak{e}}_i$ by $\mathbf{F}$. We are left with a vector field in $\mathbb{R}^3$ that is the curl of the vector field $\mathbf{V}$.
- For a smooth vector field $\mathbf{V} : \mathbb{R}^2 \to \mathbb{R}^2$, we have $\mathrm{div}(\mathbf{V}) = \mathrm{curl}(\mathbf{R}_{\frac{\pi}{2}}(\mathbf{V}))$, where $\mathbf{R}_{\frac{\pi}{2}}$ is counter-clockwise rotation by $\frac{\pi}{2}$. This is because if $\mathbf{V} = (V_1, V_2)$, then $\mathbf{R}_{\frac{\pi}{2}}(\mathbf{V}) = (-V_2, V_1)$. Thus, we can apply the second line of the boxed equation to prove the theorem when $\mathbf{V}$ is a 2-dimensional vector field.

□

Theorem 8.36. [Lee, p. 368] (Vector calculus commutative diagram).

The following commutative diagram summarizes how div, grad, and curl correspond to the exterior derivative action on differential 0-, 1-, 2-, and 3- forms on $\mathbb{R}^3$.

$$\begin{tikzcd}
C^{\infty}(\mathbb{R}^3 \to \mathbb{R}) \arrow[r, "\mathrm{grad}"] \arrow[d, "\mathbf{I}"'] & \text{vector fields on } \mathbb{R}^3 \arrow[r, "\mathrm{curl}"] & \text{vector fields on } \mathbb{R}^3 \arrow[r, "\mathrm{div}"] & C^{\infty}(\mathbb{R}^3 \to \mathbb{R}) \arrow[d, "*"] \\
\Omega^0(\mathbb{R}^3) \arrow[r, "d"'] & \Omega^1(\mathbb{R}^3) \arrow[r, "d"'] & \Omega^2(\mathbb{R}^3) \arrow[r, "d"'] & \Omega^3(\mathbb{R}^3)
\end{tikzcd}$$

Note that the fact $d \circ d = \mathbf{0}$ implies $\mathrm{curl} \circ \mathrm{grad} = \mathbf{0}$ and $\mathrm{div} \circ \mathrm{curl} = \mathbf{0}$.

8.4 The generalized Stokes' theorem

Having set up integration on manifolds and having defined the exterior derivative, we are almost ready to present the generalized Stokes' theorem. To prove the generalized Stokes' theorem, we will use a tool called a *partition of unity.* A partition of unity is essentially the "theoretically nice" way to break up the support of a compactly supported differential form. Instead of breaking up the support into distinct "chunks", partitions of unity can be thought of as "smoothly fading the differential form in and out".

Definition 8.37. [Lee, p. 43] (Partitions of unity).

If $\{U_\alpha\}$ is an open cover of M and $\{f_\alpha\}$ is a set of functions $M \to \mathbb{R}$, we say that $\{f_\alpha\}$ is a *partition of unity of M (subordinate to $\{U_\alpha\}$)* iff

- $\text{supp}(f_\alpha) \subseteq U_\alpha$ for all α
- $f_\alpha(\mathbf{p})$ is nonzero for only finitely many α, for each $\mathbf{p} \in M$
- $(\sum_\alpha f_\alpha(\mathbf{p}) = 1 \text{ for all } \mathbf{p} \in M) \iff (\sum_\alpha f \text{ is the multiplicative identity of } \{\text{functions } M \to \mathbb{R}\})$.

A *smooth partition of unity* is a partition of unity in which each function f_α is smooth.

Every open cover of a smooth manifold WWBOC admits a smooth partition of unity ([Lee, p. 43]). Since every manifold has an open cover, this means that every smooth manifold WWBOC admits a smooth partition of unity.

Theorem 8.38. [Lee, p. 43] (Existence of smooth partitions of unity).

We take for granted the fact from [Lee] that any open cover of any smooth manifold WWBOC has a smooth partition of unity.

Theorem 8.39. (Integral of a differential n-form on a smooth n-manifold via partition of unity).

Let M be an oriented smooth n-manifold WWBOC and let ω be a compactly supported n-form on M. Since $\text{supp}(\omega)$ is compact, there is a finite collection $\{(U_i, \mathbf{x}_i)\}_{i=1}^k$ of charts on M for which $\{U_i\}_{i=1}^k$ is an open cover of $\text{supp}(\omega)$, with the pairwise intersections having measure zero. Additionally, let $\{f_i\}_{i=1}^k$ be a smooth partition of unity subordinate to $\{U_i\}_{i=1}^k$.

Simply repeating the previous definition, we have,

$$\int_M \omega = \sum_{i=1}^k \int_{U_i} \omega.$$

Now, since each $f_i\omega$ is supported in U_i, we have

$$\int_{U_i} \omega = \int_{U_i} f_i\omega = \int_M f_i\omega$$

The last equality of the line above applies the definition of the integral over a manifold of a differential form that is compactly supported in a single chart (see Definition 8.22).

Summing over i, we obtain

$$\sum_{i=1}^k \int_{U_i} \omega = \sum_{i=1}^k \int_M f_i\omega.$$

Therefore the integral of a differential form over a manifold can be computed with use of the partition of unity:

$$\boxed{\int_M \omega = \sum_{i=1}^k \int_M f_i\omega}$$

Remark 8.40. (Integration with partitions of unity).

Integration of differential forms on manifolds is most often *defined* in terms of partitions of unity. We prefer to view the partition of unity method as a consequence of the definition $\int_M \omega := \sum_{i=1}^k \int_{U_i} \omega$ (see Definition 8.26), as this definition is an intuitive starting point.

Theorem 8.41. [Lee, p. 407] (Properties of integrals of differential forms).

Theorem 8.42. [HH, p. 561] (The generalized Stokes' theorem on a single smooth chart).

Let M be a smooth n-manifold and consider a smooth oriented chart $(U, \mathbf{x})$ on M. Let ω be a smooth differential $(k-1)$-form that is compactly supported in U. Then

$$\int_M d\omega = \int_{\partial M} \omega.$$

Intuitively, this theorem is true because

$$\int_M d\omega \approx \sum_i d\omega(P_i) \approx \sum_i \int_{\partial P_i} \omega \approx \int_{\partial M} \omega.$$

The first approximation holds because integrals are limits of Riemann sums. The second approximation holds because the definition of the exterior derivative implies $d\omega(C) \approx \int_{\partial C} \omega$, where C is one of the k-cubes in the previous Riemann sum (a k-cube is a k-parallelapiped, and differential forms evaluate k-parallelapipeds). The third approximation holds because each boundary ∂P_i in the sum of integrals corresponds to exactly one oppositely oriented boundary $\partial P_i = -\partial P_j$ that occupies the same space: we have $\sum_{C \in D_N(\mathrm{cl}(\mathbb{H}^k))} \int_{\partial C} \omega = \sum_{C' \in D_N(\partial \mathrm{cl}(\mathbb{H}^k)} \int_{C'} \omega$, which is equal to $\int_{\partial \mathrm{cl}(\mathbb{H}^k)} \omega$ by Definition 8.26.

Proof. For simplicity, assume $M = \mathbb{R}^n$ or $\mathbb{H}^n$.

Each "approximation" above translates loosely translates to a statement of the form $f(M, \omega, N_1) \approx g(M, \omega, N_2)$. A statement such as this further translates to the formal statement
(for all M, ω, N_1 and for all $\epsilon > 0$, there exists an $N_2 \in \mathbb{N}$ for which $|f(M, \omega, N_1) - g(M, \omega, N_2)| < \epsilon$). It suffices to prove each formal statement individually, because "approximation" when treated this way is associative.

($\int_U d\omega \approx \sum_i d\omega(P_i)$). Let $\epsilon > 0$. Then, by the definition of the integral via Riemann sums, there exists an N large enough such that the dyadic paving of U of fineness 2^{-N} ensures $|\int_U d\omega - \sum_{C_i \in D_N(\mathrm{cl}(\mathbb{H}^k))} d\omega(P_i)| < \epsilon$.

($\sum_i d\omega(P_i) \approx \sum_i \int_{\partial P_i}$). Let $\epsilon > 0$ (forget about the previous ϵ). Take the dyadic decomposition[3] of U of fineness 2^{-N}, so that U is a countable disjoint union of k-cubes with side length 2^{-N}. Each k-cube is a k-parallelapiped of the form $C_i = P_{\mathbf{p}}(h\hat{\mathfrak{e}}_1, ..., h\hat{\mathfrak{e}}_k)$ for some $\mathbf{x}$, and where $h = 2^{-N}$. By the definition of the exterior derivative (Definition 8.28), there exist $K, \delta > 0$ such that when $|h| = 2^{-N} < \delta$, we have $|\int_{C_i} d\omega - \sum_i d\omega(C_i)| < Kh^{k+1}$. By taking N sufficiently large, $|h|$ becomes sufficiently small, and we get $|\int_{C_i} d\omega - \sum_i d\omega(C_i)| < Kh^{k+1} < \epsilon$.

($\sum_i \int_{\partial P_i} \omega = \int_{\partial U} \omega$). We prove this last step with with direct equality rather than by using a converging approximation. Written out more formally, $\sum_i \int_{\partial P_i} \omega$ is $\sum_{C \in D_N(\mathrm{cl}(\mathbb{H}^k))} \int_{\partial C} \omega$. Applying Theorem 7.49, which gives the oriented boundary of a k-parallelapiped, all the internal boundaries in the sum $\sum_{C \in D_N(\mathrm{cl}(\mathbb{H}^k))} \int_{\partial C}$ cancel, since each boundary appears twice with opposite orientations. Thus $\sum_{C \in D_N(\mathrm{cl}(\mathbb{H}^k))} \int_{\partial C} \omega = \sum_{C' \in D_N(\partial \mathrm{cl}(\mathbb{H}^k))} \int_{C'} \omega = \int_{\partial M} \omega$. □

Remark 8.43. The above theorem is the real "heart" of Stokes' theorem. The proof of the full-blown Stokes' theorem uses partitions of unity to extend the previous result to an entire manifold.

