Tensors, Differential Forms, and Computer Graphics

Ross Grogan-Kaylor

September 2020

Acknowledgements

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

To do

Checklist for turn in

- Are all instances of lists such as $x_1...x_r...x_n$ replaced with $x_1...x_{k-1}x_rx_{k+1}...x_n$?
- When defining maps on elementary tensors, is the elementary tensor of the form $\mathbf{v}_1 \otimes ... \otimes \mathbf{v}_p \otimes \phi^1 \otimes ... \otimes \phi^q$?
- Make sure "if" is not used in any definition, and that "iff" is always used.

To do

Make connections between tangent vectors, cotangent vectors, vector fields, covector fields. Define differential

- Write up proof of derivation of differential on a manifold
- Write up derivation of Jacobian matrix as a result of searching for the directional derivative of a function $\mathbb{R}^m \to \mathbb{R}^n$

Read Ch 16 of Lee. Be on the look out for:

- interpreting volume form as tensor product
- constructing Hodge dual in terms of volume form
- diff between volume form and det

Decide whether to treat volume forms and hodge dual in ext algebra section or in diff forms section

- check that there's enough stars in lin alg. add stars to other review sections?
- revamp notation page
- mention that dual transf is a type of pullback
- standardize the way I type "pull back", "pullback", "pull-back"
- add technical hypotheses to calc chapter
- make sure n is used for dimension of M and m is used for dimension of N. it does kind of make sense even though it's confusing. make a remark about it
- emphasize "for all" as encompassing entire proof, like brackets in programming.
- Might delete $\mathfrak{F}$ from notation. Also make note about Russian letter, and how I don't actually ever talk about elements of W^{**}
- Linear algebra

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

- Define set of permutations S_n somewhere
- Look at Hubbard for definition of integral. Do they use an actual Riemann sum?
- Cross product and Hodge dual

Outline

A motivated introduction to tensor product spaces and dual spaces

- Preview: two main concepts underlying the idea of a "tensor." Tensor product spaces are constructed for the purpose of treating multilinear functions as linear functions. Dual spaces allow us to treat linear functions as vectors, since every linear function $V \to W$ is a sum of rank 1 linear functions, i.e., a sum of elements from the dual space.
- Tensor product spaces
  - Multilinear functions (k-linear functions)
  - Tensors as multilinear elements imply quotient space definition of tensor product space. Address associativity of $\otimes$.
  - Universal property of tensor product spaces
- Motivation: discover tensor product spaces and dual spaces by discovering $\mathcal{L}(V \to W) \cong V^* \otimes W$.
  - Mention outer product, and link to physics/engineering chapter.
- Finish up discussing the various ways to prove $\mathcal{L}(V \to W) \cong V^* \otimes W$. In hindsight, there are three *natural* isomorphisms: two "forward" and one "backward." The one we discovered went "forward." The one that's easy to find in hindsight goes "backward." Lastly, asking about how coordinates of a linear function translate over yields another "forward" isomorphism.
  - Mention how the outer product arises out of choosing a basis. And then mention how it shows up later in section on physics/engineering.

Dual spaces

- Reminder: the importance of dual spaces in the context of tensors is that 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^*.
- Remark on the phrases "dual vector," "covector," and "linear functional."
- Dual basis
  - Bases aren't necessarily self-dual (orthonormal); requiring this implies a choice of basis
- $V \cong V^{**}$ naturally
- Orthonormal $\iff$ self dual
- Dual transformation
  - Let E be an orthonormal basis for V. If the matrix of $\mathbf{f}$ with respect to E and E is $\mathbf{A}$, then the matrix of $\mathbf{f}^* : V \to V$, after identifying $V \cong V^*$, with respect to E and E is $\mathbf{A}^\top$.

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

- Begin the convention of denoting contravariance and covariance with upper and lower indicies. Note that p is above and q is below in "$\binom{p}{q}$ tensor."

- Valence ($\binom{p}{q}$) and order ($p+q$) of a tensor.

- The three most important isomorphisms

  - Universal property of tensor product spaces (already proved)
  - $\mathcal{L}(V \to W) \cong W \otimes V^*$ (already proved)
  - Prove that dual distributes over tensor product: $(V \otimes W)^* \cong V^* \otimes W^*$

- Other useful natural isomorphisms: $V \otimes F \cong V$, $V \otimes W \cong W \otimes V$, $(V \otimes W)^* \cong V^* \otimes W^*$.

- Definition: $\binom{p}{q}$ tensor. Note that $\binom{p}{q}$ tensors are generalized linear maps in the sense that they are identifiable with elements of $\text{Hom}(V^{\otimes q}, V^{\otimes p}$.

- Define of covariant and contravariant tensors as $\binom{0}{q}$ and $\binom{p}{0}$ tenors, respectively. Mention "vectors" and "dual vectors" ("covectors") again.

- Push-forward of vectors, $\otimes^p f : T^p_0(V) \to T^p_0(W)$, and pull-back of dual vectors, $\otimes_q f^* : T^0_q(W) \to T^0_q(V)$. Note that the pull-back is often denoted (misleadingly) as f^*T.

- Revisit the definition of $\binom{p}{q}$ tensors.

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

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

Bilinear forms and metric tensors

- Bilinear form defn, metric tensor defn, inner product defn (is a positive-definite metric tensor)

  - Positive definite implies nondegenerate

- dual transf after identifications

  Theorem 0.1. *If V is a finite-dimensional vector space and U is any orthonormal basis for V, then the matrix of $\mathbf{f}^* : V \to V$ (after identifying $V \cong V^*$) with respect to U and U is $[\mathbf{f}(E)]_E^\top$., where $[\mathbf{f}(E)]_E$ is the matrix of $\mathbf{f}$ with respect to E and E.*

  Proof. Let $U = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$. The jth column of $[\mathbf{f}(U)]_U$ is $[\mathbf{f}(\hat{\mathbf{u}}_j)]_E$, so the ij entry of $[\mathbf{f}(U)]_U$ is $\langle \mathbf{f}(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle$. Similarly, the ij entry of the matrix of $\mathbf{f}^*$ with respect to E and E is $\langle \mathbf{f}^*(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle$. By the condition on $\langle, \rangle$ imposed by using the dual transformation after identifying $V \cong V^*$, $\langle \mathbf{f}^*(\hat{\mathbf{u}}_j), \hat{\mathbf{u}}_i \rangle = \langle \hat{\mathbf{u}}_j, \mathbf{f}(\hat{\mathbf{u}}_i) \rangle = \langle \mathbf{f}^(\hat{\mathbf{u}}_i), \hat{\mathbf{u}}_j \rangle$. But this is the ji entry of $[\mathbf{f}(U)]_U$, i.e., the ij entry of $[\mathbf{f}(U)]_U^\top$. Thus $[\mathbf{f}(U)]_U = [\mathbf{f}(U)]_U^\top$. □

- Bilinear forms and natural "musical" isomorphisms $V \cong W^*$, $W \cong V^*$; induced bilinear form on the duals

- Special case of orthonormal bases: $g_{ij} = \delta_{ij}$

- Talk about bilinear form induced by choice of basis (dot product?)

- $B(\mathbf{v}, \mathbf{w}) = [\mathbf{v}]_E^\top \mathbf{g} [\mathbf{w}]_F$, where $g_{ij} = B(\mathbf{e}_i, \mathbf{f}_j)$. In this formula, $\mathbf{g}$ is the matrix that represents a $\binom{1}{1}$ tensor. Despite this, we think of $\mathbf{g}$ as being $\binom{0}{2}$ tensor due to the fact that it expresses the action of a bilinear form on V, which is an element of $\mathcal{L}(V \times V \to F) \cong \mathcal{L}(V \otimes V \to F) = (V \otimes V)^* \cong V^* \otimes V^* \cong T^0_2(V)$.

  - We could technically think of $\mathbf{g}$ as obtained from B by tracing through the previously mentioned isomorphisms, but explicitly doing this would be difficult, since we only understand the "forward" version of the isomorphism $(V \otimes V)^* \to V^* \otimes V^*$ as the inverse of its "reverse." Instead, we can notice that every bilinear form is determined by how it acts on the tuples of basis vectors, $(\mathbf{e}_i, \mathbf{f}_j)$. Therefore it is identifiable with a matrix, i.e., with a $\binom{1}{1}$ tensor, whose ${}^i_j$ entry is $B(\mathbf{e}_i, \mathbf{f}_j)$.

- Def: a metric tensor on V is a nondegenerate symmetric bilinear form on V. (So every inner product is a metric tensor, but not every metric tensor is an inner product. An inner product is a positive-definite metric tensor- in Appendix). Covariant metric tensor $\mathbf{g}$, with ${}_{ij}$ entry g_{ij} and contravariant metric tensor $\mathbf{g}^{-1}$ with ^{ij} entry g^{ij}

Coordinates of tensors

- Two facts about coordinates: $([\mathbf{v}]_E)_i = ([\mathbf{v}]_{E^{**}})_i = \epsilon_i(\mathbf{v})$ and $([\phi]_{E^*})_i = \phi(\mathbf{e}_i)$.
- In general (when we might not have a metric tensor), define $v^i := ([\mathbf{v}]_E)_i$, $\phi_i := ([\phi]_{E^*})_i$.
- When we do have a metric tensor, then additionally define $v_i := (\phi_{\mathbf{v}})_i$. Here $\phi_{\mathbf{v}}$ is obtained through the musical isomorphism, which is natural. Prove $v^i = \sum_j v_j g^{ij}$ and $v_i = \sum_j v^j g_{ij}$.
- When we have a metric tensor g, then $V \cong V^*$ naturally, so we can convert between covariant and contravariant tensors.
  - Staggered index notation for $\binom{p}{q}$ tensors. Raise index i of T_{ij}: ${T^i}_j := \sum_k g^{ik} T_{kj}$. Raise index i of T_{ji}: ${T_j}^i := \sum_k g^{ik} T_{jk}$.
    To write an arbitrary index for a $\binom{p}{q}$ tensor, we *could* come up with notation for arbitrary staggering of indcies. Instead, to keep things neater, we use the convention of using ${T^{i_1 \dots i_p}}_{j_1 \dots j_q}$; then we can raise/lower indices in whatever order necessary by multiplying by g^{ij} or g_{ij} and summing as necessary.
- Def: $[\cdot]_{E,F}$
- δ^i_j vs. δ^{ij} vs. δ_{ij}
- Changing basis
  - Induced change of basis on dual spaces
  - Ricci transformation for $\binom{p}{q}$ tensors, justification of the words "covariant" and "contravariant."
- Tensor contraction
  - Result of asking: "how does composition composition of linear maps generalize?"
  - There is a natural bilinear form C on V and V^* defined by $C(\mathbf{v}, \phi) = \phi(\mathbf{v})$. Define $\binom{k}{\ell}$ tensor contraction on an elementary tensor $\mathbf{v}^1 \otimes \dots \otimes \mathbf{v}^p \otimes \phi_1 \otimes \dots \otimes \phi_q$ with use of C. This implies an equivalent definition on basis elements.
  - Trace as a special case. An element of $\mathcal{L}(V, V)$ with matrix $\mathbf{A} = (a_{ij})$ coorresponds to the $\binom{1}{1}$ tensor $\sum_{ij} a^j_i \epsilon_j \otimes \mathbf{e}^i$, which contracts to $\sum_{ij} a^j_i \epsilon_j(\mathbf{e}^i) = \sum_{ij} a^j_i \delta_{ji} = \sum_i a^i_i = \mathrm{tr}(\mathbf{A})$.
  - Discuss what happens to components; upper indices get "contracted against" lower indices.
  - Contraction is basis-independent
  - Induced contraction between any two $\binom{p}{q}$ tensors, given a bilinear g form on V. (Recall that g allows for conversion between covariant and contravariant).
    - Example: double contraction (double dot product) is the induced contraction between order 2 tensors. (The valence doesn't matter because the result is a scalar).

Equivalent definitions of "tensor"

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

Exterior powers

- Define alt, and $T \wedge S := \text{alt}(T \otimes S)$
- Properties of $\wedge$. $\wedge$ is still "multilinear." Address associativity of $\wedge$, alternatingness of $\wedge$. Show alternatingness iff $\mathbf{v} \wedge \mathbf{v} = 0$ when $\text{char}(F) > 2$.
- $\Lambda^k(V) := \text{alt}(V^{\otimes k}) = \text{alt}(T_0^k(V))$.
- Basis, dimension of $\Lambda^k(V)$.
- The three most important isomorphisms, translated to exterior powers
  - Universal property of kth exterior power. An alternating multilinear map uniquely $f : V^{\times k} \to W$ uniquely corresponds to a linear map $h : \Lambda^k(V) \to W$, where $f = h \circ g$. Proof similar to that for tensor product spaces.
  - $\text{alt}(V^* \otimes W) \cong \text{alt}(\mathcal{L}(V \to W))$.
  - $\Lambda^k(V^*) \cong \Lambda^k(V)^*$. Proof analagous to the one for tensors. Define an isomorphim on elementary tensors by $\mathbf{T} = \phi^1 \wedge ... \wedge \phi^k \mapsto f_{\mathbf{T}}, f_{\mathbf{T}}(\mathbf{v}_1, ..., \mathbf{v}_k) = (\phi^1 \overline{\wedge} ... \overline{\wedge} \phi^k)(\mathbf{v}_1, ..., \mathbf{v}_k)$, and extend using alternatingness and multilinearity.
- The determinant
  - Define det of a square matrix axiomatically. Mention how volume being additive implies that volume is signed, or oriented. Talk about permutation formula as showing existence and uniqueness. Mention that $\text{sgn}(\sigma)$ is well-defined.
  - Laplace expansion: use ALA notes
  - Det of upper triangular
  - Define det of a linear function $V \to V$ as det of the linear function's matrix with respect to E and E, where E is a basis for V. Prove that, since we used same basis twice, det of a linear function is basis-independent.
  - Show that given a linear function $\mathbf{f} : V \to V$, the induced map $\Lambda^{\dim(V)} f$ on $\Lambda^k(V)$ is multiplication by $\det(\mathbf{f})$. Prove using same process used to derive permutation formula of det for matrices, but to divide out by $\mathbf{e}_1 \overline{\wedge} ... \overline{\wedge} \mathbf{e}_k$ at end.

- Properties of det
  - $\mathbf{f}$ is invertible iff $\det(\mathbf{f}) \neq 0$
  - Product rule
  - $\det(\mathbf{f}^*) = \det(\mathbf{f})$. Follows because of two facts. (1) if $\mathbf{A}$ is matrix of $\mathbf{f}$ wrt orthonormal bases U_1, U_2, then $\mathbf{A}^\top$ is matrix of $\mathbf{f}^*$ wrt U_2^*, U_1^*. (2) $\det(\mathbf{A}^\top) = \det(\mathbf{A})$, which is true because a determinant is a sum of determinants of diagonal matrices $\mathbf{D}$, which have the property $\det(\mathbf{D}^\top) = \det(\mathbf{D})$.
  - $\det(\mathbf{f}^{-1}) = \frac{1}{\det(\mathbf{f})}$. Follows because determinant is a group homomorphism (the product rule).
  - Mention that the following are in appendix: Laplace expansion, adjoint, Cramer's rule
- Lemma. Suppose $\mathbf{f} : V \to V$, so then $\mathbf{f}^* : V^* \to V^*$. Then $\det(\mathbf{f}^*)$ is the determinant of the matrix of $\mathbf{f}^*$ relative to E^* and E^*, which is

$$\Big([\mathbf{f}^*(\epsilon^1)]_{E^*} \quad \dots \quad [\mathbf{f}^*(\epsilon^n)]_{E^*}\Big) = \Big([\phi^1]_{E^*} \quad \dots \quad [\phi^n]_{E^*}\Big).$$

  Take a moment to talk about geometric interpretation of this.
  We have $([\phi_i]_{E^*})_j = \phi^i(\mathbf{e}_j)$ (see Theorem 5.1). Therefore

$$\det(\mathbf{f}^*) = \det(\phi^i(\mathbf{e}_j)),$$

  where $(\phi^i(\mathbf{e}_j))$ is the matrix with ij entry $\phi^i(\mathbf{e}_j)$.
- Lemma. $(\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k)(\mathbf{e}_1, ..., \mathbf{e}_k) = \det(\epsilon^i(\mathbf{e}_j) = \det(\mathbf{I}) = 1$.