[3] Intuitively, the dyadic decomposition of a subset of $\mathbb{R}^k$ of 2^{-N} is a partition of that subset obtained by halving the subset N times. See [HH, p. 356].

Theorem 8.44. [Lee, p. 419] (The generalized Stokes' theorem).

Let M be an oriented smooth manifold with corners, and let ω be a compactly supported smooth differential $(n-1)$-form on M. Then

$$\int_M d\omega = \int_{\partial M} \omega.$$

Proof. The proof of this theorem is adapted from [HH, p. 661 - 665].

Using the theorem which expresses $\int_M \omega$ in terms of a partition of unity, and using the linearity of the exterior derivative, we have

$$\int_M d\omega = \int_M \sum_{i=1}^k f_i d\omega = \int_M \sum_{i=1}^k d(\underbrace{f_i\omega}_{\omega_i}) = \int_M \sum_{i=1}^k d\omega_i.$$

Here, we have[4] defined $\omega_i := f_i\omega$.

Using the linearity of the integral, we have

$$\int_M \sum_{i=1}^k d\omega_i = \sum_{i=1}^k \int_M d\omega_i.$$

Recalling the definition of an integral of a differential form over a manifold the definition of the integral of a differential form over a manifold that is compactly supported in a single chart (Definitions 8.22 and 8.26, respectively) this becomes

$$\sum_{i=1}^k \int_M d\omega_i = \sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(d\omega_i).$$

Because the differential commutes with the pullback of a differential (Theorem 8.30), we have

$$\sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(d\omega_i) = \sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} d\Big((\mathbf{x}_i^{-1})^*(\omega_i)\Big).$$

We apply the previous theorem- Stokes' theorem on a single smooth chart- to the integral inside the sum to obtain

$$\sum_{i=1}^k \int_{\mathbf{x}_i(U_i)} d\Big((\mathbf{x}_i^{-1})^*(\omega_i)\Big) = \sum_{i=1}^k \int_{\partial\mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(\omega_i).$$

We use that $\partial\mathbf{x}_i(U_i) = \mathbf{x}_i(\partial U_i)$ and apply the definition of a differential form over a manifold that is compactly supported in a single chart, again, to get

$$\sum_{i=1}^k \int_{\partial\mathbf{x}_i(U_i)} (\mathbf{x}_i^{-1})^*(\omega_i) = \sum_{i=1}^k \int_{\mathbf{x}_i(\partial U_i)} (\mathbf{x}_i^{-1})^*(\omega_i) = \sum_{i=1}^k \int_{\partial U_i} \omega_i.$$

Finally, since $\{U_i\}_{i=1}^k$ is a finite open cover for $\text{supp}(\omega) \subseteq M$, where the pairwise intersections have measure zero, then $\{\partial U_i\}_{i=1}^k$ is a finite open cover for $\partial\text{supp}(\omega) \subseteq \partial M$, where the pairwise intersections have measure zero. Thus, this last expression is

[4]This notation reflects that, in some sense, using a partition of unity on a topological manifold is similar to choosing a basis for a vector space. If we *really* wanted to emphasize this fact, we might have denoted $f_i\omega$ by $([\omega]_{\{f_i\}_{i=1}^k})_i$.

$$\sum_{i=1}^{k} \int_{\partial U_i} \omega_i = \sum_{i=1}^{k} \int_{\partial M} f_i \omega = \int_{\partial M} \omega.$$

The last equality follows by Theorem 8.39. $\square$

Remark 8.45. ("∂ is the adjoint of d").

If we define a "bilinear function" on differential forms ω and smooth manifolds M by "$\langle \omega, M \rangle := \int_M \omega$", then Stokes' theorem is stated as

$$\text{"}\langle d\omega, M \rangle = \langle \omega, \partial M \rangle\text{"}.$$

Recalling Definition 3.4, it *appears* that the boundary operator ∂ is the "adjoint" of the exterior derivative d. (Quotation marks have been used excessively in this remark because the statements of this remark are only heuristic and not mathematically precise).

Part IV

Computational applications of differential forms

9

Discrete differential geometry and computer graphics

This chapter is divided into three parts. First, we introduce basic notions of *discrete* differential geometry, which correspond to the ideas of "regular" differential geometry we seen. The next part relies on the first part, and shows how to "smooth out a surface" in a discrete setting by using the Laplacian operator Δ. The third part describes how to take advantage of the complex numbers to set up a minimization problem, that when solved, produces a chart mapping into $\mathbb{C}$ that preserves its manifold's notion of angle as best as possible.

9.1 Discrete differental geometry

In this section, we follow Section 4.8 of [Cra19], and translate concepts from "regular" differential geometry over to their discrete counterparts. To begin our translation of concepts, we recall that the "regular" differential forms of the previous chapter lived on smooth manifolds WWBOC. *Discrete differential forms* will live on *simplicial complexes*, or *meshes*, that are built out of *simplices*.

Definition 9.1. (Oriented k-simplex in $\mathbb{R}^n$).

An *oriented k-simplex $k \leq n$, in* $\mathbb{R}^n$ is a tuple (K, Or), where $K \subseteq \mathbb{R}^n$ is a subset of the form

$$\Big\{ \sum_{i=1}^{k+1} c_i \mathbf{q}_i \mid \sum_{i=1}^{k+1} c_i = 1, c_i \geq 0 \Big\},$$

and where Or is the standard orientation of $\{\mathbf{p}_1, ..., \mathbf{p}_{k+1}\}$ inherited from $\mathbb{R}^k$.

When treated as a manifold embedded in $\mathbb{R}^n$, an oriented k-simplex can be given a smooth structure so that it is a smooth oriented manifold with corners.

Example 9.2. (0-, 1-, and 2- simplices).

A 0-simplex is a vertex (a point with positive orientation), a 1-simplex is an edge (an oriented line segment), a 2-simplex is an oriented triangle, and a 3-simplex is an oriented tetrahedron.

Definition 9.3. (Oriented simplicial k-complex in $\mathbb{R}^n$).

An *oriented simplicial k-complex in* $\mathbb{R}^n$ is a union of oriented k-simplices in $\mathbb{R}^n$, where the pairwise intersections have measure zero, such that any two faces that fill the same space are oppositely oriented.

We often refer to an oriented simplicial k-complex in $\mathbb{R}^n$ as a *k-mesh*, or simply as a *mesh*.

Now that the setting for discrete differential forms has been established, we move to define discrete differential forms themselves. [MDT] notes that "finding a discrete counterpart to the notion of differential forms is a delicate matter. If one was to represent differential forms using their coordinate values and approximate the exterior derivative using finite differences, basic theorems such as Stokes' theorem would not hold numerically." Recognizing this issue, we now present a definition of discrete differential k-form that *does* work.

Definition 9.4. (Discrete differential k-form).

Let K be a k-mesh, considered as a smooth k-manifold with corners embedded in $\mathbb{R}^n$, and let ω be a differential k-form on K. We define the *discrete differential k-form $\widehat{\omega}$ on K corresponding to ω* to be the function which integrates ω over an arbitrary collection of k-simplices in K. That is, if $\{K_i\}$ is a finite collection of k-simplices in K, then the *evaluation $\widehat{\omega}_{\{K_i\}}$ of $\widehat{\omega}$ on* $\{K_i\}$ is defined to be

$$\widehat{\omega}_{\{K_i\}} := \int_{\{K_i\}} \omega = \sum_i \int_{K_i} \omega.$$

Here, the hat $\widehat{\ }$ is meant to denote "discrete", *not* "normalized".

Definition 9.5. (Convention for discrete 0-forms).

By the previous definition, a discrete 0-form is a 0-form (i.e. a function) that's been integrated over every 0-simplex in a 1-simplex. The integral of a function "over" a point is zero, but we will use the convention of treating the value of a discrete 0-form $\widehat{f}|_{\mathbf{p}_i}$ at a point to be the value of the corresponding function at that point, $\widehat{f}|_{\mathbf{p}_i} := f|_{\mathbf{p}_i}$.

Now that we have defined discrete differential forms, we expect a corresponding definition of discrete exterior derivative.

Definition 9.6. (Discrete exterior derivative).

Let K be a k-mesh, and consider a discrete differential ℓ-form $\widehat{\omega}$, $\ell < k$, on K. The *discrete exterior derivative* $\widehat{d}$ acts on $\widehat{\omega}$ to produce a discrete differential $(\ell+1)$-form $\widehat{d}\widehat{\omega}$ on K. The value of $\widehat{d}\widehat{\omega}$ at an $(\ell+1)$-simplex L of K is given by

$$(\widehat{d}\widehat{\omega})_L := \int_L d\omega.$$

The following theorem reveals why we chose to define the discrete exterior derivative as we did.

Theorem 9.7. (Stokes' theorem for discrete differential forms).

If the $(\ell+1)$-simplex L of the previous definition is the finite union of $\{L_i\}$, where the pairwise intersections of the $\{L_i\}$ have measure zero, then we have

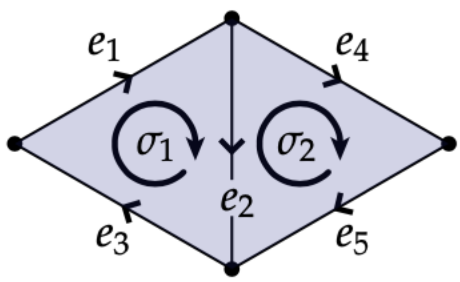
$$(\widehat{d}\widehat{\omega})_L = \int_L d\omega = \int_{\partial L} \omega = \sum_i \int_{(\partial L)_i} \omega = \sum_i \widehat{\omega}_{(\partial L)_i}.$$

(We have assumed existence of a finite collection $\{(\partial L)_i\}$, where the pairwise intersections of $\{(\partial L)_i\}$ have measure zero, whose union is ∂L).

Thus, we have recovered a discrete version of Stokes' theorem:

$$\boxed{(\widehat{d}\widehat{\omega})_L = \sum_i \widehat{\omega}_{(\partial L)_i}}$$

Example 9.8. (Computing the discrete exterior derivative).