  Proof.

$$\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k = \overline{\text{alt}}(\epsilon^1 \overline{\otimes} ... \overline{\otimes} \epsilon^k) = \frac{1}{k!} \sum_{\sigma \in S_n} \text{sgn}(\sigma) \epsilon^{\sigma(1)} \overline{\otimes} ... \overline{\otimes} \epsilon^{\sigma(k)}$$

$$\begin{aligned}(\epsilon^{\sigma(1)} \overline{\otimes} ... \overline{\otimes} \epsilon^{\sigma(k)})(\mathbf{e}_1, ..., \mathbf{e}_k) &= \Big(\frac{1}{k!} \sum_{\sigma \in S_n} \text{sgn}(\sigma) \epsilon^{\sigma(1)} \overline{\otimes} ... \overline{\otimes} \epsilon^{\sigma(k)}\Big)(\mathbf{e}_1, ..., \mathbf{e}_k) \\ &= \frac{1}{k!} \sum_{\sigma \in S_n} \text{sgn}(\sigma) \epsilon^{\sigma(1)}(\mathbf{e}_1) ... \epsilon^{\sigma(k)}(\mathbf{e}_k).\end{aligned}$$

  Since $\epsilon^i(\mathbf{e}_j) = \delta^i_j$, the only permutation $\sigma \in S_n$ for which the sum argument is nonzero is the identity permutation. Therefore the sum is $k!$, so the final result is $\frac{k!}{k!} = 1$. □

- Theorem. Let $\mathbf{f} : V \to V$, so $\mathbf{f}^* : V^* \to V^*$. Consider $\Lambda^n(V^*) = \text{alt}(T_k^0(V))$. If $\phi^i = \mathbf{f}^*(\epsilon^i)$, then $\phi^1 \overline{\wedge} ... \overline{\wedge} \phi^n = \det(\mathbf{f}^*) \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n$. Apply both sides to $(\mathbf{e}_1, ..., \mathbf{e}_n)$ to see $(\phi^1 \overline{\wedge} ... \overline{\wedge} \phi^n)(\mathbf{e}_1, ..., \mathbf{e}_n) = \det(\phi^i(\mathbf{e}_j))$. Extend with multilinearity and alternatingness to see $(\phi^1 \overline{\wedge} ... \overline{\wedge} \phi^n)(\mathbf{v}_1, ..., \mathbf{v}_n) = \det(\phi^i(\mathbf{v}_j))$.
- Above thm holds for any $k \leq n$. Argue using vector subspaces.
- $\Lambda^k(V^*) \cong (\Lambda(V))^*$ naturally? `https://math.stackexchange.com/questions/18595/exterior-powe` `18628#18628`
- Push-forward $\Lambda^k f$ and pullback $\Lambda^k f^*$

- Orientation of finite-dimensional vector spaces.
  - The notion of "same orientation" for a 2-dimensional vector space; swap-negate and linearly combine principle for 2D.

- Consider ordered permuted standard bases of $\mathbb{R}^n$. We say that two ordered bases $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ and $F = \{\mathbf{f}_1, ..., \mathbf{f}_n\}$ have the same orientation iff the 2-dimensional subspaces $\text{span}(\mathbf{e}_i, \mathbf{e}_j)$ and $\text{span}(\mathbf{f}_i, \mathbf{f}_j)$ have the same orientation for all $i, j \in \{1, ..., n\}$.
- The above defn implies ordered orthonormal bases of an n-dimensional vector space have the same orientation iff the swap-negate and linearly combine thing holds for ordered bases of n vectors.
- There are two equivalence classes of "having the same orientation."
- Each equivalence class can be identified with an element of $\Lambda^n(V^*)$, up to a scalar multiple.
- The above implies that det tracks orientation: the equivalence class $[\{\mathbf{f}_1, ..., \mathbf{f}_n\}]_\sim$ can be identified with the equivalence class $[\mathbf{f}_1 \wedge ... \wedge \mathbf{f}_n]_\sim$. If $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is another basis for V and $\mathbf{f}$ is the linear function satisfying $\mathbf{f}(\mathbf{e}_i) = \mathbf{f}_i$ (i.e. $\mathbf{f} = (\mathbf{I}_V)_{E,F}$) then $\mathbf{f}_1 \wedge ... \wedge \mathbf{f}_n$ has the same orientation as $\mathbf{e}_1 \wedge ... \wedge \mathbf{e}_n$ iff $\det(\mathbf{f}) > 0$.
  - More concrete way of seeing how the determinant is involved: think about the determinant of the matrix with columns; it's invariant under swap-negate and linearly combining columns into others.
- Translate the above into a statement about $\Lambda^n(V)^* \cong \text{alt}(\mathcal{L}(V^{\times n} \to K))$. See Lee.

- From here, probably switch over to using actual function interpretation of exterior powers.
- Volume form μ
  - Define volume form as n-form μ for which $\mu(\mathbf{X}_1, ..., \mathbf{X}_n) = \det(\mathbf{X}_1, ..., \mathbf{X}_n)$, where the det on the RHS is the function on $\mathfrak{X}(M)^{\times n} \times M$ sending $(\mathbf{X}_1, ..., \mathbf{X}_n, \mathbf{x}) \mapsto \det(\mathbf{X}_1(\mathbf{x}), ..., \mathbf{X}_n(\mathbf{x}))$. Use "8.15. Proposition" of Sjmaar.
  - Sjmaar p. 108 examples about how μ relates to ds, dA, dV
  - Use Lee Prop. 15.29 and 15.31 give formulas about volume forms
- Hodge duality.
  - Let V be a finite-dimensional vector space and consider an ordered basis $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ for V. Let $\sigma_W \in S_k$ and $\sigma_{W^\perp} \in S_{n-k}$. Set W to be the oriented subspace of V spanned by the ordered basis $\{\mathbf{e}_{\sigma(1)}, ..., \mathbf{e}_{\sigma(k)}\}$, and give $W^\perp \subseteq V$ the orientation specified by $\{\mathbf{e}_{\sigma(k+1)}, ..., \mathbf{e}_{\sigma(n)}\}$.

    Equivalently, the orientations chosen on W and $W^\perp$ can be represented as elements of $\Lambda^k(W) \subseteq \Lambda^k(V)$ and $\Lambda^{n-k}(W^\perp) \subseteq \Lambda^{n-k}(V)$, respectively. The orientation on W is specified by $\mathbf{e}_{\sigma(1)} \wedge ... \wedge \mathbf{e}_{\sigma(k)}$ and the orientation of $W^\perp$ is specified by $\mathbf{e}_{\sigma(k+1)} \wedge ... \wedge \mathbf{e}_{\sigma(n)}$.

    The *Hodge dual*, or *Hodge star*, is the map $* : \Lambda^k(W) \to \Lambda^{n-k}(W^\perp)$ which acts on the orientation chosen for a subspace $W \subseteq V$ to the orientation chosen for $W^\perp$.

    It is enough to say that $*$ "acts" on $\mathbf{e}_{\sigma(1)} \wedge ... \wedge \mathbf{e}_{\sigma(k)}$ as the action on this element of $\Lambda^k(W)$ is implied: $*$ is a map of one-dimensional vector spaces, so we know

$$*(\mathbf{e}_{\sigma(1)} \wedge ... \wedge \mathbf{e}_{\sigma(k)}) =$$

Differential forms

- Don't have to use $\barwedge$ (interpretation of diff form at a point as actual alteranting multilinear function), can use $\wedge$ (interpretation of diff form at a point as alternating tensor)!
- A differential form at $\mathbf{x} \in M$ is an element of $\Omega^k(T_\mathbf{x}(M)^*)$. Set of differential forms on M denoted $\Omega^k(M)$.
- Theorem on differential forms at a point $\mathbf{x}$ using basis

- (This bullet point uses actual fn interp of diff forms).

  We have $d\mathbf{f} : \Omega^k(M) \to \Omega_k(N)$, $d\mathbf{f}_\mathbf{x} : T_\mathbf{x}(M) \to T_\mathbf{x}(N)$, and $(d\mathbf{f}_\mathbf{x})^* : T_{\mathbf{f}(\mathbf{x})}(N)^* \to T_\mathbf{x}(M)^*$.

  The pull-back $\overline{\Omega^k}\mathbf{f}^* : \Omega^k(N) \to \Omega_k(N)$ of a differential k-form ω on N is defined by $\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)(\mathbf{x}) = \Big(\Lambda^k(d\mathbf{f}_\mathbf{x})^*\Big)(\omega(\mathbf{f}(\mathbf{x}))$. Using the argument as was made for the pull-back of an element of $\Lambda^k(W^*)$ (see [...]), we have $\Big(\Lambda^k(d\mathbf{f}_\mathbf{x})^*\Big)(\omega(\mathbf{f}(\mathbf{x}))) = \omega(\mathbf{f}(\mathbf{x})) \circ d\mathbf{f}_\mathbf{x} = (d\mathbf{f}_\mathbf{x})^*(\omega(\mathbf{f}(\mathbf{x})))$. More explicitly, $\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)(\mathbf{x})(\mathbf{v}_1, ..., \mathbf{v}_k) = \omega(\mathbf{f}(\mathbf{x}))(d\mathbf{f}_\mathbf{x}(\mathbf{v}_1), ..., d\mathbf{f}_\mathbf{x}(\mathbf{v}_k))$.

  Also by looking back at how def of $\Lambda^k\mathbf{f}^*$ translates over to $\overline{\Lambda^k}\mathbf{f}^*$, we see that if $\omega(\mathbf{x}) = \alpha(\mathbf{x})\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k$, then

  $$\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)(\mathbf{x}) = \det((d\mathbf{f}_\mathbf{x})^*)\alpha(\mathbf{f}(\mathbf{x}))\ (d\mathbf{f}_\mathbf{x})^*(\epsilon^1)\overline{\wedge}...\overline{\wedge}(d\mathbf{f}_\mathbf{x})^*(\epsilon^k).$$

  Set $\delta^i = (d\mathbf{f}_\mathbf{x})^*(\epsilon^i)$, i.e., $\delta^i = \epsilon^i \circ d\mathbf{f}_\mathbf{x}$ to restate this as

  $$\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)(\mathbf{x}) = \det((d\mathbf{f}_\mathbf{x})^*)\alpha(\mathbf{f}(\mathbf{x}))\ \delta^1\overline{\wedge}...\overline{\wedge}\delta^k.$$

  Therefore, using that $\det((d\mathbf{f}_\mathbf{x})^*) = \det(d\mathbf{f})$ and suppressing dependence on $\mathbf{x}$, we have

  $$(\overline{\Omega^k}\mathbf{f}^*)(\omega) = \det(d\mathbf{f})(\alpha \circ \mathbf{f})\ \delta^1\overline{\wedge}...\overline{\wedge}\delta^k.$$

  Here $d\mathbf{f}$ denotes the function which sends $\mathbf{x} \mapsto d\mathbf{f}_\mathbf{x}$.

  - Proof that $\det(d\mathbf{f}_\mathbf{x})$ is involved in the change of variables theorem.
    By Lemma [...] we have $\det((d\mathbf{f}_\mathbf{x})^*) = \det(\delta^i(\mathbf{e}_j))$. Then $\delta^i(\mathbf{e}_j) = (\epsilon^i \circ d\mathbf{f}_\mathbf{x})(\mathbf{e}_j) = \epsilon^i(d\mathbf{f}_\mathbf{x}(\mathbf{e}_j))$. But $d\mathbf{f}_\mathbf{x}(\mathbf{e}_j)$ is the directional derivative of $\mathbf{f}$ in the $\mathbf{e}_j$ direction, $d\mathbf{f}_\mathbf{x}(\mathbf{e}_j) = \frac{\partial \mathbf{f}(x_1,...,x_j,...,x_k)}{\partial x_j}$. We also have $\epsilon^i(\mathbf{v}) = ([\mathbf{v}]_E)^i$, so $\delta^i(\mathbf{e}_j) = ([\frac{\partial \mathbf{f}(x_1,...,x_j,...,x_k)}{\partial x_j}]_E)^i$. If $\mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_k(\mathbf{x}) \end{pmatrix}$, then by definition of partial derivative of a vector-valued function, we have $([\frac{\partial \mathbf{f}(x_1,...,x_j,...,x_k)}{\partial x_j}]_E)^i = \frac{\partial f_i(\mathbf{x})}{\partial x_j}$. Therefore $\det((d\mathbf{f})^*) = \det(\frac{\partial f_i(\mathbf{x})}{\partial x_j})$.
  - $\Omega^k\mathbf{f}^*$ and $\overline{\Omega^k\mathbf{f}^*}$ are nonstandard notation; people denote $\overline{\Omega^k}\mathbf{f}^*$ by $\mathbf{f}^*$, even though this can be confused with the dual transformation.

- Notation: $dx_i := \epsilon^i$ is a basis for $\Lambda^k(T_\mathbf{x}(M))^*$; $df_i, y_i := \delta^i$ is a basis for $\Lambda^k(T_{\mathbf{f}(\mathbf{x})}(N))^*$

- Change of variables theorem. Consider open subsets U, V of $\mathbb{R}^k$, a diffeomorphism $\mathbf{f} : V \to U$, and an integrable function $\alpha : U \to \mathbb{R}$ is integrable on U. Then

  $$\int_U \alpha = \int_V (\alpha \circ \mathbf{f})|\det(d\mathbf{f})|.$$

  So, we define, for $\omega = \alpha\ dx_1\overline{\wedge}...\overline{\wedge}dx_k$,

  $$\int_U \alpha\ dx_1\overline{\wedge}...\overline{\wedge}dx_k := \int_U \alpha.$$

Using $dx_1...dx_n$ as a "placeholder" on the right side (but not on the left), as is often done in an integral, this definition is

$$\int_U \alpha \; dx_1 \overline{\wedge} ... \overline{\wedge} dx_n := \int_U \alpha dx_1 ... dx_n.$$

The idea behind the definition is that $dx_1...dx_n$ are "secretly" $dx_1 \overline{\wedge} ... \overline{\wedge} dx_n$. While the definition technically defines the LHS in terms of the RHS, you might think of it as "secretly" giving a new meaning to the old placeholder notation of the RHS.

Then when $\mathbf{f} : V \to U$ is orientation preserving, $\det(d\mathbf{f}) > 0$, so $|\det(d\mathbf{f})| = \det(d\mathbf{f})$, and the change of variables theorem is restated as

$$\int_U \omega = \int_V \Big(\Omega^k \mathbf{f}^*\Big)(\omega).$$

Manifolds

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

- Outer product
- If we have identified $V \cong V^*$, then every linear transformation $V \to W$ has a matrix of the form $\sum_k \mathbf{w}_k \mathbf{v}_k^\top =^\top$.
  - By the discussion on rank 1 tensors above, every element of $\text{Hom}(V, W)$ has a matrix of the form $\sum_k [\mathbf{w}_k]_F [\mathbf{v}_k]_E^\top$, where $[\mathbf{v}_k]_F = (v_{k1}, ..., v_{kn})^\top, [\mathbf{w}_k]_E = (w_{k1}, ..., w_{kn})^\top$. The matrix $\sum_k [\mathbf{w}_k]_F [\mathbf{v}_k]_E^\top$ has ij entry $\sum_k w_{ki} v_{kj}$. Let be the matrix whose ith column is $[\mathbf{w}_i]_F$ and be the matrix whose ith column is $[\mathbf{v}_i]_E$. Then the ij entry of $\sum_k [\mathbf{w}_k]_F [\mathbf{v}_k]_E^\top$ is $\sum_k w_{ki} v_{kj} =_{:i} \cdot_{:j} =_{i:}^\top \cdot_{:j} = ij$ entry of $^\top$. Therefore $\sum_k [\mathbf{v}_k]_E [\mathbf{w}_k]_F^\top =^\top$.

References

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

- Exterior powers (https://kconrad.math.uconn.edu/blurbs/linmultialg/extmod.pdf)

References not used

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

About this book

Main goals

Tentative prerequisites and reading advice

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

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

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

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

On the prominence of algebraic structure

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

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

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

Other

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

Contents

Notation

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

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

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

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

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

1

Review of logic, proofs, and functions

1.1 Propositions and predicates

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

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

1.1.1 Logical operators

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

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

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

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

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

P	not P
T	F
F	T

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

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

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

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

1.2 Quantificational logic

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

1.2.1 Quantifiers

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

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

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

Nested quantifers

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

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

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

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

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

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

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

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

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

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

Negating quantifiers

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

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

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

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

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

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

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

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

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

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

1.3 Implications with sets

1.3.1 Sets

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

1.3.2 Constructing sets

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

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

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

1.3.3 Implications

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

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

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

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

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

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

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

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

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

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

"Necessary" and "sufficient"

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

Quantifiers with set membership

We define

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

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

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

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

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

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

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

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

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

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

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

Common shorthand

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

Converses

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

1.3.4 "If and only if"

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

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

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

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

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

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

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

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

1.3.5 The contrapositive and proof by contradiction

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

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

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

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

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

Proof by contradiction

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

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

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

1.4 Sets

1.4.1 Set equality, subsets

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

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

1.4.2 Common sets

We define notation for many common infinite sets.

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

1.4.3 Indexing sets

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

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

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

1.4.4 Union and intersection

We define the following operations on sets.

1.4.5 There is no universal set in ZFC

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

Other versions of set theory can handle

1.4.6 Set difference and set complement

DeMorgan's laws for sets

1.4.7 Ordered pairs

1.4.8 Cartesian product

1.5 Relations

inverse relations

inverse functions <=> bijection

1.5.1 Equivalence relations

Definition 1.1. (Equivalence relation).

Definition 1.2. (Quotient set).

1.5.2 Functions

Latin meanings of injective and surjective.

Definition 1.3. (Function).

Definition 1.4. (Uniqueness).
$\exists! x \ P(x) :\cong (\exists x_0 \ P(x_0))$ and $(\forall x \ P(x) \implies x = x_0)$.

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

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

Definition 1.7. (Onto, surjection).

Definition 1.8. (Bijection).

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

Definition 1.9. (Inverse function).

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

1.5.3 Cardinality of sets

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

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

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

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

Definition 1.12. (Proof by induction).
Defn?

2

Review of linear algebra

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

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

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

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

2.1 Vector spaces, span, and linear independence

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.1.1 Span and linear independence

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

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

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

where the c_i's are some scalars in K.

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

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

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

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

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

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

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

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

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

2.1.2 Bases and dimension of vector spaces

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

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

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

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

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

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

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

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

Lemma 2.20. (Linear dependence lemma). Let V and W be vector spaces over a field K. If $\mathbf{v}_1, ..., \mathbf{v}_k \in V$ and $\mathbf{w}_1, ..., \mathbf{w}_\ell \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_k)$ 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 $c_iq_i\mathbf{w}_\ell$ terms, and obtain an equation of the form $d_1\mathbf{w}_1 + ...d_\ell\mathbf{w}_\ell + d_{\ell+1}\mathbf{w}_{\ell+1} = \mathbf{0}$ for some scalars d_i. As the c_i's are not all zero, then the d_i's must also not all be zero. This means the $\mathbf{w}_i$'s are linearly dependent. □

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

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

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

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

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

2.2 Linear functions

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

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

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

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

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

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

Another equivalent condition is

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

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

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

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

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

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

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

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

2.2.1 Kernel and image of a linear function

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Theorem 2.32. (A linear function is one-to-one iff it is onto).

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

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

$\mathbf{f}$ is not one-to-one $\iff$ $\mathbf{f}$ has a nontrivial kernel $\iff$ $\dim(\mathbf{f}^{-1}(\mathbf{0})) \neq 0$
$\iff$ $\dim(\mathbf{f}(V)) < \dim(V)$ $\iff$ $\mathbf{f}$ is not onto.

Above, we have used the fact that $\mathbf{f}$ is one-to-one iff it has a trivial kernel (2.29) as well as the main dimension theorem (the previous theorem).

The rightmost equivalence can be formally checked with use of Lemma 2.21. $\square$

Definition 2.33. ($\star$ Linear isomorphism $\star$).

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

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

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

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

Definition 2.34. ($\star$ Natural linear isomorphism $\star$).

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

2.2.2 Inverse of a linear function

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

Theorem 2.35. (The inverse of a linear function is a linear function). Consider an invertible linear function $\mathbf{f} : V \to W$. The inverse $\mathbf{f}^{-1}$ is also a linear function.

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

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

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

Theorem 2.36. (Dimensions of spaces must be equal when linear function invertible).

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

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

Theorem 2.37. (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

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

$$\text{if and only if}$$

$$(\mathbf{f} \text{ is invertible})$$

Proof.

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

($\implies$). Suppose that if $\mathbf{v}_1, ..., \mathbf{v}_k$ are linearly independent, then $\mathbf{f}(\mathbf{v}_1), ..., \mathbf{f}(\mathbf{v}_n)$ are linearly independent. We need to show $\mathbf{f}$ is invertible; 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. Using this together with the assumption from the beginning of the ($\implies$) direction implies that $([\mathbf{v}]_E)_i = 0$ for all i is the only solution to $\sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{f}(\mathbf{e}_i) = \mathbf{0}$. Therefore $([\mathbf{v}]_E)_i = 0$ for all i, so $\mathbf{v} = \mathbf{0}$. $\square$

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

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

2.2.3 Matrices and coordinatization

⋆ 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 2.38. (Coordinates of a vector relative to a basis). Given a vector $\mathbf{v} \in V$, we define $[\mathbf{v}]_E$ to be the vector in $K^{\dim(V)}$ that stores the *coordinates of* $\mathbf{v}$ *relative to the basis* E. Formally, $[\mathbf{v}]_E$ is the tuple of scalars

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

for which

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

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

Definition 2.39. (Linear function acting on a list of vectors). Consider a linear function $\mathbf{f} : V \to K^m$. We define the notation

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

Note that when $\mathbf{f}$ is invertible and E is a basis of V, then the columns of the matrix $\mathbf{f}(E)$ are a basis of K^m, since invertible linear functions preserve linear independence (see Theorem 2.37).

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

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

Theorem 2.40. Taking coordinates relative to a basis is an invertible linear operation. Put differently, $[\cdot]_E$ is an invertible linear function.

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

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

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

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

Matrices as representative of linear functions

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

Derivation 2.41. ($\star$ Primitive matrix of a linear function $V \to K^m$ $\star$).

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

To start, consider a linear function $\mathbf{f} : V \to K^m$. Then from the definition of $[\cdot]_E$ (see Definition 2.38) we have $\mathbf{v} = \sum_{i=1}^n ([\mathbf{v}]_E)_i \mathbf{e}_i$, so

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

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

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

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

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

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

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

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

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

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

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

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

Derivation 2.42. ($\star$ Matrix of a linear function relative to bases $\star$).

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

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

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

We call this matrix *the matrix of* $\mathbf{f}$ *relative to the bases* E *and* F. This matrix is $[\mathbf{f}(E)]_F$, where we have used Definition 2.39 to define $[\mathbf{f}(E)]_F := [\cdot]_F(E)$. That is,

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

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

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

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

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

Remark 2.43. (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 $[\mathbf{f}(E)]_{\mathfrak{E}} = \mathbf{f}(E)$ of $\mathbf{f} : V \to K^m$ relative to the bases E and $\mathfrak{E}$, where $\mathfrak{E}$ is the standard basis for K^m. The key fact here is that $\left[\begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}\right]_{\mathfrak{E}} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$.

Remark 2.44. (Matrix-vector product pedagogy). All the linear algebra texts I have read always present the relationship between linear functions and matrices in the following way: first define matrices as grids of scalars (often in the context of systems of linear equations), then define a linear function as satisfying the second condition of Definition 2.23, and then prove that each linear function has a matrix. I believe this is bad pedagogy; there should be no need to conjecture and prove that a matrix-vector product corresponds to the action of a linear function, because this fact is apparent from Derivation 2.41.

It's important to emphasize that definition of matrix-vector product is therefore really no more than a definition. We decided that writing a grid of scalars next to a vector in the specific way of Derivation 2.41 should produce the output of Derivation 2.41 because this is what formalizes the correspondence between linear functions and the lists of their transformed basis vectors. Now, it is true that this definition of matrix-vector product leads to somewhat complicated formulas for the ith entry of a matrix-vector product and for the ij entry of a matrix-matrix product (these formulas will be presented in Theorems 2.49 and 2.50). Upon seeing these index-notation formulas, remember that they are consequences, and are not "just the way things are"!

Theorem 2.45. (Matrices with respect to bases are also primitive matrices).

Let V and W be a finite-dimensional vector spaces with bases E and F, respectively. We just learned that every linear function $V \to W$ corresponds to a matrix with respect to bases. Here we see that every matrix with respect to bases can be viewed as a primitive matrix (see Derivation 2.41).

The previous derivation showed $[\mathbf{f}(\mathbf{v})]_F = [\mathbf{f}(E)]_F[\mathbf{v}]_E$. We rephrase this 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

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

so that the above rephrases as

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

So, $[\mathbf{f}(E)]_F$ is the "primitive" matrix of $\mathbf{f}_{E,F} : K^{\dim(V)} \to K^{\dim(W)}$ relative to E, in the sense which is applicable only to linear functions $V \to K^m$ (see Derivation 2.41).

We can understand $\mathbf{f}_{E,F}$ to be the "induced" linear function for which this diagram commutes:

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

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

Definition 2.46. ($\star$ Composition of linear functions that map into K^m and K^p $\star$).

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

The composition $\mathbf{g} \circ \mathbf{f} : V \to K^p$ is also a linear function (prove this as an exercise). Since $\mathbf{g} \circ \mathbf{f}$ is a linear function $V \to K^p$, its primitive matrix relative to E (see Derivation 2.41) is

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

But $\mathbf{f} : V \to K^m$ and $\mathbf{g} : K^m \to K^p$, so $\mathbf{f}$ and $\mathbf{g}$ also have primitive matrices. The primitive matrix of $\mathbf{f}$ relative to E is $[\mathbf{f}(E)]_{\mathfrak{E}} = \mathbf{f}(E)$ and the primitive matrix of $\mathbf{g}$ relative to $\mathfrak{E}$ is $[\mathbf{g}(\mathfrak{E})]_{\mathfrak{F}} = \mathbf{g}(\mathfrak{E})$ (see Remark 2.43). The above then becomes

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

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

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

so that

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

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 $\mathbf{g}(\mathfrak{E})$ and $\mathbf{f}(E)$.

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

Remark 2.47. (Compatibility of matrices for matrix-matrix products). The composition of two linear functions is only defined when the output space of one is the entire input space of the other. Thus, the matrix-matrix product $\mathbf{BA}$ of an $m \times n$ matrix with an $r \times s$ matrix $\mathbf{B}$ is only defined when $r = n$.

Theorem 2.48. (Matrix-matrix products of linear functions mapping into finite-dimensional vector spaces).

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

The matrix of $\mathbf{g} \circ \mathbf{f}$ relative to E and G is the same as the primitive matrix for $(\mathbf{g} \circ \mathbf{f})_{E,G}$ relative to E (see Theorem 2.45). We will therefore compute this later matrix. We have

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

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

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

Therefore

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

Theorem 2.49. ($\star$ ith entry of matrix-vector product $\star$).

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix with entries in a field K and let $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in K^n$ be a column vector. Referring to the definition of matrix-vector product in Derivation 2.41, we see the matrix-vector product $\mathbf{Ac}$ has the following ith entry:

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

Therefore,

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

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

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

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

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

Theorem 2.50. ($\star$ ij entry of matrix-matrix product $\star$).

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix with entries in a field K and $\mathbf{B} = (b_{ij})$ be an $n \times p$ matrix with entries in K. Then the ij entry of the matrix-matrix product $\mathbf{BA}$ can be computed using the definition of matrix-matrix product (Theorem 2.48) and the previous theorem, which gives a formula for the ith entry of a vector:

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

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

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

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

Change of basis

Theorem 2.51. ($\star$ Change of basis for vectors $\star$).

Let V be a finite-dimensional vector space with bases E and F, and let W be a different finite-dimensional vector space. In Theorem 2.45, we saw that the primitive matrix of $\mathbf{f}_{E,F} : K^{\dim(V)} \to K^{\dim(W)}$ is $[\mathbf{f}(E)]_F$. In order to convert the coordinates of a vector relative to E into coordinates relative to F, we consider the case $V = W$. Recall from Derivation 2.42 that $[\mathbf{f}(E)]_F$ is characterized by the fact that

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

This statement is equivalent to

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

In the special case when $\mathbf{f} = \mathbf{I}_V$ is the identity on V, then the characterizing properties of $[\mathbf{f}(E)]_F$ and $\mathbf{f}_{E,F}$ imply

$$\boxed{\begin{aligned} [\cdot]_F &= (\mathbf{I}_V)_{E,F} \circ [\cdot]_E \\ [\mathbf{v}]_F &= [\mathbf{E}]_F [\mathbf{v}]_E \end{aligned}}$$

The second line of the boxed equation is obtained from the first by using the fact that the primitive matrix of $(\mathbf{I}_V)_{E,F}$ relative to E is $[\mathbf{I}_V(E)]_F = [\mathbf{E}]_F$.

Theorem 2.52. (Primitive matrix of $[\cdot]_F$ relative to E, primitive matrix of $[\cdot]_F^{-1}$ relative to $\mathfrak{E}$).

Let V be a finite-dimensional vector space over a field K with bases E and F. As was shown in the previous theorem, the primitive matrix of $[\cdot]_F$ relative to E is $[\mathbf{E}]_F$.

This means the primitive matrix of $[\cdot]_F^{-1} : K^n \to V$ relative to $\mathfrak{E}$ is $[\mathbf{E}]_F^{-1} = [\mathbf{F}]_E$. (The last equality follows from Theorem 2.54).

There is also a special case worth mentioning. As was alluded to in Remark 2.43, if $\mathfrak{E}$ denotes the standard basis of K^n, then $[\cdot]_{\mathfrak{E}}$ is the identity on K^n, $[\cdot]_{\mathfrak{E}} = \mathbf{I}_{K^n}$. The primitive matrix of $[\cdot]_{\mathfrak{E}}$ relative to $\mathfrak{E}$ is then the $n \times n$ identity matrix $\mathbf{I}$ (which has ij entry δ_{ij}).

Remark 2.53. (Change of basis for vectors when $V = K^n$). Consider the context of Theorem 2.51. Let $V = K^n$, and let $F = \mathfrak{E} = \{\mathfrak{e}_1, ..., \mathfrak{e}_n\}$ be the standard basis of K^n. Then the boxed equations of Theorem 2.51 simplify to

$$\mathbf{c} = \mathbf{E}[\mathbf{c}]_E$$
$$\mathbf{F} = \mathbf{E}[\mathbf{F}]_E.$$

Proof. These equations can be quickly proved from the definition of $[\cdot]_E$. First we prove the first equation:

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

The second one follows as as a special case of the first, since the first equation implies in particular that $\mathbf{f}_i = \mathbf{E}[\mathbf{f}_i]_E$:

$$\begin{pmatrix}\mathbf{f}_1 & \dots & \mathbf{f}_n\end{pmatrix} = \begin{pmatrix}\mathbf{E}[\mathbf{f}_1]_E & \dots & \mathbf{E}[\mathbf{f}_n]_E\end{pmatrix} = \mathbf{E}[\mathbf{F}]_E.$$

The last equality follows from the definition of matrix-matrix product (see Theorem 2.48). □

Theorem 2.54. ($\star$). Let V be a vector space. The identity function $\mathbf{I}_V : V \to V$ on V satisfies $(\mathbf{I}_V)_{E,F}^{-1} = (\mathbf{I}_V)_{F,E}$. As a corollary, we have $[\mathbf{E}]_F = [\mathbf{F}]_E$.

Proof. Given any bases E, F of V, Theorem 2.45 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. □

Theorem 2.55. (Change of basis for linear functions). Let V and W be finite-dimensional vector spaces. Let E, F be bases of V, let G, H be bases of W, and consider a linear function $\mathbf{f} : V \to W$. Then $\mathbf{f}_{E,F}$ and $\mathbf{f}_{G,H}$ are related by

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

This is because $\mathbf{f}_{E,F}$ was defined as $\mathbf{f}_{E,F} := [\cdot]_F \circ \mathbf{f} \circ [\cdot]_E^{-1}$. (See Theorem 2.45). But $[\cdot]_H \circ [\cdot]_{F^{-1}} = (\mathbf{I}_W)_{F,H}$ and $[\cdot]_E \circ [\cdot]_{G^{-1}} = (\mathbf{I}_V)_{G,F}$, so

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

We now translate the above equation into a statement about primitive matrices. Since the primitive matrix of a composition of functions is the product of matrices taken relative to the appropriate bases (see Theorem 2.48), 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 Theorem 2.52.

Theorem 2.56. (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 $G = H$ and $E = F$, we have

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

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

Theorem 2.57. ($\star$ Change of basis in terms of basis vectors $\star$).

Let V be a vector space with bases E and F. We now derive an alternative way to change the coordinates of a vector (Theorem 2.51). Theorem 2.51 showed that $[\mathbf{v}]_F = [\mathbf{E}]_F[\mathbf{v}]_E$ for any bases E, F of V. Similarly to as in Theorem 2.54, we can swap E and F to obtain $[\mathbf{v}]_E = [\mathbf{F}]_E[\mathbf{v}]_F$. Then in particular we have $[\mathbf{f}_i]_E = [\mathbf{F}]_E[\mathbf{f}_i]_F = [\mathbf{F}]_E\mathfrak{f}_i = ([\mathbf{F}]_E)_i$, where $([\mathbf{F}]_E)_i$ is the ith column of $[\mathbf{F}]_E$. So

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

Remark 2.58. (On the order of proving change of basis theorems). Most linear algebra texts first prove the previous theorem and use it to show a version of the first equation in the box of Theorem 2.51. This approach for proving Theorem 2.51 was not used because it involves quite a bit more matrix algebra than the approach supplied in this text. However, it good to know that these theorems are equivalent.

2.3 The dot product and cross product

2.3.1 Orthogonal linear functions

Definition 2.59. (Length of a vector). Let V be an n-dimensional vector space, and let U be any orthonormal basis for V. In analogy to the Pythagorean theorem, we define the *length* of a vector $\mathbf{v} \in V$ to be $||\mathbf{v}|| := \sqrt{\sum_{i=1}^{n}([\mathbf{v}]_U)_i^2}$.