In the above figure, take $\mathbf{p}_1$ to be the leftmost vertex and $\mathbf{p}_2$ to be the topmost vertex. We have

$$\begin{aligned}(\widehat{d\omega})_{\mathbf{p}_1} &= \widehat{\omega}_1 + \widehat{\omega}_2 + \widehat{\omega}_3 \\ (\widehat{d\omega})_{\mathbf{p}_2} &= \widehat{\omega}_4 + \widehat{\omega}_5 - \widehat{\omega}_1.\end{aligned}$$

Before introducing the notion of *mesh duality*, we need to know what a "k-cell in $\mathbb{R}^n$" is.

Definition 9.9. (k-cell in $\mathbb{R}^n$).
A *k-cell in* $\mathbb{R}^n$ is a subset of $\mathbb{R}^n$ that is a Cartesian product of k closed intervals.

Definition 9.10. [MDT] (Mesh duality).
Consider a mesh L (recall Definition 9.3). We will consider L to be the "original", or *primal mesh*. The *dual mesh* to the primal mesh that is L is obtained by replacing each k-simplex in L with an $(n-k)$-cell. As stated in [MDT], the dual mesh "inherits several properties and operations" from the primal mesh. "Most important is the notion of incidence. For instance, if two primal edges are on the same primal face, then the corresponding dual faces are incident, that is, they share a common dual edge (which is the dual of the primal common face)".

Since a dual mesh is constructed out of $(n-k)$-cells rather than $(n-k)$-simplices, dual meshes are not necessarily simplicial complexes, but are instead examples of what are called *cell complexes*.

Definition 9.11. [Sol13] (Dual mesh conventions).
The previous definition gives topological rules for constructing the dual complex, but it does not specify where the dual cells are located in space. As is done in [Sol13], "we'll choose the dual vertices to be located at the circumcenter of their primal face counterparts (circumcenter is the center of the circle that passes through all three triangle vertices)." Doing this ensures that a dual edge is perpendicular to its corresponding primal edge.

As it deals with relating k-dimensional space to $(n-k)$-dimensional space, the above definition is reminiscent of Hodge-duality. The Hodge-dual is indeed what will establish a correspondence between so-called *primal discrete differential forms* and *dual discrete differential forms*.

Definition 9.12. [Sol13] (Discrete Hodge-duals)
Let K be a 2-mesh embedded in $\mathbb{C}$. Denote the 2-simplices (faces) of K by F_i, the 1-simplices (edges) of K by e_i, and the 0-simplices (vertices) of K by $\mathbf{p}_i$. We define the *discrete Hodge-dual* $\widehat{*}$ on discrete primal differential forms in a piecewise manner:

$$\begin{aligned}(\widehat{*}\widehat{\omega})_{F_i} &= \frac{1}{|F_i|}\widehat{\omega}, && \widehat{\omega} \text{ is a discrete primal differential 2-form} \\ (\widehat{*}\widehat{\omega})_{e_i} &= \frac{|e_i^*|}{|e_i|}\widehat{\omega}, && \widehat{\omega} \text{ is a discrete primal differential 1-form} \\ (\widehat{*}\widehat{f})_{\mathbf{p}_i} &= |\mathbf{p}_i^*|\widehat{\omega}, && \widehat{f} \text{ is a discrete primal differential 0-form (i.e. a function)}\end{aligned}$$

In the above, $|\cdot|$ takes the volume of its argument (so, in the above, it only ever returns 2-dimensional area or the length of an edge), e_i^* is the dual edge to e_i (so e_i^* is really a 1-cell in the dual mesh), and $\mathbf{p}_i^*$ is the dual vertex to $\mathbf{p}_i$ (so $\mathbf{p}_i^*$ is really a 2-cell in the dual mesh).

The middlemost definition of $\widehat{*}$ (which acts on a discrete primal differential 1-form) is often referred to as the *diagonal Hodge-dual*.

9.2 Smoothing with the Laplacian

In this section, our goal is to discretize the PDE

$$\frac{\partial \mathbf{F}}{\partial t} = \Delta \mathbf{F},$$

where M is a smooth 2-manifold embedded in $\mathbb{R}^3$, $\mathbf{F} : M \to \mathbb{R}^3$ is sufficently smooth, and where the *Laplacian* operator Δ on vector-valued functions is defined as

$$\Delta \mathbf{F} = \Delta \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} := \begin{pmatrix} \Delta F_1 \\ \Delta F_2 \\ \Delta F_3 \end{pmatrix}.$$

Here, the Δ on the right hand side of this definition is the Laplacian operator on scalar-valued functions, $\Delta = \text{div} \circ \text{grad}$.

It can be shown that $\Delta \mathbf{F} = 2H\hat{\mathbf{n}}$, where H is the mean curvature[1] and $\hat{\mathbf{n}}$ is the unit normal (see [Cra19, p. 114]). Thus $\mathbf{F}$ satisfies the above PDE iff, one can stay on the surface by going in the direction of the normal at a point, with strength proportional to the mean curvature at that point. More briefly, solving the above PDE is a formal way to "smooth out" the surface M parameterized by $\mathbf{F}$.

In order to solve the PDE $\frac{\partial \mathbf{F}}{\partial t} = \Delta \mathbf{F}$, we will set $\mathbf{G} := \Delta \mathbf{F}$ and consider the more abstract problem of

$$\Delta \mathbf{F} = \begin{pmatrix} \Delta F_1 \\ \Delta F_2 \\ \Delta F_3 \end{pmatrix} = \begin{pmatrix} G_1 \\ G_2 \\ G_3 \end{pmatrix} = \mathbf{G}.$$

This problem further reduces to the problem $\Delta F_i = G_i$, that is,

$$\Delta f = g,$$

where $f, g : M \to \mathbb{R}$ are sufficiently smooth real valued functions on M. So, our goal has turned into discretizing the Laplacian Δ on scalar-valued functions.

Discretizing the Laplacian

We present two approaches for turning the "regular" Laplacian operator Δ on scalar-valued fucntions into the discrete Laplacian operator $\widehat{\Delta}$ on scalar valued functions. The relatively standard first approach is to use the *finite element method*. The novel second approach utilizes discrete differential geometry.

Discretizing the Laplacian with the finite element method

Consider the PDE

$$\Delta f = g.$$

The basic of the idea of the *finite element method* is to consider "approximate solutions" to the PDE in question that are defined piecewise on finitely many regions, called *elements*. The vector space of functions that are defined piecewise on elements is called the *discrete approximation space*. The

[1] There are various ways to measure the curvature of a surface at a point. The mean curvature is one of these ways.

best *approximate solution* is the function in the discrete approximation space that has the smallest normed *residual*. The norm we use to measure "normed residuals" is induced by an inner product on the discrete approximation space.

For our particular problem, our functions are defined on a 2-mesh K embedded in $\mathbb{R}^3$ (recall Definition 9.3). We take the elements to be 2-simplices on the mesh. Thus, our discrete approximation space is the vector space of sufficiently smooth functions that are defined piecewise on 2-simplices.

For the purpose of establishing a norm, we will use the L^2 *inner product* $\langle \cdot, \cdot \rangle$ on functions in our discrete approximation space, defined by

$$\langle f, g \rangle = \int_K fg.$$

The residual of a function h in our discrete approximation space is $\Delta h - g$. (The residual is the zero function when h is a solution to the PDE). The normed residual of h is $||\Delta h - g||$, where the norm $|| \cdot ||$ is induced by the L^2 inner product.

In the finite element method, one also makes use of a basis for the discrete approximation space. We will use the basis $\Phi = \{\phi_{\mathbf{p}_i}\}_{i=1}^n$ of *hat functions*, where the hat function $\phi_{\mathbf{p}_i} : K \to \mathbb{R}$ is defined to be 1 at the vertex $\mathbf{p}_i$ and 0 everywhere else.

To discretize the Laplacian operator Δ, we determine the Laplacian of a function h in the discrete approximation space by using the basis of hat functions. We need the following lemma before doing so.

Lemma 9.13. (Green's first identity).

Let M be a smooth 2-manifold embedded in $\mathbb{R}^3$. If $f, g : M \to \mathbb{R}$ are sufficiently smooth, then

$$\langle \Delta f, g \rangle = -\langle \nabla f, \nabla g \rangle + \langle \nabla f \cdot \hat{\mathbf{n}}, g \rangle_\partial,$$

where $\langle \cdot, \cdot \rangle_\partial$ is the L^2 inner product on the boundary ∂K, and where we've made use of the L^2 inner product on vector fields that is defined by

$$\langle \mathbf{V}_1, \mathbf{V}_2 \rangle := \int \mathbf{V}_1 \cdot \mathbf{V}_2,$$

Here, $\mathbf{V}_1 \cdot \mathbf{V}_2$ denotes the function $\mathbf{p} \mapsto \mathbf{V}_1|_\mathbf{p} \cdot \mathbf{V}_2|_\mathbf{p}$.

Proof. In general, $d(f\omega) = df \wedge d\omega + fd\omega$. So $d(g * df) = dg \wedge *df + gd * df$. Integrate over M and apply Stokes' theorem to the left hand side to obtain

$$\int_{\partial M} g * df = \int_M dg \wedge *df + \int_M gd * df.$$

Another general fact is that $**\omega = -1^{k(n-k)}\omega$ (see [Sch15] for a proof of this), so $**(d*df) = d*df$ (use $k = 3$, $n = 3$). That is, $(d * d)f = *(*d * d)f = *\Delta f$. So the above becomes

$$\int_{\partial M} g * df = \int_M dg \wedge *df + \int_M g * \Delta f.$$

This is the same as

$$\int_{\partial M} g * df = \langle dg, df \rangle + \langle g, \Delta f \rangle.$$

Rearranging, we have

$$\langle \Delta f, g \rangle = -\langle df, dg \rangle + \int_{\partial M} g * df.$$

To complete the proof, observe that $\langle df, dg \rangle = \langle \nabla f, \nabla g \rangle$, and that after identifying $dx^{\sigma(i)} \wedge dx^{\sigma(j)}$ in $*df$ with $\operatorname{sgn}(\sigma)\hat{\mathfrak{e}}_{\sigma(\{1,2,3\}-\{i,j\})}$, $*df$ becomes $\nabla f \cdot \hat{\mathbf{n}}$. $\square$

We are now ready to take the first step in computing a discretization of the Laplacian with the finite element method.

Derivation 9.14. (Discrete Laplacian in terms of L^2 inner products of gradients).

Let K be a 2-mesh embedded in $\mathbb{R}^3$, and let K_j be the 2-simplices (the triangles) on K. Using the bilinearity of the inner product $\langle \cdot, \cdot \rangle$, we see the *discretization (via the finite element method)* $\widehat{\Delta}$ *of the Laplacian* Δ satisfies

$$\langle \widehat{\Delta} h, \phi_{\mathbf{p}_i} \rangle = \sum_j \langle \widehat{\Delta} h, \phi_{\mathbf{p}_i} \rangle_{K_j}.$$

Applying Green's first identity to the argument of the sum yields

$$\langle \widehat{\Delta} h, \phi_{\mathbf{p}_i} \rangle = -\sum_j \langle \widehat{\nabla} h, \widehat{\nabla} \phi_{\mathbf{p}_i} \rangle_{K_j} + \sum_j \langle \hat{\mathbf{n}} \cdot \widehat{\nabla} h, \phi_{\mathbf{p}_i} \rangle_{\partial K_j}.$$

If the mesh has no boundary, then the inner product on the boundary $\langle \cdot, \cdot \rangle_{\partial K_j}$ vanishes, and we're left with

$$\langle \widehat{\Delta} h, \phi_{\mathbf{p}_i} \rangle = -\sum_j \langle \widehat{\nabla} h, \widehat{\nabla} \phi_{\mathbf{p}_i} \rangle_{K_j} = -\langle \widehat{\nabla} h, \widehat{\nabla} \phi_{\mathbf{p}_i} \rangle.$$

Expanding h in the basis $\Phi = \{\phi_{\mathbf{p}_i}\}_{i=1}^n$ as $h = \sum_j ([h]_\Phi)_j \phi_{\mathbf{p}_j}$, using the linearity of $\widehat{\nabla}$, and using the bilinearity of $\langle \cdot, \cdot \rangle$, we have

$$\langle \widehat{\Delta} h, \phi_{\mathbf{p}_i} \rangle = \langle \widehat{\nabla} \Big(\sum_j ([h]_\Phi)_j \phi_{\mathbf{p}_j} \Big), \widehat{\nabla} \phi_{\mathbf{p}_i} \rangle = \sum_j ([h]_\Phi)_j \langle \widehat{\nabla} \phi_{\mathbf{p}_j}, \widehat{\nabla} \phi_{\mathbf{p}_i} \rangle.$$

So it remains to compute $\langle \widehat{\nabla} \phi_{\mathbf{p}_j}, \widehat{\nabla} \phi_{\mathbf{p}_i} \rangle$. This will be accomplished by the next several lemmas.

Lemma 9.15. (Cotan formula lemma 1).

The "aspect ratio" $\frac{w}{h}$ of a triangle is the sum of the cotangents of the interior angles at its base. That is,

$$\frac{w}{h} = \cot(\alpha) + \cot(\beta).$$

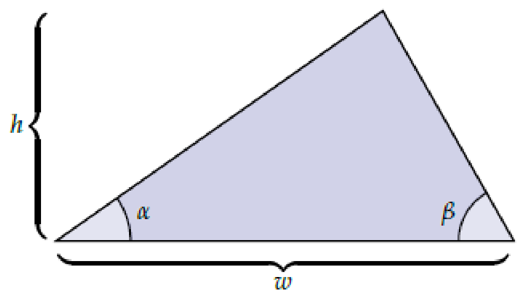


Figure 9.1: [Cra19, p. 105]

Proof. Drop an altitude from the top vertex of the triangle so that it intersects the base. Let x denote the length from this intersection point to the right vertex of the triangle. Then $\cot(\alpha) = \frac{w-x}{h}$ and $\cot(\beta) = \frac{x}{h}$, so $\cot(\alpha) + \cot(\beta) = \frac{w}{h}$. $\square$

Lemma 9.16. (Cotan formula lemma 2).

If $\mathbf{e}$ is the edge vector along the base of a triangle, and $\mathbf{e}_\perp$ is the result of rotating $\mathbf{e}$ counterclockwise by $\frac{\pi}{2}$, then the gradient of the hat function ϕ associated with the vertex opposite the side corresponding to $\mathbf{e}$ is

$$\widehat{\nabla}\phi = \frac{\mathbf{e}_\perp}{2A},$$

where A is the area of the triangle.

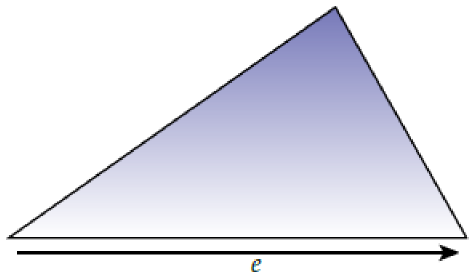


Figure 9.2: [Cra19, p. 106]

Proof. Denote the left, top, and right vertices of the triangle by $\mathbf{x}_i$, $i = 1, 2, 3$, respectively. Let $\phi = \phi_2$ denote the hat function associated with $\mathbf{x}_2$, the top vertex. We have $\phi(\mathbf{x}_1) = \phi(\mathbf{x}_3) = 0$ and $\phi(\mathbf{x}_2) = 1$.

Since ϕ is linear, it is of the form $\phi(\mathbf{x}) = \phi(\mathbf{p}) + (\widehat{\nabla}_\mathbf{x}\phi)|_\mathbf{p} \cdot (\mathbf{x} - \mathbf{p})$, for any $\mathbf{p} \in \mathbb{R}^3$. (Here, $\cdot$ is the dot product). It follows that the gradient of ϕ is the constant function $\widehat{\nabla}_\mathbf{x}\phi \equiv (\widehat{\nabla}_\mathbf{x}\phi)|_\mathbf{p}$. Using $\mathbf{p} = \mathbf{x}_2$, we have

$$\phi(\mathbf{x}) = 1 + (\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x} - \mathbf{x}_2).$$

Now use $\mathbf{x} = \mathbf{x}_1$ and $\mathbf{x} = \mathbf{x}_3$ to obtain

$$\begin{aligned} 0 &= 1 + (\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_1 - \mathbf{x}_2) \\ 0 &= 1 + (\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_3 - \mathbf{x}_2). \end{aligned}$$

These equations are equivalent to

$$\begin{aligned} 1 &= (\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_2 - \mathbf{x}_1) \\ 1 &= (\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_2 - \mathbf{x}_3). \end{aligned}$$

Subtract the second equation from the first and use the linearity of $\widehat{\nabla}_\mathbf{x}$ to obtain

$$(\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_3 - \mathbf{x}_1) = 0.$$

Thus, $(\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2}$ is perpendicular to $\mathbf{e} = \mathbf{x}_3 - \mathbf{x}_1$, which is the edge vector of the base of the depicted triangle. Since $(\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2}$ is the direction of greatest *increase*, not decrease, in ϕ, then $(\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2}$ must point *into* the triangle rather than out. (ϕ increases as we go "into" the triangle, since $\phi(\mathbf{x}_1) = \phi(\mathbf{x}_3) = 0 < \phi(\mathbf{x}_2) = 1$). This implies that the direction of $(\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2}$ is the same as the direction of the $\mathbf{e}_\perp$ described in the theorem.

To complete the proof, we will find the magnitude of $(\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2}$ and then show $\widehat{\nabla}_\mathbf{x}\phi \equiv (\widehat{\nabla}_\mathbf{x}\phi)|_{\mathbf{x}_2} = \frac{\mathbf{e}_\perp}{2A}$.

Consider the previous equation

$$1 = (\widehat{\nabla}_{\mathbf{x}}\phi)|_{\mathbf{x}_2} \cdot (\mathbf{x}_2 - \mathbf{x}_1).$$

The dot product on the right hand side can be re-expressed as

$$1 = ||(\widehat{\nabla}_{\mathbf{x}}\phi)|_{\mathbf{x}_2}||\ ||\mathbf{x}_2 - \mathbf{x}_1|| \cos\Big(\frac{\pi}{2} - \alpha\Big),$$

where α is the angle between $(\widehat{\nabla}_{\mathbf{x}}\phi)|_{\mathbf{x}_2}$ and $\mathbf{x}_2 - \mathbf{x}_1$. Equivalently, α is the angle between $\mathbf{e}$ and the left side of the triangle. Using this reinterpretation of α, we solve for $||(\widehat{\nabla}_{\mathbf{x}}\phi)|_{\mathbf{x}_2}||$ and get

$$||(\widehat{\nabla}_{\mathbf{x}}\phi)|_{\mathbf{x}_2}|| = \frac{1}{||\mathbf{x}_2 - \mathbf{x}_1|| \cos\Big(\frac{\pi}{2} - \alpha\Big)} = \frac{1}{||\mathbf{x}_2 - \mathbf{x}_1|| \sin(\alpha)} = \frac{1}{h},$$

where h is the height of the triangle.

Putting everything together, and using that $||\mathbf{e}_\perp|| = ||\mathbf{e}||$, we have

$$\widehat{\nabla}_{\mathbf{x}}\phi \equiv (\widehat{\nabla}_{\mathbf{x}}\phi)|_{\mathbf{x}_2} = ||(\widehat{\nabla}_{\mathbf{x}}\phi)|_{\mathbf{x}_2}||\hat{\mathbf{e}}_\perp = ||(\widehat{\nabla}_{\mathbf{x}}\phi)|_{\mathbf{x}_2}||\frac{\mathbf{e}_\perp}{||\mathbf{e}_\perp||} = \frac{1}{h||\mathbf{e}||}\mathbf{e}_\perp = \frac{1}{2(\frac{1}{2}h||\mathbf{e}||)}\mathbf{e}_\perp = \frac{\mathbf{e}_\perp}{2A}.$$

□

Lemma 9.17. (Cotan formula lemma 3).