Definition 2.60. (Orthogonal linear function). Let V be a vector space over a field K. A linear function $\mathbf{f} : V \to V$ is said to be *orthogonal* iff $\mathbf{f}$ is length-preserving, i.e., iff for all $\mathbf{v} \in V$, we have $||\mathbf{v}|| = ||\mathbf{f}(\mathbf{v})||$.

Remark 2.61. We don't know yet that the length of a vector doesn't depend on the orthonormal basis chosen for V.

Theorem 2.62. Let V be a vector space and consider a linear function $\mathbf{f} : V \to V$.

The following conditions on $\mathbf{f}$ are equivalent:

1. $\mathbf{f}$ preserves length.
2. $\mathbf{f}$ preserves dot product.
3. $\mathbf{f}$ preserves length and angle.

Proof. Blah blah blah □

Theorem 2.63. The length of a vector is basis-independent.

Proof. Any orthonormal basis U' can be obtained from any other orthonormal basis U by applying an orthogonal linear function to every basis vector in U. We just saw that orthogonal linear functions preserve length, so the length of a vector must be the same in any orthonormal basis. □

2.3.2 The dot product

Definition 2.64. (Unit vector hat notation). Let V be a finite-dimensional vector space. We $\wedge : V \to V$ to be the function $\wedge(\mathbf{v}) = \frac{\mathbf{v}}{||\mathbf{v}||}$. We denote $\wedge(\mathbf{v}) := \hat{\mathbf{v}}$.

Definition 2.65. ($\perp$ 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 2.66. When we consider $\wedge : \mathbb{R}^2 \to \mathbb{R}^2$, we have $\perp \circ \wedge = \wedge \circ \perp$, so writing $\hat{\mathbf{v}}_\perp$ is unambiguous.

Definition 2.67. (Vector projection). Let V be a vector space over K, and consider vectors $\mathbf{v}, \mathbf{w}, \mathbf{w}_\perp \in V$.

The *vector projection* of $\mathbf{v}$ onto $\mathbf{w}$ is the unique vector $\text{proj}(\mathbf{v} \to \mathbf{w}) := v_{||}\hat{\mathbf{w}}$ such that $\mathbf{v} = v_{||}\hat{\mathbf{w}} + v_\perp \hat{\mathbf{w}}_\perp$, where $v_{||}, v_\perp \in K$.

Remark 2.68. Note that $\text{proj}(\mathbf{v} \to \mathbf{w}) = \text{proj}(\mathbf{v} \to \hat{\mathbf{w}})$ because $\hat{\hat{\mathbf{w}}} = \hat{\mathbf{w}}$.

Lemma 2.69. (Length of projection is rotation-invariant). $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. If $\mathbf{f}$ is a rotation, then $||\mathbf{f}(\text{proj}(\mathbf{v} \to \mathbf{w})|| = ||\text{proj}(\mathbf{f}(\mathbf{v}) \to \mathbf{f}(\mathbf{w}))||$.

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

Lemma 2.70. (Polar representation of a vector). Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. Then $\mathbf{v} = v_{||}\hat{\mathbf{w}} + v_\perp \hat{\mathbf{w}}_\perp$, where $v_{||} = \cos(\theta)$ and $v_\perp = \sin(\theta)$, and where θ is the signed counterclockwise angle from $\mathbf{w}$ to $\mathbf{v}$.

In particular, the length of the projection of $\mathbf{v}$ onto $\mathbf{w}$ is $||\text{proj}(\mathbf{v} \to \mathbf{w})|| = ||\mathbf{v}|| \cos(\theta)$.

Remark 2.71. The length of the projection of $\mathbf{v}$ onto $\mathbf{w}$ is also $||\text{proj}(\mathbf{v} \to \mathbf{w})|| = ||\mathbf{v}|| \cos(|\theta|)$. This means that $||\text{proj}(\mathbf{v} \to \mathbf{w})||$ depends only on the angle from $\mathbf{w}$ to $\mathbf{v}$: we need not specify "signed" or "counterclockwise."

Proof. We show the "in particular" part. To show $v_\perp = \sin(\theta)$, use the fact that $v_\perp = \text{proj}(\mathbf{v} \to \mathbf{w}_\perp)$.

The lemma holds in the special case when $\mathbf{w} = \mathfrak{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; draw a right triangle to see this. For the general case, consider the rotation $\mathbf{f}$ that satisfies $\mathbf{f}(\hat{\mathbf{w}}) = \mathfrak{e}_1$, that is, $\mathbf{f}(\mathbf{w}) = ||\mathbf{w}||\mathfrak{e}_1$. Then because rotations are length-preserving and with use of the previous lemma, $v_{||} = ||\text{proj}(\mathbf{v} \to \mathbf{w})|| = ||\mathbf{f}(\text{proj}(\mathbf{v} \to \mathbf{w}))|| = ||\text{proj}(\mathbf{f}(\mathbf{v}) \to \mathbf{f}(\mathbf{w}))||$. This is the same as $||\text{proj}(\mathbf{f}(\mathbf{v}) \to ||\mathbf{w}||\mathfrak{e}_1)|| = ||\text{proj}(\mathbf{f}(\mathbf{v}) \to \mathfrak{e}_1)|| = ||\mathbf{f}(\mathbf{v})|| \cos(\phi)$, where ϕ is the signed counterclockwise angle from $\mathfrak{e}_1$ to $\mathbf{f}(\mathbf{v})$. We have $||\mathbf{f}(\mathbf{v})|| = \mathbf{v}$ because rotations are length-preserving, and $\phi = \theta$, where θ is the signed counterclockwise angle from $\mathbf{w}$ to $\mathbf{v}$, because length-preserving linear functions also preserve angle (see Theorem 2.62). Therefore $||\text{proj}(\mathbf{v} \to \mathbf{w})|| = ||\mathbf{v}|| \cos(\theta)$. □

Definition 2.72. (Geometric dot product on $\mathbb{R}^2$). Let U be an orthonormal basis for $\mathbb{R}^2$, where $\mathbb{R}^2$ is considered as a vector space over $\mathbb{R}$. The *geometric dot product on* $\mathbb{R}^2$ is the function $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined by $\mathbf{v} \cdot \mathbf{w} := ||\mathbf{w}||\text{proj}(\mathbf{v} \to \mathbf{w})$.

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

Lemma 2.73. Projection onto a vector is a linear function.

Proof. Define $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}$ by $\mathbf{f}(\mathbf{v}) = \text{proj}(\mathbf{v} \to \mathbf{w})$. We show $\mathbf{f}(\mathbf{v}+\mathbf{w}) = \mathbf{f}(\mathbf{v})+\mathbf{f}(\mathbf{w})$ and $\mathbf{f}(c\mathbf{v}) = c\mathbf{f}(\mathbf{v})$.

$$\begin{aligned}
\mathbf{f}(\mathbf{v}+\mathbf{w}) &= \mathbf{f}(v_{||}\hat{\mathbf{w}} + v_\perp \hat{\mathbf{w}}_\perp + w_{||}\hat{\mathbf{w}} + w_\perp \hat{\mathbf{w}}_\perp) = \mathbf{f}((v_{||} + w_{||})\hat{\mathbf{w}} + (v_\perp + u_\perp)\hat{\mathbf{w}}_\perp) \\
&= \text{proj}\Big(\Big((v_{||} + w_{||})\hat{\mathbf{w}} + (v_\perp + u_\perp)\hat{\mathbf{w}}_\perp\Big) \to \mathbf{w}\Big) \\
&= (v_{||} + w_{||})\hat{\mathbf{w}} = v_{||}\hat{\mathbf{w}} + w_{||}\hat{\mathbf{w}} = \text{proj}(\mathbf{v} \to \mathbf{w}) + \text{proj}(\mathbf{w} \to \mathbf{w}) = \mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{w}).
\end{aligned}$$

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

□

Theorem 2.74. The geometric dot product on $\mathbb{R}^2$ is a bilinear form. 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 $\mathbf{f}(\mathbf{v}) = \mathbf{v} \cdot \mathbf{w}$ is a linear function. Well, $\mathbf{f}(\mathbf{v}) = ||\mathbf{w}||\text{proj}(\mathbf{v} \to \mathbf{w})$, where $\text{proj}(\mathbf{v} \to \mathbf{w})$ is linear in $\mathbf{w}$. Therefore, since $\mathbf{f}$ is the result of scaling a linear function by $||\mathbf{w}||$, it too is a linear function. □

Remark 2.75. 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 $\text{proj}(\mathbf{v} \to \mathbf{w})$ is linear in $\mathbf{v}, \mathbf{w}$ when $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$.

Lemma 2.76. Consider $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ where $||\mathbf{v}|| = 1$. Then, in applying the formula $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||||\mathbf{w}|| \cos(\theta)$, we have

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

Derivation 2.77. (Algebraic dot product on $\mathbb{R}^2$). We can now derive an "algebraic" formula for the dot product, using its bilinearity (Theorem 2.74) 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

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

(This is covered in Theorem [...] of the chapter on bilinear forms).

Let $U = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ be an orthonormal basis for $\mathbb{R}^2$. The geometric dot product is a bilinear function $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, so the above implies

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

Therefore

$$\boxed{\mathbf{v} \cdot \mathbf{w} = ([\mathbf{v}]_U)_1 ([\mathbf{w}]_U)_1 + ([\mathbf{v}]_U)_2 ([\mathbf{w}]_U)_2}$$

Theorem 2.78. (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 U is an orthonormal basis of $\mathbb{R}^2$, then defining $\cdot : \mathbb{R}^2 \to \mathbb{R}^2$ by $\mathbf{v} \cdot \mathbf{w} = ([\mathbf{v}]_U)_1([\mathbf{w}]_U)_1 + ([\mathbf{v}]_U)_2([\mathbf{w}]_U)_2$ implies $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{w}||\text{proj}(\mathbf{v} \to \mathbf{w}) = ||\mathbf{v}||||\mathbf{w}|| \cos(\theta)$, where θ is the angle from $\mathbf{w}$ to $\mathbf{v}$.

Proof. Let $U = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2\}$ be an orthonormal basis for $\mathbb{R}^2$, consider $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$, and let $\mathbf{f}$ be the rotation satisfying $\mathbf{f}(\hat{\mathbf{w}}) = \hat{\mathbf{u}}_1$, that is, $\mathbf{f}(\mathbf{w}) = ||\mathbf{w}||\hat{\mathbf{u}}_1$.

We showed in Theorem 2.62 that orthogonal linear functions preserve dot product, so $\mathbf{v} \cdot \mathbf{w} = \mathbf{f}(\mathbf{v}) \cdot \mathbf{f}(\mathbf{w}) = \mathbf{f}(\mathbf{v}) \cdot ||\mathbf{w}||\hat{\mathbf{u}}_1 = \begin{pmatrix} ([\mathbf{f}(\mathbf{v})]_U)_1 \\ ([\mathbf{f}(\mathbf{v})]_U)_2 \end{pmatrix} \cdot \begin{pmatrix} ||\mathbf{w}|| \\ 0 \end{pmatrix} = ||\mathbf{w}||([\mathbf{f}(\mathbf{v})]_U)_1$. We have $([\mathbf{f}(\mathbf{v})]_U)_1 =$ $\text{proj}(\mathbf{f}(\mathbf{v}) \to \hat{\mathbf{u}}_1) = \text{proj}(\mathbf{f}(\mathbf{v}) \to \mathbf{f}(\hat{\mathbf{w}})) = \text{proj}(\mathbf{f}(\mathbf{v}) \to \mathbf{f}(\mathbf{w})) = \text{proj}(\mathbf{f}(\mathbf{v}) \to \widehat{\mathbf{f}(\mathbf{w})})$
$= \text{proj}(\mathbf{f}(\mathbf{v}) \to \mathbf{f}(\mathbf{w})) = \text{proj}(\mathbf{v} \to \mathbf{w})$, where the last equality is by Lemma 2.69.
Therefore $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{w}||\text{proj}(\mathbf{v} \to \mathbf{w})$, which is the definition of the geometric dot product on $\mathbb{R}^2$. $\square$

Remark 2.79. 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. See Theorem 2.84.

Definition 2.80. (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 U be any orthonormal basis of $\mathbb{R}^n$, and let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. We define $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by $\mathbf{v} \cdot \mathbf{w} := \sum_{i=1}^n ([\mathbf{v}]_U)_i([\mathbf{w}]_U)_i$.

Note that the dot product doesn't depend on the choice of the orthonormal basis U, because showed that orthogonal linear functions preserve dot product (see Theorem 2.62).

Theorem 2.81. (Dot product on $\mathbb{R}^n$ as matrix-matrix product). $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1^\top \mathbf{v}_2$. (Proof left as exercise).

Theorem 2.82. (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 $\mathfrak{E} = \{\mathfrak{e}\}_{i=1}^3$ is $\sqrt{\sum_{i=1}^3 ([\mathbf{v}]_\mathfrak{E})_i^2} = \sqrt{\mathbf{v} \cdot \mathbf{v}}$. Therefore, since orthogonal linear functions preserve length, the length of $\mathbf{v} \in \mathbb{R}^3$ is $\sqrt{\mathbf{v} \cdot \mathbf{v}}$, where the dot product is taken with respect to any orthonormal basis U of $\mathbb{R}^3$. This motivates us to define the length of a vector $\mathbf{v} \in \mathbb{R}^n$ to be $||\mathbf{v}|| := \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

Definition 2.83. (Angle in $\mathbb{R}^n$). The dot product on $\mathbb{R}^2$ satisfied $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||||\mathbf{w}|| \cos(\theta)$, so $\theta = \cos^{-1}\left(\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||||\mathbf{w}||}\right)$. 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 $\theta := \cos^{-1}\left(\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||||\mathbf{w}||}\right)$, where the dot product here is the dot product on $\mathbb{R}^n$.

Theorem 2.84. (Law of cosines in $\mathbb{R}^n$).

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

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

The only reason this theorem was included was to demonstrate the point of Remark 2.79.

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

Remark 2.85. The above theorem reveals that the algebraic dot product on $\mathbb{R}^2$ can also be discovered as an orthogonality condition between vectors. When $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ are orthogonal, they form a right triangle, so Pythagorean theorem gives $||\mathbf{v}||^2 + ||\mathbf{w}||^2 = ||\mathbf{v} - \mathbf{w}||^2$. Use $\mathbf{v} = \sqrt{\sum_{i=1}^2 (([\mathbf{v}]_U)_i)^2}$ to discover that we must have $([\mathbf{v}]_U)_1 + ([\mathbf{v}]_U)_2 = 0$.

Theorem 2.86. (Vector projection in terms of algebraic dot product).

Theorem 2.87. (Vector projection is bilinear).

Theorem 2.88. (The algebraic dot product is an inner product).

Theorem 2.89. (Cancelability of inner product).
Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$, and consider $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$. If $\langle \mathbf{v}_1, \mathbf{v} \rangle = \langle \mathbf{v}_2, \mathbf{v} \rangle$, then $\mathbf{v}_1 = \mathbf{v}_2$. Similarly, if $\langle \mathbf{v}, \mathbf{v}_1 \rangle = \langle \mathbf{v}, \mathbf{v}_2 \rangle$, then $\mathbf{v}_1 = \mathbf{v}_2$.

Proof. We prove the first statement; the proof of the second is analogous. Suppose $\langle \mathbf{v}_1, \mathbf{v} \rangle = \langle \mathbf{v}_2, \mathbf{v} \rangle$. Then $\langle \mathbf{v}_1 - \mathbf{v}_2, \mathbf{v} \rangle = 0$. Since $\mathbf{v}$ is arbitrary, we can choose $\mathbf{v} \neq 0$; this forces $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$ since $\langle \cdot, \cdot \rangle$ is positive-definite. Thus $\mathbf{v}_1 = \mathbf{v}_2$. □

2.3.3 The cross product

For completeness, we include a derivation of the cross product. Note that this derivation cannot be fully understood until after reading Chapter 6.

Definition 2.90. (Cross product). The *cross product* of vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ is the unique vector $(\mathbf{v}_1 \times \mathbf{v}_2) \in \mathbb{R}^3$ for which

$$(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{x} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^3.$$

Proof. Let $\mathbf{x} \in \mathbb{R}^3$. We need to show that the vector $\mathbf{c} \in \mathbb{R}^3$ for which $\mathbf{c} \cdot \mathbf{x} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{c})$ is unique. So, let $\mathbf{d} \in R^3$ satisfy this relation as well; suppose $\mathbf{d} \cdot \mathbf{x} = \det(\mathbf{v}_1, \mathbf{v}_2, \mathbf{c})$. Then $\mathbf{c} \cdot \mathbf{x} = \mathbf{d} \cdot \mathbf{x}$. Since $\cdot$ is an inner product, then $\mathbf{c} = \mathbf{d}$. (See Theorem 2.89). □

Theorem 2.91. (Magnitude, direction of cross product).
right hand rule
sin(theta) formula

2.4 Symmetric and orthogonal linear functions

This section is presented for completeness. It is entirely optional, since the material it covers is not necessary to understand tensors or differential forms. If you do choose to read this section, note that you should do so after reading chapters 3 through 4. Technically, this section only depends on Ch. **??** and Ch. 4, but these chapters are written in the context of the other first few chapters, so it is best to just read chapters 3 through 4.