The L^2 inner product of the gradient of a hat function ϕ corresponding to some vertex is

$$\langle \widehat{\nabla}\phi, \widehat{\nabla}\phi \rangle = \frac{1}{2}\Big(\cot(\alpha) + \cot(\beta)\Big),$$

where α and β are the interior angles at the other two vertices.

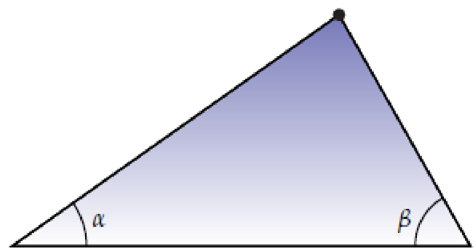


Figure 9.3: [Cra19, p. 106]

Proof. We have

$$\langle \widehat{\nabla}\phi, \widehat{\nabla}\phi \rangle = \int_{\text{triangle}} (\widehat{\nabla}\phi) \cdot (\widehat{\nabla}\phi) = \int_{\text{triangle}} ||\widehat{\nabla}\phi||^2 = A||\widehat{\nabla}\phi||^2,$$

where A is the area of the triangle. In the proof of the previous theorem, we showed $||\widehat{\nabla}\phi|| = \frac{1}{h}$, where h is the height of the triangle. So

$$\langle \widehat{\nabla}\phi, \widehat{\nabla}\phi \rangle = \frac{A}{h^2} = \frac{\frac{1}{2}||\mathbf{e}||h}{h^2} = \frac{1}{2}\frac{||\mathbf{e}||}{h}.$$

Apply Lemma 9.15, which states that the width-height ratio $\frac{||\mathbf{e}||}{h}$ is $\frac{||\mathbf{e}||}{h} = \cot(\alpha) + \cot(\beta)$, to obtain the result. □

Lemma 9.18. (Cotan formula lemma 4).

Consider the hat functions $\phi_{\mathbf{p}_i}$ and $\phi_{\mathbf{p}_j}$ corresponding to vertices $\mathbf{p}_i$ and $\mathbf{p}_j$ of the same triangle. We have

$$\langle \widehat{\nabla}\phi_{\mathbf{p}_i}, \widehat{\nabla}\phi_{\mathbf{p}_j} \rangle = -\frac{1}{2}\cot(\theta),$$

where θ is the angle between opposing edge vectors.

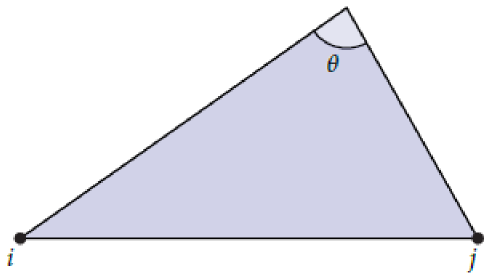


Figure 9.4: [Cra19, p. 106]

Proof. We have

$$\langle \widehat{\nabla}\phi_{\mathbf{p}_i}, \widehat{\nabla}\phi_{\mathbf{p}_j} \rangle = \int_{\text{triangle}} (\widehat{\nabla}\phi_{\mathbf{p}_i}) \cdot (\widehat{\nabla}\phi_{\mathbf{p}_i}) = A(\widehat{\nabla}\phi_{\mathbf{p}_i}) \cdot (\widehat{\nabla}\phi_{\mathbf{p}_i}),$$

where A is the area of the triangle. Applying Lemma 9.16, we have $\widehat{\nabla}\phi_{\mathbf{p}_i} = \frac{(\mathbf{e}_{\mathbf{p}_i})_\perp}{2A}$ and $\widehat{\nabla}\phi_{\mathbf{p}_j} = \frac{-(\mathbf{e}_{\mathbf{p}_j})_\perp}{2A}$, so

$$\langle \widehat{\nabla}\phi_{\mathbf{p}_i}, \widehat{\nabla}\phi_{\mathbf{p}_j} \rangle = A\frac{(\mathbf{e}_{\mathbf{p}_i})_\perp}{2A} \cdot \frac{-(\mathbf{e}_{\mathbf{p}_j})_\perp}{2A} = -\frac{\mathbf{e}_{\mathbf{p}_i} \cdot \mathbf{e}_{\mathbf{p}_j}}{4A} = -\frac{||\mathbf{e}_{\mathbf{p}_i}||||\mathbf{e}_{\mathbf{p}_j}||\cos(\theta)}{4A}.$$

To complete the proof, we do some trigonometry. Let h be the height of the triangle, let α be the angle at vertex $\mathbf{p}_i$ and let β be the angle at vertex $\mathbf{p}_j$. Then

$$\langle \widehat{\nabla}\phi_{\mathbf{p}_i}, \widehat{\nabla}\phi_{\mathbf{p}_j} \rangle = -\frac{||\mathbf{e}_{\mathbf{p}_i}||\;||\mathbf{e}_{\mathbf{p}_j}||\cos(\theta)}{4A} = -\frac{\frac{h}{\sin(\beta)}\frac{h}{\sin(\alpha)}\cos(\theta)}{4(\frac{1}{2}wh)} = -\frac{h\cos(\theta)}{2w\sin(\alpha)\sin(\beta)}$$

where w is the length of the base of the triangle. From Lemma 9.15, we know $w = h(\cot(\alpha) + \cot(\beta))$. Thus, the above becomes

$$\begin{aligned}\langle \widehat{\nabla}\phi_{\mathbf{p}_i}, \widehat{\nabla}\phi_{\mathbf{p}_j} \rangle &= -\frac{\cos(\theta)}{2(\cot(\alpha)+\cot(\beta))\sin(\alpha)\sin(\beta)} = -\frac{\cos(\theta)}{2\Big(\frac{\cos(\alpha)}{\sin(\alpha)} + \frac{\cos(\beta)}{\sin(\beta)}\Big)\sin(\alpha)\sin(\beta)} \\ &= -\frac{\cos(\theta)}{2(\cos(\alpha)\sin(\beta)+\cos(\beta)\sin(\alpha))} = -\frac{\cos(\theta)}{2\sin(\alpha+\beta)}.\end{aligned}$$

Since $\alpha + \beta = \theta$, we obtain the claimed result. □

Recall, the purpose of these lemmas was to determine $\langle \phi_{\mathbf{p}_i}, \phi_{\mathbf{p}_j} \rangle$. We needed to determine $\langle \phi_{\mathbf{p}_i}, \phi_{\mathbf{p}_j} \rangle$ so that we can finish the computation from the end of Derivation 9.14:

$$\langle \widehat{\Delta} h, \phi_{\mathbf{p}_i} \rangle = \sum_j ([h]_\Phi)_j \langle \widehat{\nabla} \phi_{\mathbf{p}_j}, \widehat{\nabla} \phi_{\mathbf{p}_i} \rangle.$$

Lemma 9.17 gives us a formula for $\langle \phi_{\mathbf{p}_i}, \phi_{\mathbf{p}_j} \rangle$ when $i = j$, and Lemma 9.18 gives us a formula for $\langle \phi_{\mathbf{p}_i}, \phi_{\mathbf{p}_j} \rangle$ when $i \neq j$. Plugging the results of these lemmas into the above sum, we obtain the following theorem.

Theorem 9.19. (Cotan formula via FEM).

$$(\widehat{\Delta} h)_{\mathbf{p}_i} = \frac{1}{2} \sum_j \Big(\cot(\alpha_j) - \cot(\beta_j) \Big) (h|_{\mathbf{p}_i} - h|_{\mathbf{p}_j})$$

Discretizing the Laplacian with discrete differential geometry

We now use discrete differential geometry to produce a nearly identitical discretization for the Laplacian (the $\widehat{\Delta}$ derived in this section will differ from the $\widehat{\Delta}$ derived with the finite element method by a scalar multiple).

Definition 9.20. (Discrete Laplacian via discrete differential geometry).

Let K be a 2-mesh embedded in $\mathbb{R}^3$. Since Theorem 8.35 tells us that $\text{div}(\mathbf{V}) = *d * \mathbf{V}^\flat$ and $\text{grad}(f) = (df)^\sharp$, then the "regular" Laplacian must act on f by $\Delta f = (*d * d)f$. Thus, the "regular" Laplacian is

$$\Delta = *d * d.$$

This "regular" Laplacian is easily discretized- to discretize, we just put hats on the $*$'s and d's! Formally, the *discrete Laplacian (via discrete differential geometry)* is defined to be

$$\widehat{\Delta} := \widehat{*} \ \ \widehat{d}_{\text{dual}} \ \ \widehat{*} \ \ \widehat{d}_{\text{primal}},$$

where $\widehat{d}_{\text{primal}}$ and $\widehat{d}_{\text{dual}}$ are the discrete exterior derivatives on the primal and dual meshes, respectively. Referring back to Definition 9.12, we see that the rightmost $\widehat{*}$ must be the diagonal Hodge-dual and that the leftmost $\widehat{*}$ must be the inverse of the third definition of the Hodge-dual from Definition 9.12.

Theorem 9.21. (Formula for the discrete Laplacian on a 2-mesh embedded in $\mathbb{R}^3$).

Let K be a 2-mesh embedded in $\mathbb{R}^3$, and consider a discrete differential 0-form $\widehat{f}$ on K. Then

$$\boxed{(\widehat{\Delta} \widehat{f})_{\mathbf{p}_i} = (\widehat{*} \ \widehat{d}_{\text{dual}} \ \widehat{*} \ \widehat{d}_{\text{primal}} \ \widehat{f})_{\mathbf{p}_i} = \frac{1}{|\mathbf{p}_i^*|} \sum_j \frac{|e_{ij}^*|}{|e_{ij}|} (f|_{\mathbf{p}_j} - f|_{\mathbf{p}_i})}$$

where $\mathbf{p}_i^*$ is the 2-cell in the dual mesh that is dual to the vertex $\mathbf{p}_i$ in the primal mesh K.

Proof. First, we evaluate $\widehat{d}\widehat{f}$ on an arbitrary edge extending from $\mathbf{p}_i$; let e_{ij} be the edge connecting vertices $\mathbf{p}_i$ and $\mathbf{p}_j$ that has the standard orientation of $\{\mathbf{p}_i, \mathbf{p}_j\}$ inherited from $\mathbb{R}^2$. To take the discrete exterior derivative $\widehat{d}_{\text{primal}}$, we integrate along the single edge e_{ij}:

$$(\widehat{d}_{\text{primal}} \widehat{f})_{e_{ij}} = \int_{e_{ij}} df = \int_{\partial e_{ij}} f = f|_{\mathbf{p}_j} - f|_{\mathbf{p}_i}.$$

Then, applying the diagonal Hodge-dual $\widehat{*}$, we have

$$(\widehat{*}\, \widehat{d}_{\mathrm{primal}} \widehat{f})_{e_{ij}} = \frac{|e^*_{ij}|}{|e_{ij}|}(f|_{\mathbf{p}_j} - f|_{\mathbf{p}_i}).$$

Here, e^*_{ij} is the 1-cell that is dual to the 1-simplex (the edge) e_{ij}.

Next, we apply $\widehat{d}_{\mathrm{dual}}$ to $\widehat{*}\, \widehat{d}_{\mathrm{primal}} \widehat{f}$. To do so, we only have to observe that $\widehat{d}_{\mathrm{dual}}(\widehat{*}\, \widehat{d}_{\mathrm{primal}} \widehat{f})$ will be evaluated at the 2-cell $\mathbf{p}^*_i$ in the dual mesh; Stokes' theorem for discrete differential forms (Theorem 9.7) takes care of the rest:

$$(\widehat{d}_{\mathrm{dual}}\ \widehat{*}\ \widehat{d}_{\mathrm{primal}}\ \widehat{f})_{\mathbf{p}^*_i} = \sum_j \frac{|e^*_{ij}|}{|e_{ij}|}(f|_{\mathbf{p}_j} - f|_{\mathbf{p}_i}).$$

The last Hodge-dual is the inverse of the third Hodge-dual presented in Definition 9.12. Thus, applying the last Hodge-dual divides the previous result by the area of $|\mathbf{p}^*_i|$, and we obtain the claimed result. □

We only need one more lemma to complete our discretization of the Laplacian with discrete differential geometry.

Lemma 9.22. (Cotan formula lemma).

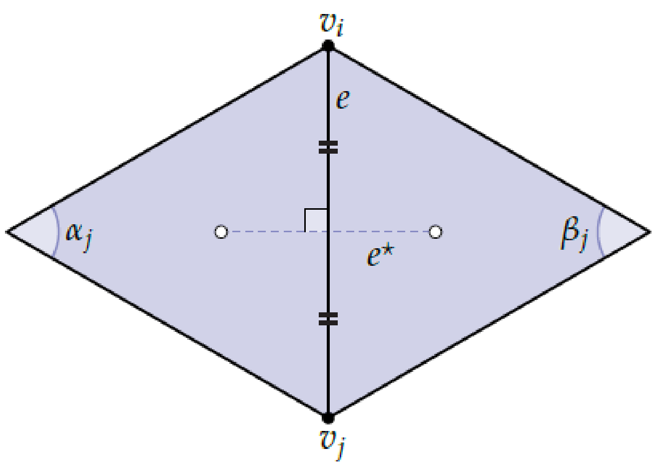


Figure 9.5: The diagram, which is from [Cra19, p. 110], uses e and e^* to denote e_{ij} and e^*_{ij}.

Since Lemma 9.15 shows that the width-height ratio $\frac{2|e^*_{ij}|}{|e_{ij}|}$ is $\frac{2|e^*_{ij}|}{|e_{ij}|} = \cot(\alpha_j) + \cot(\beta_j)$, then

$$\frac{|e^*_{ij}|}{|e_{ij}|} = \frac{1}{2}(\cot(\alpha_j) + \cot(\beta_j)).$$

Combining the previous lemma and theorem, we obtain the following theorem, which finalizes our differential geometric discretization of the Laplacian.

Theorem 9.23. (Cotan formula via discrete differential geometry).

$$\boxed{(\widehat{\Delta}\widehat{f})_{\mathbf{p}_i} = \frac{1}{2|\mathbf{p}^*_i|}\sum_j \Big(\cot(\alpha_j) + \cot(\beta_j)\Big)(\widehat{f}|_{\mathbf{p}_j} - \widehat{f}|_{\mathbf{p}_i})}$$

Solving the problem

We've seen that the discrete Laplacian can be expressed using the "cotan formula", and that this "cotan formula" can be derived using either the finite element method or with discrete differential geometry. How do we use this formula in a computer to smooth out surfaces in real time? The following theorem helps with this.

Theorem 9.24. (Cotangent of the angle between vectors).
Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$, and consider the angle θ between $\mathbf{v}_1$ and $\mathbf{v}_2$. We have

$$\cot(\theta) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_1 \times \mathbf{v}_2||}.$$

Proof. By definition of cot, $\cot(\theta) = \frac{||(\mathbf{v}_1)_{||}||}{||(\mathbf{v}_1)_{\perp}||}$. Using Lemma 1.89, we have $||(\mathbf{v}_1)_{||}|| = ||\text{proj}(\mathbf{v}_1 \to \mathbf{v}_2)|| = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_2||}$. Also, $||\mathbf{v}_1 \times \mathbf{v}_2|| = ||(\mathbf{v}_1)_{\perp}||||\mathbf{v}_2||$, so $||(\mathbf{v}_1)_{\perp}|| = \frac{||\mathbf{v}_1 \times \mathbf{v}_2||}{||(\mathbf{v}_1)_{\perp}||}$. Combining these two facts, we obtain $\cot(\theta) = \frac{(\mathbf{v}_1 \cdot \mathbf{v}_2)/||\mathbf{v}_2||}{||\mathbf{v}_1 \times \mathbf{v}_2||/||\mathbf{v}_2||}$, so $\cot(\theta) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{||\mathbf{v}_1 \times \mathbf{v}_2||}$. □

This theorem guarantees that if all lengths of all edges in our mesh are all the same, then the cotangent of the angle between any two edge vectors is a bilinear function. More specifically, since the cotangent produces a scalar, the cotangent of the angle between edges is a bilinear form in this situation. We learned in Chapter 3 that a bilinear form can be identified with a $\binom{1}{1}$ tensor. Thus, provided that the edge lengths are all the same, we can produce a matrix which describes the action of the discrete Laplacian at each vertex.

9.3 Surface parameterization

In this section, we explore how to wrap a 2-dimensional map onto a 3-dimensional surface (such as a globe) in an angle preserving manner[2]. Actually, instead of wrapping a 2-dimensional map onto a 3-dimensional surface, we will think about *unwrapping* a 3-dimensional surface into a flat 2-dimensional map. That is, we will be considering smooth charts (recall Definition 7.7) on a 3-dimensional manifold that map into a 2-dimensional space.

Since we are dealing with the problem of angle-preservation, we will need to deal with rotations in the "flat" unwrapped domain. One easy way to deal with such 2-dimensional rotations is to use the complex numbers $\mathbb{C}$, since multiplying by $i = \sqrt{-1}$ corresponds to performing a counterclockwise rotation by $\frac{\pi}{2}$ in the complex plane. Thus, given a smooth 2-manifold M embedded in $\mathbb{C}$, we will look for a global smooth chart $z : M \to \mathbb{C}$ that is angle preserving.

Before we proceed looking for such $z : M \to \mathbb{C}$, we need to define *complex differential forms.*

Definition 9.25. (Complex differential k-form).

Earlier, we defined a *real* differential k-form on a smooth manifold WWBOC M is a continuous function $\Lambda^k(T^*(M)) \to M$, where $\Lambda^k(T^*(M))$ is the subset of $T^0_k(T(M))$ consisting of antisymmetric tensors. These differential forms were *real* in the sense that $T^*_{\mathbf{p}}(M)$ was considered the double dual to the vector space $C^\infty(M \to \mathbb{R})$. (This consideration ripples upwards because $T^*(M) := \bigsqcup_{\mathbf{p} \in M} T^*_{\mathbf{p}}(M)$).

Let M be a smooth manifold WWBOC. A *complex* differential k-form on M is a continuous function $\Lambda^k(T^*(M)) \to M$, where $\Lambda^k(T^*(M))$ is the subset of $T^0_k(T(M))$ consisting of antisymmetric tensors, and where $T^*_{\mathbf{p}}(M)$ is considered as the double dual to the vector space $C^\infty(M \to \mathbb{C})$.

In practical terms, this definition implies the following analogue of Theorem 8.2. Let M be a smooth manifold WWBOC. Given any smooth chart $(U, \mathbf{x})$ on M, where x^i is the ith component function of $\mathbf{x}$, a complex differential k-form ω can be expressed as

$$\omega = \sum_{1 \leq i_1 < \ldots < i_k \leq n} (f_{i_1 \ldots i_k} + i g_{i_1 \ldots i_k}) dx^{i_1} \wedge \ldots \wedge dx^{i_k},$$

where each $f_{i_1 \ldots i_k}, g_{i_1 \ldots i_k} : U \to \mathbb{R}$.

Now that we have defined complex differential forms, we are able to speak of the differential of a complex-valued function. This is necessary for formally defining what it means for a chart on a smooth 2-manifold embedded in $\mathbb{C}$ to preserve angles.

Definition 9.26. (Angle-preserving complex chart).

Let M be a smooth 2-manifold embedded in $\mathbb{C}$. A chart $z : M \to \mathbb{C}$ is *angle-preserving* iff the following condition is satisfied:

$$\hat{\mathbf{n}} \times dz(v_{\mathbf{p}}) = i dz(v_{\mathbf{p}}) \text{ for all } v_{\mathbf{p}} \in T_{\mathbf{p}}(M).$$

The left side is the result of pushing forward a tangent vector with the differential of z and then rotating that tangent vector counterclockwise by $\frac{\pi}{2}$. The right side is the same thing, but the rotation has just been expressed using multiplication by i rather than with a cross product.