Let V be a vector space over $\mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle$, with $\mathbf{f} : V \to V$ linear and $\mathbf{f}^* : V \to V$ the transpose after identifying V with V^*. From before, we had $\langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.

Definition 2.92. (Definition, conditions for a symmetric linear function).
$\mathbf{f}$ is defined 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. If V is finite-dimensional and U is an orthonormal basis of V, then the matrix $[\mathbf{f}(U)]_U$ of $\mathbf{f}$ relative to U and U is a *symmetric matrix*. That is, it is its own transpose: $[\mathbf{f}(U)]_U = [\mathbf{f}(U)]_U^\top$.

Proof.
($1 \iff 2$).

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

($\impliedby$). We have $\langle \mathbf{f}(\mathbf{v}_1), \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{f}(\mathbf{v}_2) \rangle$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $\langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle = \langle \mathbf{v}_2, \mathbf{f}(\mathbf{v}_1) \rangle$. Therefore $\langle \mathbf{v}_1, \mathbf{f}(\mathbf{v}_2) \rangle = \langle \mathbf{v}_1, \mathbf{f}^*(\mathbf{v}_2) \rangle$. Due to the cancelability of inner products (see Theorem 2.89), we have $\mathbf{f}(\mathbf{v}_2) = \mathbf{f}^*(\mathbf{v}_2)$ for all $\mathbf{v}_2 \in V$. So $\mathbf{f} = \mathbf{f}^*$.

($2 \iff 3$).

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

($\impliedby$). Let $U = \{\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_n\}$ be an orthonormal basis for V, and let the matrix $\mathbf{A}$ of $\mathbf{f}$ relative to U and 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. □

Theorem 2.93. (Conditions for an orthogonal linear function).

$\mathbf{f}$ is *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. $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{f}(\mathbf{v}_1), \mathbf{f}(\mathbf{v}_2) \rangle$. ($\mathbf{f}$ preserves inner products).
4. If V is finite-dimensional and U is an orthonormal basis of V, then the matrix $[\mathbf{f}(U)]_U$ of $\mathbf{f}$ relative to U and U has orthonormal columns.
5. If V is finite-dimensional and U is an orthonormal basis of V, then $[\mathbf{f}(U)]_U^\top [\mathbf{f}(U)]_U = \mathbf{I}$ and $[\mathbf{f}(U)]_U [\mathbf{f}(U)]_U^\top = \mathbf{I}$, so $[\mathbf{f}(U)]_U^{-1} = [\mathbf{f}(U)]_U^\top$.

Proof.
($1 \iff 2$).

($\implies$). 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 $\mathbf{v}_1, \mathbf{v}_3 \in V$.

($3 \iff 4$).

($\implies$). 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 U is orthonormal, $\langle \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j \rangle = \delta_{ij}$. Therefore $\langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \delta_{ij}$, so the columns $[\mathbf{f}(\hat{\mathbf{u}}_i)]_U$ of the matrix of $\mathbf{f}$ relative to U and U are orthonormal.

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

($4 \iff 5$).

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

($\impliedby$). Reversing the logic of the forward direction, we know $\langle \mathbf{f}(\hat{\mathbf{u}}_i), \mathbf{f}(\hat{\mathbf{u}}_j) \rangle = \delta_{ij}$. Therefore (3) is satisfied. Then we use (3) $\implies$ (4) $\implies$ (5).

($5 \implies 1$). Use the fact that $[\mathbf{f}(U)]_U^\top$ is the matrix of $\mathbf{f}^* : V \to V$ (after identifying $V \cong V^*$) with respect to U. □

Part I

Multilinear algebra and tensors

3

A motivated introduction to tensors

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

One of the ideas is that of a “multilinear element”. Recall that elements of vector spaces (vectors) can be thought of as “linear elements” because linear functions respect the decomposition of vectors. After defining the notion of *multilinear function*, we will see there is a corresponding notion of “multilinear element”, and that these 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”.

3.1 Multilinear functions and tensor product spaces

Definition 3.1. (Multilinear function).

Let $V_1, ..., V_k, W$ be vector spaces over a field K. We say a function $\mathbf{f} : V_1 \times ... \times V_k \to W$ is 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 3.2. (Examples of multilinear functions). The dot product is a bilinear function on $\mathbb{R} \times \mathbb{R}$. If you have encountered the determinant before, you might recall that it is a multilinear function.

Definition 3.3. (Vector space of multilinear functions).

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

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

Definition 3.4. (Tensor product space).

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

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

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 3.5. (Tensor terminology). Some authors use the word "tensor" to mean "$\binom{p}{q}$ tensor". (We have not defined $\binom{p}{q}$ tensors yet, but we will in Definition 3.25). We will use the word "tensor" to either mean an element of a tensor product space or a $\binom{p}{q}$ tensor, but we only do this when the meaning is clear from context.

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

Theorem 3.7. (Associativity of tensor product).

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

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

That is, these spaces are "the same", since an element of one can "naturally" be identified as an element of the other. (See Definition 2.33 for a discussion of linear isomorphisms). These identifications are "natural" in the sense that they do not depend on a choice of basis (see Definition 2.34).

Proof. Since an isomorphism of vector spaces is a linear map, it is enough to define an isomorphism on elementary tensors and "extend with linearity". To construct these isomorphisms, we will recall the definition of a tensor product space as a quotient space, so that elementary tensors of $(V_1 \otimes V_2) \otimes V_3$ are of the form $((\mathbf{v}_1, \mathbf{v}_2), \mathbf{v}_3)$, elementary tensors of $V_1 \otimes V_2 \otimes V_3$ are of the form $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$, and elementary tensors of $V_1 \otimes (V_2 \otimes V_3)$ are of the form $(\mathbf{v}_1, (\mathbf{v}_2, \mathbf{v}_3))$. For the first isomorphism, we send $((\mathbf{v}_1, \mathbf{v}_2), \mathbf{v}_3) \mapsto (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$, and for (reverse of) the second isomorphism, we send $(\mathbf{v}_1, (\mathbf{v}_2, \mathbf{v}_3)) \mapsto (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. We leave it to the reader to check that these maps are indeed linear and injective; surjectivity follows from fact that these maps "extend with linearity". When extending with linearity, it will be necessary to use the fact that $\otimes$ (that is, the outermost comma) appears to be a multilinear function. $\square$

Theorem 3.8. (Basis for a tensor product space).

Theorem 3.9. (Dimension of a tensor product of finite-dimensional vector spaces).

Theorem 3.10. (Universal property of the tensor product).

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

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

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

Theorem 3.11. (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 3.10. We already know this map is an injection because $\mathbf{h}$ is uniquely determined by $\mathbf{f}$ (see the proof of Theorem 3.10). It is a surjection because, given $\mathbf{h}$, we can choose $\mathbf{f}$ so that $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ (this is the condition that $\mathbf{h}$ uniquely satisfies); choose $\mathbf{f}$ so that $\mathbf{f}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{h}(\mathbf{v}_1 \otimes \mathbf{v}_2)$. It is linear because, given vector spaces Y, Z, W, the map $\circ$ which composes linear functions, $\circ : \mathcal{L}(Y \to Z) \times \mathcal{L}(Z \to W) \to \mathcal{L}(Y \to W)$, is a bilinear map. (Check this fact for yourself. The consequences of this are explored in [...]). Therefore the map $\mathcal{L}(V_1 \times V_2 \to W) \mapsto \mathcal{L}(V_1 \otimes V_2 \to W)$ described above is a linear isomorphism. □

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

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

The theorem we will discover is that when V and W are finite-dimensional vector spaces, there is a natural isomorphism $\mathcal{L}(V \to W) \cong W^* \otimes V$, 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 $\mathfrak{E} = \{\mathfrak{e}_1, ..., \mathfrak{e}_n\}$ be the standard basis of $K^n = K^{\dim(V)}$ and let $\mathfrak{F} = \{\mathfrak{f}_1, ..., \mathfrak{f}_m\}$ be the standard basis of $K^m = K^{\dim(W)}$. So, in the example, $\mathfrak{E} = \{\mathfrak{e}_1, \mathfrak{e}_2\}$ and $\mathfrak{F} = \{\mathfrak{f}_1, \mathfrak{f}_2, \mathfrak{f}_3\}$. The first "big leap" is to notice that the $m \times n$ matrix with ij entry 1 and all other entries 0 is $\mathfrak{f}_i \mathfrak{e}_j^\top$, where $\mathfrak{f}_i \mathfrak{e}_j^\top$ is the product of a $m \times 1$ matrix with a $1 \times n$ matrix (see Theorem 2.50). This means that the above 3×2 matrix can be expressed as

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

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

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

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

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

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

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

Consider one of the matrices in the weighted sum, $\mathbf{g}_i\mathbf{h}_j^\top$. Initially, we may be tempted to directly investigate the linear function represented by $\mathbf{g}_i\mathbf{h}_j^\top$. This will work, but we can be even more specific; $\mathbf{g}_i\mathbf{h}_j^\top$ is a matrix-matrix product, so it corresponds to a composition of linear functions (see Theorem 2.46 and possibly Theorem 2.48). Asking "to what linear functions do $\mathbf{g}_i$ and $\mathbf{h}_j^\top$ correspond?" will prove fruitful.

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

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

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

Overall, we have shown that the matrix with respect to bases of a linear function $\mathbf{f} : V \to W$ can be expressed as a linear combination of the (primitive (see Derivation 2.41)) 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 3.12. 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 3.13. (Dual space). Let V be a (not necessarily finite-dimensional) vector space over a field K. The *dual vector space* to V is the vector space over K, denoted V^*, consisting of the linear functions $V \to K$ under the operations of function addition and function scaling:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

Theorem 3.14. ($\mathcal{L}(V \to W) \cong W \otimes V^*$ naturally).

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

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

This isomorphism is natural because it does not depend on a choice of basis. (See Definition 2.34).

Remark 3.15. (Rank-1 linear transformations correspond to elementary tensors).

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

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

Remark 3.16. (Tensor product space as the structure behind composition).

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

Remark 3.17. (The two key ideas). Now that we have gone through the derivation, we can specifically see how the two key ideas of thinking of linear functions as vectors and "multilinear elements" have manifested.

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

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

3.3 Introduction to dual spaces

Recall that dual spaces are crucial to the concept of a $\binom{p}{q}$ 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^*.

Definition 3.18. repeat defn

Remark on the phrases "dual vector," "covector," and "linear functional."

Definition 3.19. (Induced dual basis).

If we have chosen a basis E for V, then...

dimension of dual space

Remark 3.20. ($V \cong V^*$ is *not* natural, and only happens when V is finite dimensional).

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

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

Bases aren't necessarily self-dual (orthonormal); requiring this implies a choice of basis for V

Theorem 3.22. (Orthonormal $\iff$ self dual).

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

Definition 3.24. (Dual transformation).

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

Definition 3.25. ($\binom{p}{q}$ tensor).

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

Remark 3.26. (The four-fold nature of $\binom{p}{q}$ tensors).

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

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

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

Generalized linear maps, Bonet and Wood?

Remark 3.27. "there's many defs of a tensor"

Definition 3.28. (Covariance and contravariance, coordinates of a $\binom{p}{q}$ tensor).

Upper and lower indices

Definition 3.29. (Valence and order of a tensor). The *valence* of a $\binom{p}{q}$ tensor is the tuple $\binom{p}{q}$. The *order* of a $\binom{p}{q}$ tensor is $p + q$.

Theorem 3.30. (Three fundamental natural isomorphisms for $\binom{p}{q}$ tensors).

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

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

Proof. The first line is Theorem 3.11, and the second line is Theorem 3.14. We need to prove the third line; 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. □

Theorem 3.31. (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 3.32. Let V and W be finite-dimensional vector spaces, and let E and F be bases for V and W, respectively. If (a_{ij}) is the matrix of $\mathbf{f}$ relative to E and F, then $\mathbf{f}$ coorresponds to the $\binom{1}{1}$ tensor $\sum_{ij} a_{ji}\epsilon_i \otimes \mathbf{f}_j$.

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

We have $\sum_{ij} a_{ji}\epsilon_i \otimes \mathbf{f}_j = \sum_{ij} \epsilon_i \otimes \mathbf{f}(\mathbf{e}_i)$. The explicit check that this is a linear isomorphism is left to the reader. It is enough to show linearity and surjectivity because $\mathcal{L}(V \to W) \cong W \otimes V^*$ is a finite-dimensional vector space; injectivity automatically follows due to Theorem 2.32.

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

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

Proof. We define an isomorphism $V^* \otimes W \to \mathcal{L}(V \to W)$ by $\phi \otimes \mathbf{w} \mapsto f_{\phi \otimes \mathbf{w}} : V \to W$, $f_{\phi\otimes\mathbf{w}}(\mathbf{v}_0) = \phi(\mathbf{v}_0)\mathbf{w}$. That is, $f_{\phi\otimes\mathbf{w}} = \mathbf{w}\phi$. Since $V^* \otimes W$ is finite-dimensional, it is enough to show that this map is linear and injective; surjectivity follows automatically from Theorem 2.32. □

3.4.1 Push-forward and pull-back

Push-forward of vectors, $\otimes^p f : T^p_0(V) \to T^p_0(W)$, and pull-back of dual vectors, $\otimes_q f^* : T^0_q(W) \to T^0_q(V)$. Note that the pull-back is often denoted (misleadingly) as f^*T.

4

Bilinear forms and metric tensors

5

Coordinates of tensors

5.1 To do

- fix upper/lower index conventions for entire book

5.2 Content

Theorem 5.1.

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

The second equality in the first line is obtained by applying the second line to the vector space V^*.

Remark 5.2. Once a basis E of V is fixed, it is common to use the convention of denoting $v^i = ([\mathbf{v}]_E)^i$ and $\phi_i = ([\phi]_{E^*})_i$. We will not use this convention.

Definition 5.3. $\mathbf{g} = (g_{ij})$ and $\mathbf{g}^{-1} = (g^{ij})$, where $g_{ij} := g(\mathbf{e}_i, \mathbf{e}_j)$.

Theorem 5.4.

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

Proof. There are two approaches.

1. Prove the third equation with the argument below. Then the fourth equation follows. The fourth equation implies the second and first equations.

2. Prove the first line with use of the induced bilinear form on V^*. Then the second equation follows. The second equation implies the fourth and third equations.

Proof of third equation. Using $([\phi]_{E^*})_i = \phi(\mathbf{e}_i)$ from Theorem 5.1, we have

$$([\phi_{\mathbf{v}}]_{E^*})_i = \phi_{\mathbf{v}}(\mathbf{e}_i) = g(\mathbf{v}, \mathbf{e}_i) = g(\sum_{j=1}^n ([\mathbf{v}]_E)^j \mathbf{e}_j, \mathbf{e}_i) = \sum_{j=1}^n ([\mathbf{v}]_E)^j g(\mathbf{e}_j, \mathbf{e}_i) = \sum_{j=1}^n ([\mathbf{v}]_E)^j g_{ji} = \sum_{j=1}^n ([\mathbf{v}]_E)^j g_{ij}$$

$\square$

Theorem 5.5. (Changing coordinates of $\binom{p}{q}$ tensors).

Given a $\binom{p}{q}$ tensor $\mathbf{T}$, we convert it to a $\binom{p}{q+1}$ tensor

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

Using the previous theorem, we compute $[\phi_{\mathbf{e}_{i_k}}]_{E^*} = \mathbf{g}[\mathbf{e}_{i_k}]_E = \mathbf{g}\mathfrak{e}_{i_k} = i_k$th column of $\mathbf{g}$. (Here, $\mathfrak{e}_{i_k}$ is a standard basis vector of $K^{\dim(V)}$). Therefore $\phi_{\mathbf{e}_{i_k}} = \sum_{r=1}^n ([\phi_{\mathbf{e}_{i_k}}]_{E^*})_r \epsilon^r = \sum_{r=1}^n g_{ri_k} \epsilon^r = \sum_{r=1}^n g_{i_k r} \epsilon^r$, so the result of the mapping is

$$
\begin{aligned}
&\mathbf{e}_{i_1} \otimes ... \sum_{r=1}^n \left(g_{i_k r} \epsilon^r \right) ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \\
&= \sum_{r=1}^n \left(g_{i_k r} \mathbf{e}_{i_1} \otimes ... \otimes \epsilon^{i_k - 1} \otimes \epsilon^r \otimes \epsilon^{i_k + 1} \otimes ... \otimes \mathbf{e}_{i_p} \otimes \epsilon^{j_1} \otimes ... \otimes \epsilon^{j_q} \right)
\end{aligned}
$$

This $\binom{p}{q}$ tensor is identifiable with the tensor

...

Therefore if the coordinates of T were originally $T^{i_1 ... i_p}_{j_1 ... j_q}$, then they get sent to $g_{i_k r} T^{i_1 ... i_{k-1} \ r \ i_{k+1} ... i_p}_{j_1 ... j_q}$.