We now turn the above condition into one that is stated in terms of differential geometric concepts. This is accomplished by the following theorem, definition, and theorem.

Theorem 9.27. (Cauchy-Riemann equation).

Let M be as in the previous definition, and suppose $z : M \to \mathbb{C}$ is angle-preserving. Notice that instead of pushing $v_{\mathbf{p}}$ forward by dz and then rotating with the cross product, we could have rotated and *then* pushed forward $v_{\mathbf{p}}$ with dz. More formally, there must be some $J : T_{\mathbf{p}}(M) \to T_{\mathbf{p}}(M)$ for which $\hat{\mathbf{n}} \times d\mathbf{F}(v_{\mathbf{p}}) = d\mathbf{F}(J(v_{\mathbf{p}}))$. Thus, the definition of an angle preserving chart is equivalent to

[2] Our "wrapping" will not preserve lengths.

$$dz(J(v_{\mathbf{p}})) = idz(v_{\mathbf{p}}) \text{ for all } v_{\mathbf{p}} \in T_{\mathbf{p}}(M).$$

This equation is known as the *Cauchy-Riemann equation*.

To obtain a differential geometric statement of the Cauchy-Riemann equation, we need to define the *complex Hodge-dual*, which acts on complex differential forms.

Definition 9.28. (Complex Hodge-dual).
Let M be a smooth 2-manifold embedded in $\mathbb{C}$, and consider complex differential forms on M. We define the complex Hodge-dual $*$ to act on complex differential 1- and 2- forms, as follows. To specify the action of the complex Hodge-dual, we treat differential forms as objects that are, pointwise, actual functions. (Recall Derivation 8.5).
Let $v_{\mathbf{p}}$ be any tangent vector with[3] unit length.

$$\begin{aligned} * \, \omega(v_{\mathbf{p}}) &:= \omega(J(v_{\mathbf{p}})), \quad \omega \text{ is a complex differential 1-form} \\ * \, \omega &:= \omega(v_{\mathbf{p}}, J(v_{\mathbf{p}})), \quad \omega \text{ is a complex differential 2-form} \end{aligned}$$

As a sanity check, note that the complex Hodge-dual of a complex differential 1-form is a complex differential $(2-1=1)$-form, and the complex Hodge-dual of a complex differential 2-form is a complex differential $(2-2=0)$-form.

We can now present our final restatement of the angle-preservation condition.

Theorem 9.29. (Cauchy-Riemann with the Hodge-dual).
Let M be a smooth 2-manifold embedded in $\mathbb{C}$. The Cauchy-Riemann equation is

$$dz(J(v_{\mathbf{p}})) = idz(v_{\mathbf{p}}) \text{ for all } v_{\mathbf{p}} \in T_{\mathbf{p}}(M).$$

The left hand side can be written as a contraction: $dz(J(v_{\mathbf{p}})) = C(dz, J(v_{\mathbf{p}}))$. Using the previous definition of the complex Hodge-dual, we have $C(dz, J(v_{\mathbf{p}})) = C(*dz, v_{\mathbf{p}})$. Then, converting back to evaluation by the differential, we have $C(*dz, v_{\mathbf{p}}) = *dz(v_{\mathbf{p}})$. Thus, the above is
so the above becomes

$$*dz(v_{\mathbf{p}}) = idz(v_{\mathbf{p}}) \text{ for all } v_{\mathbf{p}} \in T_{\mathbf{p}}(M).$$

So the Cauchy-Riemann equation is equivalent to

$$\boxed{*dz = idz}$$

Having successfully translated the condition for angle-preservation into differential geometric terms, we will now find charts $z : M \to \mathbb{C}$ that "almost" solve the Cauchy-Riemann equation; we find charts that are as close to angle-preserving as possible.

Just as we did in our brief encounter with the finite element method, we will minimize the normed residual $|| * dz - idz||$ of a chart $z : M \to \mathbb{C}$ to produce the chart which fails to solve the Cauchy-Riemann equation with by the smallest margin. (This time, we will not assume that the minimum such "margin" is zero, however).

To do this, we need to define an inner product, so that we can use its induced norm $|| \cdot ||$. Since we are working with complex numbers, we will actually be defining a *Hermitian inner product*, which is similar to but not exactly the same as a "regular" inner product.

[3]Since the smooth 2-manifolds we are considering are embedded in $\mathbb{C}$, their tangent spaces at any point can be identified with $\mathbb{C}$. The length of a tangent vector is then computed using the norm induced by the analogue of the dot product for $\mathbb{C}$. See example 9.31.

Definition 9.30. (Hermitian inner product).

Let V be a vector space over $\mathbb{C}$. A *Hermitian inner product* on V is a binary function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ satisfying the following properties:

1. $\langle \cdot, \cdot \rangle$ respects addition. That is...
   (a) $\langle \mathbf{v}_1 + \mathbf{v}_2, \mathbf{v} \rangle = \langle \mathbf{v}_1, \mathbf{v} \rangle + \langle \mathbf{v}_2, \mathbf{v} \rangle$ for all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$.
   (b) $\langle \mathbf{v}, \mathbf{v}_1 + \mathbf{v}_2 \rangle = \langle \mathbf{v}, \mathbf{v}_1 \rangle + \langle \mathbf{v}, \mathbf{v}_2 \rangle$ for all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$.
2. $\langle c\mathbf{v}_1, \mathbf{v}_2 \rangle = c \langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $c \in \mathbb{C}$.
3. $\langle \mathbf{v}_1, c\mathbf{v}_2 \rangle = \overline{c} \langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $c \in \mathbb{C}$.
4. $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \overline{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.
5. $\langle \cdot, \cdot \rangle$ is positive-definite. That is, for all $\mathbf{v} \in V$, $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, and $\langle \mathbf{v}, \mathbf{v} \rangle = \mathbf{0}$ iff $\mathbf{v} = \mathbf{0}$.

Example 9.31. The traditional example of a Hermitian inner product on an n-dimensional vector space V over $\mathbb{C}$ is following "complex analogue" of the dot product: $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle := \sum_{i=1}^{n} ([\mathbf{v}_1])^i \overline{([\mathbf{v}_2])^i}$.

We now define the Hermitian product on complex differential 1-forms that we will use to measure the normed residual.

Definition 9.32. (Hodge Hermitian inner product on complex differential 1-forms).

Let M be a smooth 2-manifold embedded in $\mathbb{C}$. The *Hodge Hermitian inner product* $\langle\langle \cdot, \cdot \rangle\rangle$ on complex differential 1-forms is the Hermitian inner product defined as

$$\langle\langle \omega, \eta \rangle\rangle := \text{Re} \int_M *\overline{\omega} \wedge \eta.$$

Proof. We will show that $\langle\langle \cdot, \cdot \rangle\rangle$ is positive-definite. The other conditions are simple to check.

Consider $*\overline{\omega} \widetilde{\wedge} \omega(v_{\mathbf{p}}, J(v_{\mathbf{p}}))$. (Recall Derivation 8.5 for what it means to interpret a differential form as an object that is, pointwise, an actual function. Also recall that we use $\widetilde{\wedge}$ rather than $\wedge$ for such differential forms). We have

$$*\overline{\omega} \widetilde{\wedge} \omega(v_{\mathbf{p}}, J(v_{\mathbf{p}})) = \det \begin{pmatrix} *\overline{\omega}(v_{\mathbf{p}}) & \omega(v_{\mathbf{p}}) \\ *\overline{\omega}(J(v_{\mathbf{p}})) & \omega(J(v_{\mathbf{p}})) \end{pmatrix} = \det \begin{pmatrix} \overline{\omega}(J(v_{\mathbf{p}})) & \omega(v_{\mathbf{p}}) \\ \overline{\omega}(J^2(v_{\mathbf{p}})) & \omega(J(v_{\mathbf{p}})) \end{pmatrix}.$$

The last equality in the above line follows by applying the definition of the complex Hodge-dual. Then, since $J^2 = i \cdot i = -1$, we have

$$*\overline{\omega} \widetilde{\wedge} \omega(v_{\mathbf{p}}, J(v_{\mathbf{p}})) = \det \begin{pmatrix} \overline{\omega}(J(v_{\mathbf{p}})) & \omega(v_{\mathbf{p}}) \\ -\overline{\omega}(v_{\mathbf{p}}) & \omega(J(v_{\mathbf{p}})) \end{pmatrix} = \overline{\omega}(J(v_{\mathbf{p}}))\omega(J(v_{\mathbf{p}})) + \overline{\omega}(v_{\mathbf{p}})\omega(v_{\mathbf{p}}).$$

Overall, we have

$$*\overline{\omega} \widetilde{\wedge} \omega(v_{\mathbf{p}}, J(v_{\mathbf{p}})) = \overline{\omega}(J(v_{\mathbf{p}}))\omega(J(v_{\mathbf{p}})) + \overline{\omega}(v_{\mathbf{p}})\omega(v_{\mathbf{p}}).$$

We claim that $\overline{\omega}(v_{\mathbf{p}})\omega(v_{\mathbf{p}}) \geq 0$ for any $v_{\mathbf{p}}$. If the claim is true, then the previous equation yields the condition

$$*\overline{\omega} \widetilde{\wedge} \omega(v_{\mathbf{p}}, J(v_{\mathbf{p}})) \geq 0, \text{with equality to zero only when } \omega = \mathbf{0}.$$

(We have equality to zero only when $\omega = \mathbf{0}$ because a sum of nonnegative terms is zero only when all the terms are zero).

If we have this condition, then we have proved $\langle\langle \cdot, \cdot \rangle\rangle$ is positive-definite, as integration over M ultimately resolves into the integral of $\overline{\omega} \widetilde{\wedge} \omega(v_{\mathbf{p}}, J(v_{\mathbf{p}}))$.