Theorem 5.6.

$$
\begin{aligned}
g(\mathbf{v}, \mathbf{w}) &= [\mathbf{v}]_E^\top \mathbf{g} [\mathbf{w}]_F \\
g(\boldsymbol{\psi}, \boldsymbol{\phi}) &= ...
\end{aligned}
$$

Prove by starting with fact that $\psi_{\mathbf{v}}$ has matrix $(\mathbf{g}[\mathbf{v}]_E)^\top$, and similar fact for $\phi_{\mathbf{w}}$. Need to prove that <-

6

Exterior powers

7

Tensors in physics and engineering

- In physics/engineering, we are usually in the situation in which a basis has been chosen for V. (Since V is usually $\mathbb{R}^n$, for which we usually choose the standard basis). So there is always an isomorphism $V \cong V^*$. Two ways to see this (that are really the same): (1) choice of basis allows for sending basis vectors to dual vectors in the way you'd expect, (2) choice of basis induces the bilinear form that is the dot product.
- Einstein notation
  - Levi-Civita symbol. (It is not a tensor!)
- Example: Cauchy stress tensor
- Summarize things that have been mentioned before
  - Outer product. It is denoted by $\mathbf{v}_1 \otimes \mathbf{v}_2$.
  - Double contraction
    - Example: Hooke's law
  - Metric tensor and slanted indices
- Abstract index notation? (`https://math.stackexchange.com/questions/455478/what-is-the-practic` `rq=1`)

Part II

Differential forms

8

Review of 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.

Definition 8.1. (The derivative).

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

Definition 8.2. (Common notation for derivatives).

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

FIX ALIGNMENT

Leibniz notation

$$\frac{df}{dt} := f'$$
$$\frac{df}{dt}\Big|_{t=t_0} = f'(t_0)$$

Operator notation

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

Definition 8.3. (Derivative with respect to a function).

This definition formalizes a convention that is often used but rarely explained.

Suppose $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ and $x : V \subseteq \mathbb{R} \to \mathbb{R}$ satisfy the differentiability conditions of the chain rule (see Theorem [...]), so that $f \circ x$ is differentiable. We define $\frac{df}{dx} :? \to ?$ to be the function defined by

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

That is, $\frac{df}{dx} := f' \circ x$.

With this notation, the chain rule is

$$\frac{d(f \circ x)}{dt} = \frac{df}{dx}\frac{dx}{dt}.$$

This is more elegant than the following statement of the chain rule employing a substitution, which is often presented in standard calculus textbooks:

$$\frac{d(f \circ x)}{dt}\Big|_{t=t_0} = \frac{df(u)}{du}\frac{du(t)}{dt}, \text{ where } u = x(t_0).$$

check the above standard chain rule statement

Remark 8.4. (Letters in the denominator).

The definition $\frac{df}{dt} := f'$ from Definition 8.2 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.

Lemma 8.5. (Multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}$).

Set $\mathbf{x}(t_0) = \mathbf{x}_0$.

It can be proved that

$$\frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = \frac{\partial f(\mathbf{x})}{\partial x_1}\Big|_{\mathbf{x}=\mathbf{x}_0}\frac{dx_1}{dt}\Big|_{t=t_0} + ... + \frac{\partial f(\mathbf{x})}{\partial x_n}\Big|_{\mathbf{x}=\mathbf{x}_0}\frac{dx_n}{dt}\Big|_{t=t_0}.$$

In other words,

$$\frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}(t)}{dt}\Big|_{t=t_0}.$$

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

$$\boxed{\frac{d(f \circ \mathbf{x})}{dt} = (\nabla_{\mathbf{x}} f) \cdot \frac{d\mathbf{x}(t)}{dt}.}$$

Definition 8.6. (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 $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\frac{d\mathbf{x}}{dt}\Big|_{t=t_0} = \mathbf{v}$. We define the *directional derivative* $\frac{\partial f}{\partial \mathbf{v}}$ *of* f *in the direction of* $\mathbf{v}$ to be

$$\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{x}=\mathbf{x}_0} := \frac{d(f \circ \mathbf{x})}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}(t)}{dt}\Big|_{t=t_0} = (\nabla_{\mathbf{x}} f)|_{\mathbf{x}=\mathbf{x}_0} \cdot \mathbf{v}$$

Therefore the directional derivative is expressed as

$$\boxed{\begin{aligned}&\frac{\partial f}{\partial \mathbf{v}}\Big|_{\mathbf{x}=\mathbf{x}_0} = (\nabla_{\mathbf{x}} f)_{\mathbf{x}=\mathbf{x}_0} \cdot \mathbf{v} \\ &\frac{\partial f}{\partial \mathbf{v}} = \nabla f \cdot \mathbf{v}\end{aligned}}$$

In the second line, we interpret ∇ as the function sending $\mathbf{x} \mapsto \nabla_{\mathbf{x}}$.
Most authors denote $\frac{\partial f}{\partial \mathbf{v}}$ as $D_{\mathbf{x}} f(\mathbf{v})$ or as $Df[\mathbf{v}](\mathbf{x})$.

Remark 8.7. (Directional derivative simplifies to partial derivative). We have $\frac{\partial}{\partial \mathfrak{e}_i} = \frac{\partial}{\partial x_i}$.

Lemma 8.8. (Multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}^m$).
Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be [...sufficiently differentiable...], and set $\mathbf{x}_0 = \mathbf{x}(t_0)$.

$$\frac{d(\mathbf{f} \circ \mathbf{x})(t)}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = \begin{pmatrix} \frac{d}{dt} f_1(\mathbf{x}(t)) \\ \vdots \\ \frac{d}{dt} f_m(\mathbf{x}(t)) \end{pmatrix} = \begin{pmatrix} (\nabla_{\mathbf{x}} f_1)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}}{dt}\Big|_{t=t_0} \\ \vdots \\ (\nabla_{\mathbf{x}} f_m)|_{\mathbf{x}=\mathbf{x}_0} \cdot \frac{d\mathbf{x}}{dt}\Big|_{t=t_0} \end{pmatrix} = \begin{pmatrix} \underset{\mathbf{x}}{\nabla}(f_1) \\ \vdots \\ \underset{\mathbf{x}}{\nabla}(f_m) \end{pmatrix}\Bigg|_{\mathbf{x}=\mathbf{x}_0} \frac{d\mathbf{x}}{dt}\Big|_{t=t_0}.$$

In terms of functions, we have

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

Recall from Derivation 2.41 and Theorem 2.49 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 8.9. (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 8.3), we define the *Jacobian matrix* $\frac{d\mathbf{f}}{d\mathbf{x}}$ to be

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

Now the multivariable chain rule for differentiable functions $\mathbb{R}^n \to \mathbb{R}^m$ can be succintly stated using the Jacobian as

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

Definition 8.10. (Directional derivative of a differentiable function $\mathbb{R}^n \to \mathbb{R}^m$).

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

As was done previously, let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be the curve with $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\frac{d\mathbf{x}}{dt}\Big|_{t=t_0} = \mathbf{v}$. We define the *directional derivative* $\frac{\partial f}{\partial \mathbf{v}}$ *of* f *in the direction of* $\mathbf{v}$ to be

$$\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\Big|_{\mathbf{x}=\mathbf{x}_0} := \frac{d(\mathbf{f} \circ \mathbf{x})}{dt}\Big|_{\mathbf{x}=\mathbf{x}_0} = \frac{d\mathbf{f}}{d\mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}_0} \frac{d\mathbf{x}}{dt}\Big|_{t=t_0}$$

So this most general definition of directional derivative is expressed as

$$\boxed{\begin{aligned} &\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\Big|_{\mathbf{x}=\mathbf{x}_0} = \frac{d\mathbf{f}}{d\mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}_0} \frac{d\mathbf{x}}{dt}\Big|_{t=t_0} \\ &\frac{\partial \mathbf{f}}{\partial \mathbf{v}} = \frac{d\mathbf{f}}{d\mathbf{x}} \frac{d\mathbf{x}}{dt} \end{aligned}}$$

Remark 8.11. Linearity with respect to the $\mathbf{v}$ in the denominator.

9

Basic topology

Topology in $\mathbb{R}_{\text{std}}$

Open sets, closed sets and their characterizations

Define an *open set in* $\mathbb{R}_{\text{std}}$ to be an arbitrary union of open intervals. From this, the *interior point characterization of open sets* follows: a set $U \subseteq \mathbb{R}_{\text{std}}$ is open iff each $x \in U$ is an interior point, where we have defined $x \in A \subseteq \mathbb{R}_{\text{std}}$ to be an *interior point* of A iff $\forall x \in A\ \exists \text{open } U_x \ni x$ s.t. $U_x \subseteq A$. We define the *interior* of A to be the set of all interior points of A, $\text{int}(A) := \{x \in A \mid x \text{ is an interior point of } A\}$, so that the interior point characterization of open sets can be stated as: $U \subseteq \mathbb{R}_{\text{std}}$ is open iff $U = \text{int}(U)$. (Every set A satisfies $\text{int}(A) \subseteq A$. So $A = \text{int}(A)$ iff $A \subseteq \text{int}(A)$).

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

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

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

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

Still looking at the above proof, we can notice that we accidentally proved $x \in \text{int}(U)$ iff $x \notin (U^c)'$. So, we proved $x \in \text{int}(U)$ iff $x \in ((U^c)')^c$, which means $\text{int}(U) = ((U^c)')^c \iff \text{int}(U)^c = (U^c)'$. The one to remember is $\text{int}(U)^c = (U^c)'$.

What is most important from the above is that we have seen U is open iff $(U^c)' \subseteq U^c$. In words, a set is open iff its complement contains all of its limit points. For this reason, we define a *closed set in* $\mathbb{R}_{\text{std}}$ to be any set which is the complement of an open set, or, equivalently, any set which contains all of its limit points. The fact that C is closed iff $C' \subseteq C$ is the *limit point characterization of closed sets*.

Unions and intersections of open and closed sets

It quickly follows from the definition of an open set as an arbitrary union of open intervals that an arbitrary union of open sets is an open set. DeMorgan's laws then imply that an arbitrary intersection of closed sets is a closed set.

What about intersections of open sets- or, equivalently, by DeMorgan's laws- unions of closed sets? In $\mathbb{R}_{\text{std}}$, it is apparent that any infinite sort of union of closed sets is not necessarily closed: consider $\cup_{i=1}^{\infty}\{\frac{1}{n}\} = \{\frac{1}{n} \mid n \in \mathbb{N}\}$, which does not consider its limit point of 0. Perhaps at finite unions of closed sets must be closed? This turns out to be true in $\mathbb{R}_{\text{std}}$.

When defining the general notion of a topological space, we *require* that finite unions of closed sets be closed. We will require general topological spaces to have the same fundamental property of $\mathbb{R}_{\text{std}}$ that causes finite unions of closed sets to be closed. (This fundamental property will be that "a basis refines with interior points").

Topologies

A *topology* τ on a set X is a collection of subsets of X such that...

1. There is a *cover* $\mathcal{B}$ of X that *generates* τ.
   - A collection $\mathbb{C}$ of subsets of X is a *cover* of a set X iff $\cup_{C \in \mathbb{C}} C = X$. A set $\mathbb{C}$ *generates* a topology τ iff each $U \in \tau$ is an arbitrary union of the elements of $\mathbb{C}$; each $U \in \tau$ is $U = \cup_{\alpha \in I} C_\alpha$, $\{C_\alpha\} \subseteq \mathbb{C}$. (In the example with $\mathbb{R}_{\text{std}}$, elements of $\mathbb{C}$ were open intervals).
   - For the same reasons as with $\mathbb{R}_{\text{std}}$ ($X = \mathbb{R}$ and $\tau = \{\text{open sets} \subseteq \mathbb{R}\}$), (1) is equivalent to the interior point characterization of open sets, which is in turn equivalent to the limit point characterization of closed sets. Interior points and limit points are defined in the same way as before. This is also equivalent to the fact: arbitrary unions of open sets are open $\iff$ arbitrary intersections of closed sets are closed.
2. Finite unions of closed sets are closed $\iff$ finite intersections of open sets are open.

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

The above definition of a topology can quickly be seen to be equivalent to the most common definition. The most common definition requires τ to satisfy the following:

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

Note that the most common definition makes no mention of any sort of basis.

From covers to bases

In order to derive the definition of a *basis* for a topology, we will show: if $\mathcal{B}$ is a cover that generates τ, then finite unions of closed sets are closed iff $\mathcal{B}$ "refines with interior points". We say $\mathcal{B}$ *refines with interior points* iff $\forall B_1, B_2 \in \mathcal{B}$ $\;x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B}$ s.t. $x \in B_3$. This motivates defining a *basis* of a topology τ on a set X to be a cover of X that refines with interior points: since finite unions of closed sets are closed iff $\mathcal{B}$ refines with interior points, then *every topology is generated by some basis.* In fact, the reason that finite unions of closed sets are closed in $\mathbb{R}_{\text{std}}$ is that $\mathbb{R}_{\text{std}}$ refines with interior points.

Now we show that if there $\mathcal{B}$ is the cover that generates τ, then finite unions of closed sets are closed iff $\forall B_1, B_2 \in \mathcal{B}$ $\;x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B}$ s.t. $x \in B_3$.

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

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

So, set $V = \cap_{i=1}^n U_i$, where the U_i are open. If any U_i is empty, then their intersection is $\emptyset$, which is open, so assume no U_i is empty. We show that V is open by showing it satisfies the interior point characterization of open sets. Consider $x \in V$. Then $x \in U_i$ for all i. Each U_i is a union of basis elements, so, for each U_i, we have $x \in B_i$ for some $B_i \in \mathcal{B}$. Thus $x \in \cap_{i=1}^n B_i$. Using induction on the fact that $\mathcal{B}$ refines with interior points, there is a $B_x \in \mathcal{B}$ s.t. $x \in B_x \subseteq \cap_{i=1}^n B_i$. We have $\cap_{i=1}^n B_i = \cup_{x \in V} B_x$, so $\cap_{i=1}^n B_i$ is open.

Summary

We start by defining the elements of a *topology* τ on a set X to be generated by a cover $\mathcal{B}$ of X. (In the motivating example of $\mathbb{R}_{\text{std}}$, the elements of $\mathcal{B}$ are open intervals). In noticing the *interior point characterization of open sets*, we discover the definition of an *interior point.* The interior point characterization of open sets states that U is open iff $U = \text{int}(U)$.

Next, we rephrase the characterization of open sets in terms of complements of open sets, and discover the definitions *limit point*, *closed set*, and the theorem that is the *limit point characterization of closed sets.* The limit point characterization of closed sets states that C is closed iff $C' \subseteq C$, where C' is the set of limit points of C. A closer look at the proof of this reveals that the proof hinged upon the fact that $\text{int}(U)^c = (U^c)'$.

Lastly, we require any topology τ on a set X to be such that finite unions of closed sets are closed. (We are motivated to do so because $\mathbb{R}_{\text{std}}$ demonstrates this property). We show that if $\mathcal{B}$ is a cover of X that generates τ, then finite unions of closed sets are closed iff $\mathcal{B}$ refines with interior points. Therefore, we have the three following equivalent definitions of a topology:

τ is a *topology* on X iff any one of these equivalent statements is true:

- There is a cover $\mathcal{B}$ of X that generates τ, and finite unions of closed sets are closed.
- Arbitrary unions of open sets are open and finite intersections of open sets are open. (Open sets are elements of τ).
- There is a basis for τ.
  - $\mathcal{B}$ is a *basis* for τ iff $\mathcal{B}$ is a cover of X that refines with interior points; that is, iff $\mathcal{B}$ is a cover of X for which $\forall B_1, B_2 \in \mathcal{B} \;\; x \in B_1 \cap B_2 \implies \exists B_3 \in \mathcal{B}$ s.t. $x \in B_3$.

The definitions of and relations between open and closed sets all follow from the first condition in each definition of a topology (which is: there is a cover $\mathcal{B}$ of $X \iff$ arbitrary unions of open sets are open). The necessary and sufficient condition for finite intersections of closed sets (or finite unions of open sets) can be derived with only knowledge of open sets and the interior point characterization of open sets (though, it seems best to frame the proof of this necessary and sufficient condition as the last piece of the puzzle).

Interior and closure

Define interior as set of all interior points

Eqivalent def: largest open set contained in A

Then define closure as smallest closed set which contains A

Compare $\text{cl}(A)$ and A'; maybe investigate A' vs. $\text{cl}(A) - A'$; $A' - \text{cl}(A) \subseteq A' \cap A$?

$$\begin{aligned} x \in \text{cl}(A) &\iff \forall \text{closed } C \supseteq A \; x \in C && (9.1)\\ &\iff \forall \text{open } U \; U \cap A = \emptyset \implies x \notin U && (9.2)\\ &\iff \forall \text{open } U \ni x \; U \cap A \neq \emptyset. && (9.3) \end{aligned}$$

In line (2), use $U = C^C$. Line (3) follows from (2) because $(P \text{ and } \sim Q) \iff (Q \implies P)$.

More

- mention that open sets are also called "neighborhoods"
- define what it means for a point to "have a neighborhood"
- homeomorphisms are both open and closed maps
- product topology
- projection maps

10

Manifolds

10.1 Introduction to manifolds

Definition 10.1. (Manifold). An *n-manifold* is a topological space M that is...

- Hausdorff, or "point-separable"
- *second-countable*; that is, M has a countable basis
- *locally Euclidean of dimension n* in the sense that each point in M has a neighborhood that is homeomorphic to $\mathbb{R}^n$ (where $\mathbb{R}^n$ has the standard topology)

Example 10.2. (Graphs of continuous functions are manifolds).
Example 1.3

Definition 10.3. (Closed n-dimensional upper half-space). Consider $\mathbb{R}^n$ with the standard topology. We define the *closed n-dimensional half space* to be the topology $\mathbb{H}^n \subseteq \mathbb{R}^n$,

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

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

The point of defining $\mathbb{H}^n$ is to allow for a distinction between interior points of M and boundary points of M. To see how $\mathbb{H}^n$ facilitates this, note that the interior $\mathrm{int}(\mathbb{H}^n)$ and boundary $\partial\mathbb{H}^n$ of $\mathbb{H}^n$, in the usual topological senses of "interior" and "boundary" (see Defn [...]), are

$$\begin{aligned} \mathrm{int}(\mathbb{H}^n) &= \{(x^1, ..., x^n) \in \mathbb{R}^n \mid x^n > 0\} \\ \partial\mathbb{H}^n &= \{(x^1, ..., 0) \in \mathbb{R}^n\}. \end{aligned}$$

Now let M be an n-dimensional Hausdorff second-countable space, and suppose that for some open subset $U \subseteq M$ containing a point $\mathbf{p} \in U$, there is a homeomorphism $\mathbf{x} : U \subseteq M \to \mathbb{R}^n$, where $\mathbf{x}(\mathbf{p}) = (x^1(\mathbf{p}), ..., x^{n-1}(\mathbf{p}), 0)$ lies in $\partial\mathbb{H}^n$. The

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

Definition 10.4. (Manifold with boundary). An *n-manifold with boundary* is a topological space M that is...

- Hausdorff, or "point-separable"
- second-countable; that is, M has a countable basis

- each point of M has a neighborhood that is either homeomorphic to an open subset of $\mathbb{R}^n$ (where $\mathbb{R}^n$ has the standard topology), or to an open subset $\mathbb{H}^n$ with the subspace topology (inherited from the standard topology on $\mathbb{R}^n$)

Remark 10.5. (Topological interior and boundary vs. manifold interior and boundary).

We can obtain the *topological* interior and *topological* boundary of M by regarding M as a topological space and taking $\text{int}(M), \partial M$ in the usual topological senses of interior and boundary (see Definition [...] and Definition [...]).

In general, the topological interior and topological boundary are *not* the same as the manifold interior and manifold boundary.

To see this, we first need to remember that the topological notions of interior and boundary are only applicable when M is a subset of some other topological space X. We consider the case in which $X = M$, which is a relatively "natural" (but also trivial) choice, and the case $X = \mathbb{R}^m$ for some $m \geq n$, which is less natural, but still a good example.

When $X = M$, we have $\text{int}(M) = M$ and $\partial M =$

When $X = \mathbb{R}^m$, then because $M \subseteq X$, we have $M = \mathbb{R}^n$ where $n \leq m$. When $n = m$, then the topological and manifold notions of interior and boundary are equivalent, but when $n < m$, the topological interior is always empty.

"So for Hn as a subspace of Rn, there it is true that the topological boundary is the manifold boundary. And that's the model for all manifolds with boundary. But because of global topology, it is often not possible for those local Hns to live in a global ambient Rn, and without that we no longer have topological boundary = manifold boundary. (Consider a cylinder IxS1, it's not a subspace of R2)"

10.2 Coordinatizing manifolds

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

  The component functions are local in the sense that their domain is U, rather than all of M.

  $(U, \mathbf{x})$ is an *interior chart* iff $\mathbf{x}(U)$ is an open subset of $\mathbb{R}^n$, and is a *boundary chart* iff $\mathbf{x}(U)$ is an open subset of $\mathbb{H}^n$ that intersects the boundary of $\mathbb{H}^n$, $\mathbf{x}(U) \cap \partial\mathbb{H}^n \neq \emptyset$.

  An interior point $\mathbf{p} \in M$ is one for which there is an interior chart $(U, \mathbf{x})$ with $\mathbf{p} \in U$. The *(manifold) interior of* M is the set of interior points in M, and is denoted $\text{int}(M)$.

  A *boundary point* $\mathbf{p} \in M$ is one for which there is a boundary chart $(U, \mathbf{x})$ with $\mathbf{p} \in U$. The *(manifold) boundary of* M is set of all boundary points in M, and is denoted ∂M.

  Read how "It follows from the definition that each point $\mathbf{p} \in M$ is either an interior point or a boundary point"

- An *atlas* for M is a collection of charts $\{(U_\alpha, \mathbf{x}_\alpha)\}$ whose domains cover M, $M = \cup_\alpha U_\alpha$.

- Let M be a n-dimensional smooth manifold. $C^k(M)$ is defined to be the set of functions $M \to \mathbb{R}^n$ whose 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.

  $C^\infty(M) := \cup_{k=1}^\infty C^k(M)$. That is, $C^\infty(M)$ is the set of functions $M \to \mathbb{R}^k$ whose derivatives of orders $0, 1, ..., k$ are continuous, for all $k \in \{0, 1, ...\}$.

  A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is said to be *smooth* iff $\mathbf{f} \in C^\infty(M)$. Iff $\mathbf{f}$ is smooth, bijective and has a smooth inverse, then it is said to be a *diffeomorphism.* Note, all diffeomorphisms are homeomorphisms.

- The *coordinate representation of a function* $\mathbf{f} : M \to \mathbb{R}^k$ *relative to a chart* $(U, \mathbf{x})$ is the function $\widetilde{\mathbf{f}}_{(U,\mathbf{x})} = \mathbf{f} \circ \mathbf{x}^{-1} : \mathbb{R}^n \to \mathbb{R}^k$.

10.3 Smooth manifolds

- Let M be an n-dimensional manifold. We want to define the notion of a "smooth structure" on M. A natural way to move towards such a "smooth structure" is to define the notion of a smooth function. So, we might initially try defining smooth functions $M \to \mathbb{R}^k$ to be functions whose coordinate representations are smooth in any chart. This does not actually work, though.

  Let $\mathbf{f} : M \to \mathbb{R}^k$, and let $(U, \mathbf{x})$ and $(V, \mathbf{y})$ be charts on M. We have $\widetilde{\mathbf{f}}_{(U,\mathbf{x})} = \widetilde{\mathbf{f}}_{(V,\mathbf{y})} \circ (\mathbf{y} \circ \mathbf{x}^{-1})$, where $\mathbf{y} \circ \mathbf{x}^{-1} : \mathbf{x}(U \cap V) \to \mathbf{y}(U \cap V)$ is known as the *transition map* from $(U, \mathbf{x})$ to $(V, \mathbf{y})$. Note that because $\mathbf{x}$ and $\mathbf{y}$ are homeomorphisms, the transition map $\mathbf{y} \circ \mathbf{x}^{-1}$ is also a homeomorphism.

  The problem is that 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!). We want our "smooth structure" on M to be such that if the coordinate representation of $\mathbf{f}$ relative to some chart is smooth, then the coordinate representation of $\mathbf{f}$ relative to any chart is also smooth.

  Since smoothness *is* preserved by composing with a diffeomorphism, we set up the idea of a "smooth structure" on M by requiring all transition maps to be diffeomorphisms[1].

  A *smooth atlas* is an atlas for M in which all transition maps are diffeomorphisms. That is, $\mathcal{A}$ is a smooth atlas for M iff for all $(U, \mathbf{x}) \in \mathcal{A}$ and $(V, \mathbf{y}) \in \mathcal{A}$ the transition map $\chi \circ \mathbf{x}^{-1}$ is a diffeomorphism. A chart that is an element of a smooth atlas is called a *smooth chart*.

  "In general, there will be many possible atlases that give the 'same' smooth structure, in that they all determine the same collection of smooth functions on M."

  So that we can deal with a concrete choice of atlas, we define a *smooth structure* on M to be a maximal smooth atlas. No generality is lost when we take our atlas on M to be the maximal one, since all other atlases are contained in the maximal atlas.

- Examples of smooth manifolds
  - (Example 1.30). If $U \subseteq \mathbb{R}^n$ is an open subset and $\mathbf{f} : U \to \mathbb{R}^k$ is a smooth function, we have already observed above (Example 1.3) that the graph of $\mathbf{f}$ is a topological n-manifold in the subspace topology. Since graph($\mathbf{f}$) is covered by the single graph coordinate chart $\mathbf{x}$: graph($\mathbf{f}$) $\to U$ (the restriction of $\boldsymbol{\pi}_1$), we can put a canonical smooth structure on graph($\mathbf{f}$) by declaring the graph coordinate chart (graph($\mathbf{f}$), $\mathbf{x}$) to be a smooth chart.
  - (Example 1.32). Level sets are also smooth manifolds.

10.4 Submersions, immersions, embeddings

- Let M and N be smooth manifolds with or without boundary and let $\mathbf{f} : M \to N$ be a smooth map. The *rank of* $\mathbf{f}$ *at* $\mathbf{p} \in M$ is the rank of the linear map $d\mathbf{f}_\mathbf{p} : T_\mathbf{p}(M) \to T_{\mathbf{f}(\mathbf{p})}(N)$. Iff $\mathbf{f}$ has the same rank at every point, it is said to have *constant* rank.

  $\mathbf{f}$ is a *smooth submersion* iff its differential is surjective everywhere (equivalently, rank($\mathbf{f}$) = dim(N)), and is a *smooth immersion* iff its differential is injective everywhere (equivalently, rank($\mathbf{f}$) = dim(M)).

  An *embedding* is an injective smooth immersion that is also a homeomorphism onto its image.

[1]Note, the empty function is a homeomorphism, so this definition covers the case in which $U \cap V = \emptyset$ and the transition map is the empty function.

- (Proposition 5.2). (Images of embeddings as submanifolds). Suppose M is a smooth manifold with or without boundary, N is a smooth manifold, and $F : N \to M$ is a smooth embedding. Let $S = F(N)$. With the subspace topology, S is a topological manifold, and it has a unique smooth structure making it into an embedded submanifold of M with the property that F is a diffeomorphism onto its image.

10.5 The boundary of a manifold

-

- $\dim(\partial M) = \dim(M) - 1$. We say ∂M has a *codimension* of 1.

- If M is a smooth manifold with boundary, then the boundary ∂M of M is a smooth manifold properly embedded in M. (A manifold N is said to be *embedded* in a manifold M iff there exists an embedding, which is an injective smooth immersion, $N \to M$. An *immersion* is a smooth map between manifolds whose derivative is everywhere injective. See Lee Theorem 5.11).

- "slice charts for embedded submanifolds"

10.6 Oriented manifolds and their oriented boundaries

Inward- and outward- pointing vectors

- There exists a global smooth vector field on M whose restriction to ∂M is everywhere inward-pointing, and one whose restriction to ∂M is everywhere outward-pointing. (See Lee problem 8-4).

- Let $(U, \mathbf{x})$ be a smooth chart on ∂M with $\mathbf{p} \in U$. We classify vectors in $T_{\mathbf{p}}(M)$ as follows.

  - A vector is *inward-pointing (on ∂M)* iff it has positive-x^n component.
  - A vector is *tangent to ∂M* iff it has an x^n-component of zero.
  - A vector is *outward-pointing (on ∂M) iff it has negative-x^n component*, i.e., $\mathbf{v} \in T_{\mathbf{p}}(M)$ is outward pointing iff $-\mathbf{v}$ is inward pointing.

  - Example. Consider the unit disk D in $\mathbb{R}^2$, and pick a point $\mathbf{p}$ on the boundary. Any vector thought of as anchored at $\mathbf{p}$ is either in $T_{\mathbf{p}}(D)$, which is a line, or in one of the two half-spaces resulting from the splitting of $\mathbb{R}^2$ by the line $T_{\mathbf{p}}(\partial D)$. Both halves of $\mathbb{R}^2$ are homeomorhpic to half planes. One of these half planes contains $D - \{\mathbf{p}\}$ and the other one does not, so it makes sense to call vectors in the first half-plane "inward pointing" and vectors in the second half-plane "outward pointing."

- Let ω be an orientation form on M in the actual function sense. Since there exists a vector field $\mathbf{N}$ that is nowhere tangent to ∂M, then there is an induced *induced orientation form ω' on the boundary (due to interior multiplication*[2]) defined by $\omega'(\mathbf{v}_1, ..., \mathbf{v}_{n-1}) = \omega(\mathbf{N}(\mathbf{p}), \mathbf{v}_1, ..., \mathbf{v}_{n-1})$. (Lee Proposition 15.21). The orientation induced by ω' does not depend on the vector field $\mathbf{N}$ that is nowhere tangent to ∂M. (Lee Proposition 15.24).

  Proof. We need to show (1) that ω' is indeed an orientation form on ∂M and (2) that the orientation induced by ω' is independent of the choice of the nowhere tangent vector field $\mathbf{N}$.

[2]We could have put $\mathbf{N}(\mathbf{p})$ in any of ω's n argument slots, but we chose to use the first. This choice corresponds to the operation called *interior multiplication*, which you can read about in Lee's book. See p. 358 and Corollary 14.21 on p. 362.

1. We show $\omega' > 0$ on [tangent bundle of ∂M]; it suffices to show $\omega' \neq 0$ on [tangent bundle] because $\omega > 0$ on [tangent bundle of M].

   Let $\{\mathbf{e}_1, ..., \mathbf{e}_{n-1}\}$ be an ordered basis for $T_{\mathbf{p}}(\partial M)$. For all $\mathbf{p} \in M$, $\mathbf{N}(\mathbf{p})$ is not tangent to M, so $\{\mathbf{N}(\mathbf{p}), \mathbf{e}_1, ..., \mathbf{e}_{n-1}\}$ is a basis for $T_{\mathbf{p}}(M)$, since $T_{\mathbf{p}}(M)$ is identifiable with $\mathbb{R}^n$. ω is nonvanishing, so it cannot be zero on any ordered basis. This means ω' nonvanishing as well.

2. Let $\mathbf{N}$ and $\mathbf{N}'$ be two vector fields that are both nowhere tangent to ∂M. We need to show that the ordered bases $E = \{\mathbf{N}(\mathbf{p}), \mathbf{v}_1, ..., \mathbf{v}_{n-1}\}$ and $F = \{\mathbf{N}'(\mathbf{p}), \mathbf{v}_1, ..., \mathbf{v}_{n-1}\}$ have the same orientation. To do so, we prove that the determinant of the change of basis matrix between the two ordered bases is positive.

   $\mathbf{N}$ and $\mathbf{N}'$ are both outward-pointing, so the nth component of $\mathbf{N}(\mathbf{p})$ relative to E and the nth component of $\mathbf{N}'(\mathbf{p})$ relative to F are both negative; denote these nth components by $(\mathbf{N}(\mathbf{p}))_n$ and $(\mathbf{N}'(\mathbf{p}))_n$, respectively. Relative to the bases E, F, the change of basis matrix between E and F has a first column whose nth entry is $\frac{(\mathbf{N}'(\mathbf{p}))_n}{(\mathbf{N}(\mathbf{p}))_n}$, and for $i > 1$, the ith column of this matrix is $\mathfrak{e}_i$. The change of basis matrix is therefore upper triangular, so its determinant is the product of the diagonal entries, i.e., the determinant is $\frac{(\mathbf{N}'(\mathbf{p}))_n}{(\mathbf{N}(\mathbf{p}))_n} > 0$.

□

10.7 Tangent vectors

- Geometric tangent vectors $\mathbb{R}^n_{\mathbf{p}}$
- Geometric tangent vectors can be identified with linear maps $C^\infty(\mathbb{R}^n) \to \mathbb{R}$ (recall, $C^\infty(\mathbb{R}^n)$ is the set of smooth functions $M \to \mathbb{R}^n$) via the isomorphism sending $\mathbf{v}_{\mathbf{p}} \mapsto \frac{\partial}{\partial \mathbf{v}}\Big|_{\mathbf{p}}$.
- Both of the above are real vector spaces.
- (Propositions 3.10, 3.12). If M is an n-dimensional manifold with or without boundary, then $\dim(T_{\mathbf{p}}(M)) = n$ for all $\mathbf{p} \in M$. Since all vector spaces over $\mathbb{R}$ of the dimension are isomorphic, this justifies identifying $T_{\mathbf{p}}(M) \cong \mathbb{R}^n$.
- Tangent bundle, $T(M) := \sqcup_{\mathbf{p} \in M} T_{\mathbf{p}}(M)$.

  (Proposition 3.18). For any smooth n-manifold M; the tangent bundle $T(M)$ has a natural topology and smooth structure that make it into a $2n$-dimensional smooth manifold.