So, we must prove $\overline{\omega}(v_{\mathbf{p}})\omega(v_{\mathbf{p}}) \geq 0$, where ω is a complex differential 1-form. We can express ω and $\overline{\omega}$ in coordinates as

$$\begin{aligned}\omega &= h_1 dx^1 + h_2 dx^2 \\ \overline{\omega} &= \overline{h_1} dx^1 + \overline{h_2} dx^2,\end{aligned}$$

where $h_1, h_2 : M \to \mathbb{C}$ are smooth complex functions.

We can directly compute $\overline{\omega} \widetilde{\wedge} v_{\mathbf{p}}(\omega, v_{\mathbf{p}})$ by using the above coordinate representations. Use the fact that $h_i \overline{h_i} = |h_i|^2$, where $|\cdot|$ denotes the norm defined by $|a + bi| = \sqrt{a^2 + b^2}$, to prove the claim. $\square$

We now work towards a theorem that expresses the square of the normed residual, $|| * dz - idz||^2$, in terms of the Laplacian $\Delta = - * d * d$. (So, this theorem will use a different sign convention for the Laplacian than was used in the previous section). We need a slew of lemmas to prove the theorem.

Lemma 9.33. The complex Hodge-dual preserves this norm induced by the Hodge Hermitian inner product $\langle\langle \cdot, \cdot \rangle\rangle$. That is, for all complex differential 1-forms ω, we have $|| * \omega|| = ||\omega||$.

Proof. We defined the complex Hodge-dual by treating complex differential forms as objects that are, pointwise, actual functions. In this definition, the length of vectors input into a Hodge-dual'ed differential form were the same as the length of the vectors input to the original differential form. Since the norm induced by the Hodge Hermitian inner product is an integral involving ω, and as an integrals of a differential form can be interpreted as an integral of that differential form being evaluated on tangent vectors, this implies that the norm induced by the Hodge Hermitian inner product is invariant under $*$. $\square$

Lemma 9.34. (Green's first identity for complex differential 1-forms).

Let M be a smooth 2-manifold embedded in $\mathbb{C}$. If $f, g : M \to \mathbb{C}$ are smooth and the normal derivative of either f or g vanishes on the boundary, then

$$\langle\langle df, dg \rangle\rangle = \langle\langle \Delta f, g \rangle\rangle,$$

where $\Delta = - * d * d$ is the Laplace operator on 0-forms.

Proof. From the proof of Lemma 9.13, we have

$$\int_{\partial M} g * df = \int_M dg \wedge *df + \int_M g * \Delta f.$$

The boundary term vanishes by assumption, so we have

$$0 = \int_M dg \wedge *df + \int_M g * \Delta f = - \int_M *df \wedge dg + \int_M (*\Delta f) g = -\langle\langle df, dg \rangle\rangle + \langle\langle \Delta f, g \rangle\rangle.$$

$\square$

Lemma 9.35. Let $||\cdot||$ be the norm induced by the Hodge Hermitian inner product $\langle\langle \cdot, \cdot \rangle\rangle$ on complex differential 1-forms. We have

$$|| * dz - idz||^2 = 2\Big(\langle\langle \Delta z, z \rangle\rangle - i \int_M d\overline{z} \wedge dz\Big).$$

Proof. If $\langle \cdot, \cdot \rangle$ is a Hermitian inner product, then $||z+w||^2 = \langle z+w, z+w\rangle = \langle z, z\rangle + \langle z, w\rangle + \langle w, z\rangle + \langle w, w\rangle = ||z||^2 + \langle z, w\rangle + \overline{\langle z, w\rangle} + ||w||^2 = ||z||^2 + 2\mathrm{Re}(\langle z, w\rangle) + ||w||^2$. Applying this fact to our situation, we have $|| * dz - idz||^2 = || * dz||^2 + 2\mathrm{Re}\langle\langle *dz, -idz\rangle\rangle + ||dz||^2$. Next, we use $|| * dz|| = ||dz||$ to obtain $|| * dz - idz||^2 = 2(||dz||^2 + \langle\langle *dz, -idz\rangle\rangle)$.

To finish, first note that $||dz||^2 = \langle dz, dz\rangle = \langle \Delta z, z\rangle$ by Green's first identity for complex differential 1-forms. We now handle $\langle\langle *dz, -idz\rangle\rangle$ to complete the proof. By definition,

$$\langle\langle *dz, -idz\rangle\rangle = \mathrm{Re} \int_M * * \overline{dz} \wedge (-idz).$$

From the definition of complex Hodge-dual, $* * dz = dz$ [?]. It's simple to check that $\overline{dz} = d\overline{z}$ (ultimately this is because $\overline{df} = d\overline{f}$ for a complex differential 0-form f), so we have

$$\langle\langle *dz, -idz\rangle\rangle = \mathrm{Re} \int_M d\overline{z} \wedge (-idz) = -i \int_M d\overline{z} \wedge dz.$$

This proves the theorem. □

Lemma 9.36. Let M be a smooth 2-manifold embedded in $\mathbb{C}$. If $z : M \to \mathbb{C}$ is a complex differential 1-form on M, then

$$A(z) = -\frac{i}{2} \int_M d\overline{z} \wedge dz.$$

is the surface area of M.

Proof. We need to show $-\frac{i}{2} d\overline{z} \wedge dz = dx \wedge dy$. We show the equivalent statement $d\overline{z} \wedge dz = -2idx \wedge dy$:

$$\begin{aligned} d\overline{z} \wedge dz &= d(\overline{x+iy}) \wedge d(x+iy) = d(x-iy) \wedge d(x+iy) = (dx - idy) \wedge (dx + idy) \\ &= dx \wedge dx + idx \wedge dy - idy \wedge dx + dy \wedge dy = idx \wedge dy - idy \wedge dx = 2idx \wedge dy. \end{aligned}$$

□

Finally, we are ready to present an expression for square of the normed residual, $|| * dz - idz||^2$, in terms of the Laplacian $\Delta = - * d * d$.

Theorem 9.37. Let $|| \cdot ||$ be the norm induced by the Hodge Hermitian inner product $\langle\langle \cdot, \cdot \rangle\rangle$ on complex differential 1-forms. We have

$$|| * dz - idz||^2 = 2\langle\langle \Delta z, z\rangle\rangle - 4A(z),$$

where $A(z)$ is the surface area of M,

$$A(z) = -\frac{i}{2} \int_M d\overline{z} \wedge dz.$$

Proof. Combine the previous lemma and theorem. □

Recall that $|| * dz - idz||$ measures how much a map z fails to be angle-preserving. We want to minimize $|| * dz - idz||$, i.e., minimize $|| * dz - idz||^2$, to produce charts which are as close as possible to being perfectly angle-preserving.

According to [Cra19, p. 126], the trick to solving this minimization problem is to pick appropriate constraints. We don't want to pick constraints that are too loose, and we don't want to pick constraints that are too rigid. An example of a constraint that is too rigid is that of fixing the boundary of M.

Why is this constraint "too rigid"? Well, since $d\overline{z} \wedge dz = d(\overline{z}dz)$, we can use Stokes' theorem to express $A(z)$ as an integral over the boundary:

$$A(z) = -\frac{i}{2}\int_M d\overline{z} \wedge dz = -\frac{i}{2}\int_M d(\overline{z}dz) = -\frac{i}{2}\int_{\partial M} \overline{z}dz.$$

If the boundary is fixed, then $A(z)$ is a constant, so minimizing $||*dz - idz||^2 = 2\langle\langle \Delta z, z\rangle\rangle - 4A(z)$ is the same as minimizing $\langle\langle \Delta z, z\rangle\rangle$. But $\langle\langle \Delta z, z\rangle\rangle$ is positive and quadratic, so the minimum of $\langle\langle \Delta z, z\rangle\rangle$ occurs when $\nabla z = \mathbf{0}$, i.e., when $\Delta z = \mathrm{div}(\nabla z) = 0$. The fix proposed by [Cra19] is to solve the following minimization problem:

$$\begin{aligned} &\min_z || * dz - idz|| \\ &\text{s.t. } ||z|| = 1 \text{ for all } z \\ &\quad\; \langle z, \mathbf{1}\rangle = 0 \text{ for all } z \end{aligned}$$

Since Δ can be shown to be a self-adjoint operator, the above minimization problem can be solved using eigenvalue methods. Once this minimization problem is solved, a map that is as close to angle-preserving as possible is obtained.

Bibliography

[BW97] Javier Bonet and Richard D. Wood. Nonlinear Continuum Mechanics for Finite Element Analysis. Cambridge Univeristy Press, 1 edition, 1997.

[Cra19] Keenan Crane. Discrete Differential Geometry: An Applied Introduction. 1 edition, 5 2019. `https://www.cs.cmu.edu/~kmcrane/Projects/DDG/`.

[Ebe19] Johannes Ebert. `https://mathoverflow.net/questions/21024/what-is-the-exterior-derivative-intuitively`, 2019.

[GP74] Victor Guillemin and Alan Pollack. Differential Topology. Prentice-Hall, 1 edition, 1974.

[HH] Barbara Burke Hubbard and John Hamal Hubbard. Vector Calculus, Linear Algebra, And Differential Forms. Prentice Hall, 1 edition.

[Lee] John M. Lee. Introduction to Smooth Manifolds. Springer, 2 edition.

[MDT] Eva Kanso Mathieu Desbrun and Yiying Tong. Discrete differential forms for computational modeling. `http://www.geometry.caltech.edu/pubs/DKT05.pdf`.

[Sch15] Rich Schwartz. `https://www.math.brown.edu/~res/M114/notes11.pdf`, 2015.

[Sol13] Justin Solomon. Discrete differential geometry. `https://graphics.stanford.edu/courses/cs468-13-spring/assets/lecture14-saenz.pdf`, 2013.

[War] Frank W. Warner. Foundations of Differentiable Manifolds and Lie Groups. 1 edition.

The source that is most frequently used is [Lee]; this is a comprehensive reference for differential geometry and manifolds. Explanations in [Lee] are generally good, but results sometimes are not motivated in the best way, and connections between concepts are sometimes missed. Following [Lee], we have used [HH] the second-most. The book [HH] possibly has the best motivation I have ever seen in a math book. The definition of the exterior derivative and proof of the generalized Stokes' theorem were adapted from [HH].