  Proof idea. Follow the special case of $M = \mathbb{R}^n$:

$$T(\mathbb{R}^n) = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} T_{\mathbf{p}}(\mathbb{R}^n) = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} \mathbb{R}^n_{\mathbf{p}} = \bigsqcup_{\mathbf{p} \in \mathbb{R}^n} (\mathbb{R}^n \times \{\mathbf{p}\}) = \mathbb{R}^n \times \mathbb{R}^n.$$

  (Figure out why last equality works).

- (Proposition 3.23). Every $\mathbf{v} \in T_{\mathbf{p}}(M)$ is the velocity of some smooth curve in M.

10.7.1 Vector bundles

Projection maps

(Proposition A.23). Let $X_1, ..., X_n$ be topological spaces, and consider the product topology $X_1 \times ... \times X_n$. The *projection onto the ith factor* is the map $\boldsymbol{\pi}_i : X_1 \times ... \times X_n \to X_i$ defined by $\boldsymbol{\pi}(\mathbf{p}_1, ..., \mathbf{p}_n) = \mathbf{p}_i$.

Vector bundles

- Let B be a topological space (thought of as the *base space*). A *vector bundle of rank k over B* is a tuple $(E, \boldsymbol{\sigma})$, where...

  - E is a topological space (thought of as the *entire space*, or *total space*).
  - $\boldsymbol{\sigma} : B \to \text{powerset}(E)$ is a map whose inverse restricted onto singletons, $\boldsymbol{\pi} = (\boldsymbol{\sigma}^{-1})|_{\cup_{\mathbf{p} \in E}\{\mathbf{p}\}}$ is a surjective continuous map satisfying...
    - For each $\mathbf{p} \in B$, the *fiber* $\boldsymbol{\sigma}(\mathbf{p})$ is a k-dimensional vector space.
    - For each $\mathbf{p} \in B$, there is a neighborhood $U \ni \mathbf{p}$ (called a *local trivialization*) for which there is a homeomorphism $\mathbf{F} : U \times \mathbb{R}^k \to \sigma(U)$ satisfying two more conditions:
      - .
      - .

  A vector bundle can be pictured as a hairbrush. The base space B is thought of as the surface of the hairbrush's handle, and the entire space E is thought of as the disjoint union of B with the protruding bristles of the brush (so the bristles are $E - B$). In this analogy, $\boldsymbol{\sigma}(\mathbf{p})$ is the set of points on a bristle above a particular point $\mathbf{p} \in B$ on the handle; it is therefore called the *fiber over* $\mathbf{p}$.

  The map $\boldsymbol{\pi}$ is called the *projection*.

  A *cross section*, or simply *section*, of E, is the image $\boldsymbol{\sigma}|_A(A)$ of some continuous restriction $\boldsymbol{\sigma}|_A$ of $\boldsymbol{\sigma}$ onto a subset $A \subseteq B$, where A is chosen so that $\boldsymbol{\sigma}|_A$ is one-to-one.

  The common convention is to state the definition of a vector bundle in terms of the surjective continuous restriction onto singletons $\boldsymbol{\pi}$.

- The tangent bundle is a vector bundle

- (Exercise 10.1). Suppose E is a smooth vector bundle over M. Show that the projection $\boldsymbol{\pi} : E \to M$ is a surjective smooth submersion. Submersions are embeddings, right?

  - (Theorem 4.26). (Local section theorem). Suppose M and N are smooth manifolds and $\pi : M \to N$ is a smooth map. Then π is a smooth submersion if and only if every point of M is in the image of a smooth local section of π.

- $T(M)$ is a vector bundle, and therefore is a smooth manifold. (?)

10.8 Vector fields

- A *vector field* is a continuous map $\mathbf{V} : M \to T(M)$. A smooth vector field is what it sounds like. The set of all smooth vector fields on M is denoted $\mathfrak{X}(M)$.

- The map $\mathbf{p} \mapsto \frac{\partial}{\partial x^i}$ determines a vector field, and is denoted $\frac{\partial}{\partial x^i}$

  - $\frac{\partial}{\partial x^i}$ and the chain rule from mathematical physics ch 11

- A *local frame* for M is an ordered n-tuple of vector fields $(\mathbf{V}_1, ..., \mathbf{V}_n)$ where each $\mathbf{V}_i : U \subseteq M \to \mathbb{R}^n$ is defined on an open subset $U \subseteq M$ and where, pointwise, the $\mathbf{V}_i$'s are linearly independent and span the tangent bundle. $(\mathbf{V}_1, ..., \mathbf{V}_n)$ is a *smooth frame* iff each V_i is smooth, and is a *global frame* iff $U = M$. A local frame is essentially a "basis" of vector fields.

10.9 Covector fields

- (dx^i) denotes coordinates in a coordinate *coframe*
- Let $S \subseteq M$ be an immersed submanifold. Given a smooth covector field $\mathbf{X}$ on M, the pullback by the inclusion map $i : S \to M$ of $\mathbf{X}$ is a smooth covector field $i^*\mathbf{X}$ on S: $(i^*\mathbf{X})_{\mathbf{x}}(\mathbf{v}) = \mathbf{X}_{\mathbf{x}}(di_{\mathbf{x}}(\mathbf{v})) = \mathbf{X}_{\mathbf{x}}(\mathbf{v})$. Therefore $i^*\mathbf{X}$ is called the restriction of $\mathbf{X}$ onto S. It is not a true restriction- really, this “restriction” has one less local coordinate than does $\mathbf{X}$. (CTRL-F: “Restricting Covector Fields to Submanifolds”)

11

Differential forms on manifolds

11.1 Integration of differential forms on manifolds

This section follows Ch. 16 of Lee.

There is no way to integrate *functions* on a manifold in a coordinate-independent way. For example, if B is the unit ball in $\mathbb{R}^n$ and $f \equiv 1$, then $\int_B f dV = \text{vol}(B)$. If, for example, we use the change of coordinates $\mathbf{x} \mapsto c\mathbf{x}$, then the integral changes to $c\text{vol}(B)$. (Or $c^n\text{vol}(B)$)?

Definition 11.1. (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 measure zero.

Theorem 11.2. (Pull-back of a differential k-form).

Let M and N be smooth manifolds, and let $\mathbf{f} : M \to N$ be a smooth map. Consider a differential k-form on N, $\omega \in \Omega^k(N)$. In Theorem [...], we showed how to pull back an element of $\Lambda^k(W)$ to obtain an element of $\Lambda^k(V)$. Since a differential form evaluated at a point yields an element of an exterior power, we can pull back differential forms by using the pull-back map on exterior powers.

We have $d\mathbf{f} : \Omega^k(M) \to \Omega^k(N)$, $d\mathbf{f}_\mathbf{x} : T_\mathbf{x}(M) \to T_\mathbf{x}(N)$, and $(d\mathbf{f}_\mathbf{x})^* : T_{\mathbf{f}(\mathbf{x})}(N)^* \to T_\mathbf{x}(M)^*$.

Any id with $(d\mathbf{f}_\mathbf{x})$ that needs to be mentioned?

The pull-back $\overline{\Omega^k}\mathbf{f}^* : \Omega^k(N) \to \Omega_k(N)$ of a differential k-form ω on N is defined by

$$\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)_\mathbf{x} = \Big(\Lambda^k(d\mathbf{f}_\mathbf{x})^*\Big)(\omega(\mathbf{f}(\mathbf{x})).$$

Repeating the argument made for the pull-back of an element of $\Lambda^k(W^*)$ (see [...]), the above becomes

$$\Big(\Lambda^k(d\mathbf{f}_\mathbf{x})^*\Big)(\omega(\mathbf{f}(\mathbf{x}))) = \omega(\mathbf{f}(\mathbf{x})) \circ d\mathbf{f}_\mathbf{x} = (d\mathbf{f}_\mathbf{x})^*(\omega(\mathbf{f}(\mathbf{x}))).$$

Therefore

$$(\overline{\Omega^k}\mathbf{f}^*)(\omega) = \omega(\mathbf{f}(\mathbf{x})) \circ d\mathbf{f}_\mathbf{x}.$$

More explicitly, $(\overline{\Omega^k}\mathbf{f}^*)(\omega)$ is defined at each point $\mathbf{x} \in M$ as

$$\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)_\mathbf{x}(\mathbf{v}_1, ..., \mathbf{v}_k) = \omega(\mathbf{f}(\mathbf{x}))(d\mathbf{f}_\mathbf{x}(\mathbf{v}_1), ..., d\mathbf{f}_\mathbf{x}(\mathbf{v}_k)).$$

Again referring back to how the definition of $\Lambda^k\mathbf{f}^*$ translates over to $\overline{\Lambda^k}\mathbf{f}^*$, we see that if $\omega(\mathbf{x}) = f(\mathbf{x})\epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^k$, then the determinant theorem [...] yields

$$\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)_\mathbf{x} = \det((d\mathbf{f}_\mathbf{x})^*) f(\mathbf{f}(\mathbf{x}))\ (d\mathbf{f}_\mathbf{x})^*(\epsilon^1) \overline{\wedge} ... \overline{\wedge} (d\mathbf{f}_\mathbf{x})^*(\epsilon^k).$$

Set $\delta^i = (d\mathbf{f_x})^*(\epsilon^i)$, i.e., $\delta^i = \epsilon^i \circ d\mathbf{f_x}$ to restate this as

$$\Big((\overline{\Omega^k}\mathbf{f}^*)(\omega)\Big)_{\mathbf{x}} = \det((d\mathbf{f_x})^*) f(\mathbf{f}(\mathbf{x}))\ \delta^1 \overline{\wedge} ... \overline{\wedge} \delta^k.$$

Therefore, using that $\det((d\mathbf{f_x})^*) = \det(d\mathbf{f})$ and suppressing dependence on $\mathbf{x}$, we have

$$\boxed{(\overline{\Omega^k}\mathbf{f}^*)(\omega) = \det(d\mathbf{f})(f \circ \mathbf{f})\ \delta^1 \overline{\wedge} ... \overline{\wedge} \delta^k.}$$

Here $d\mathbf{f}$ denotes the function which sends $\mathbf{x} \mapsto d\mathbf{f_x}$.

Derivation 11.3. (Change of variables theorem in light of the pullback of a map $M \to N$ and integral of a differential k-form).

[Follows Lee and Diff Topology]

The change of variables theorem says that if U and V are open domains of integration in either $\mathbb{R}^n$ or $\mathbb{H}^n$ and $\mathbf{f} : \text{cl}(U) \to \text{cl}(V)$ is a smooth map that restricts to a diffeomorphism $U \to V$, then, for every continuous function $f : \text{cl}(V) \to \mathbb{R}$, we have

$$\int_U f = \int_V (f \circ \mathbf{f}) |\det(d\mathbf{f})|.$$

Notice that, when $|\det(d\mathbf{f})| = \det(d\mathbf{f})$, i.e., when $\det(d\mathbf{f}) > 0$, then the integrand of the right hand side is exactly the pullback of $f\ \delta^1 \overline{\wedge} ... \overline{\wedge} \delta^k$. When $\det(d\mathbf{f}) < 0$, the integrand of the right hand side is the negation of this pullback. So, *in the case that* $\mathbf{f}$ *is orientation-preserving or orientation-reversing*, the change of variables theorem can be restated as

$$\int_U f = \begin{cases} \int_V (\overline{\Omega^k}\mathbf{f}^*)(f\ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n) & \mathbf{f} \text{ is orientation-preserving} \\ -\int_V (\overline{\Omega^k}\mathbf{f}^*)(f\ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n) & \mathbf{f} \text{ is orientation-reversing} \end{cases}.$$

(It is possible for $\mathbf{f}$ to be neither orientation-preserving nor orientation-reversing. In this case the integral of the pullback over V is likely unrelated to the integral of $\mathbf{f}$ over U).

If we define the *integral of a differential k-form* on elementary k-forms as

$$\int_U f\ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n := \int_U f,$$

then the above restatement of the change of variables theorem is further restated as

$$\int_U f\ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n = \begin{cases} \int_V (\overline{\Omega^k}\mathbf{f}^*)(f\ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n) & \mathbf{f} \text{ is orientation-preserving} \\ -\int_V (\overline{\Omega^k}\mathbf{f}^*)(f\ \epsilon^1 \overline{\wedge} ... \overline{\wedge} \epsilon^n) & \mathbf{f} \text{ is orientation-reversing} \end{cases}.$$

We define the integral of a nonelementary differential k-form by extending the definition of an elementary differential k-form with alternatingness and multilinearity. This yields a succinct final statement of the change of variables theorem:

$$\boxed{\int_U \omega = \begin{cases} \int_V (\overline{\Omega^k}\mathbf{f}^*)(\omega) & \mathbf{f} \text{ is orientation-preserving} \\ -\int_V (\overline{\Omega^k}\mathbf{f}^*)(\omega) & \mathbf{f} \text{ is orientation-reversing} \end{cases}}$$

[compactly supported?]

The notion of pullback not only simplifies the change of variables theorem. It also gives meaning to the "$dx^1 ... dx^n$" that is used as a placeholder in an integral. If we use the convention of writing the placeholder $dx^1 ... dx^n$ after the integrand, so that

$$\int_U f = \int_U f dx^1 ... dx^n,$$

then the definition of the integral of a differential form becomes

$$\int_U f \ dx_1 \overline{\wedge} ... \overline{\wedge} dx_n := \int_U f dx^1 ... dx^n.$$

In some sense, the placeholder $dx^1 ... dx^n$ is "secretly" $dx_1 \overline{\wedge} ... \overline{\wedge} dx_n$. So, while the definition technically defines the left hand side in terms of the right hand side, you might think of it as giving algebraic meaning to the old placeholder notation of the right hand side.

Geometric interpretation of linear functionals

Given a basis E for a finite-dimensional vector space V, a dual vector $\phi \in V^*$ acts on a vector by $\phi(\mathbf{v}) = [\phi]_E \cdot \mathbf{v}$. This is reminiscent of the equation for a plane.

Recall, the plane that contains the point $\mathbf{x}_0 \in V$ and has (not necessarily unit-length) normal $\mathbf{n}$ is the set $\{\mathbf{x} \mid (\mathbf{x} - \mathbf{x}_0) \cdot \mathbf{n} = 0\}$, which is the same set as $\{\mathbf{x} \mid \mathbf{x} \cdot \mathbf{n} = \mathbf{x}_0 \cdot \mathbf{n}\}$. This means that if we consider the level planes $\{\phi(\mathbf{v}) = [\phi]_E \cdot \mathbf{v} = k \mid c \in K\}$, then $[\phi]_E$ takes the role of $\mathbf{x}$, $\mathbf{v}$ takes the role of $\mathbf{n}$, and k takes the role of $\mathbf{x}_0 \cdot \mathbf{n}$ (for some $\mathbf{x}_0$).

Noticing this implies the following facts about $\{\phi(\mathbf{v}) = k \mid c \in K\}$.

1. The level planes are parallel to each other, because their normal vectors are all scalar multiples of $\mathbf{v}$.

2. $[\phi]_E$ is perpendicular to each level plane. (Consider that $[\phi]_E$ is perpendicular to the level plane corresponding to $k = 0$, since $[\phi]_E \cdot \mathbf{v} = 0$. This level plane is parallel to every other level plane, so $[\phi]_E$ is perpendicular to every level plane.)

Now, we construct a set of level planes analogous to a coordinate system. Consider the set of level planes for which $k = 0$, $k = 1$, and $k = 1 \cdot j$ for $j \in \mathbb{Z}$ (the homomorphism $j \mapsto j \cdot 1_F$ relates to ring characteristic!). Then to evaluate $h(\mathbf{v}) = [\phi]_E \cdot \mathbf{v}$, we count the number of these level planes that $[\phi]_E$ pierces.

This interpretation emphasizes thinking of dual vectors as corresponding to row-vectors, since row-vector times column-vector is the same as the dot product between the two corresponding column vectors.

Piponi visualization of diff forms

1-forms are integrated along paths.

To evaluate a linear functional on a vector, think of the linear functional as represented by a set of parallel $(n-1)$-dimensional planes, each a unit distance apart, and count how many planes the vector goes through.

To integrate a differential 1-form over a curve, think of the diff 1-form as represented by curving $(n-1)$-dimensional "leaves." Take the directed count of how many leaves the curve passes through.

A diff k-form on an n-manifold is represented by $(n-k)$-dimensional leaves.

A nonzero exterior product of a diff k form with a diff ℓ form is a $k+\ell$ form because intersecting a collection of $(n-k)$-dimensional leaves with a collection of $(n-\ell)$-dimensional leaves gives a collection of $(n-(k+\ell))$-dimensional leaves.

A differential n-form on an n-manifold is a "top" form whose leaves are points. Such a diff form is a volume density.

To integrate $dx \wedge dy = \epsilon^1 \wedge \epsilon^2$ over an area, draw vertical grid lines (corresponding to ϵ^1), draw horiztonal grid lines (coorresponding to ϵ^2),

Part III

Computational applications of differential forms

Part IV

Appendix

12

Appendix