MA661 course notes

Krit Boonsiriseth

Fall 2018

Abstract

Course notes for Hotchkiss class MA661 (Linear Algebra). Proofs in here are not guaranteed to be rigorous. The sections are split according to tests and projects.

Contents

1 Vector spaces **3**
- 1.1 Properties . . . 3
- 1.2 Subspaces . . . 4
- 1.3 Bases and dimensions . . . 6
- 1.4 $\dim U + W$ . . . 9

2 Linear transformations **10**
- 2.1 Kernels and images . . . 10
- 2.2 Matrices . . . 11
- 2.3 Matrices of transformations . . . 13
- 2.4 Isomorphisms . . . 15
- 2.5 Dimension formula . . . 15
- 2.6 Test review . . . 17

3 Geometry in $\mathbb{R}^n$ **20**
- 3.1 The dot product . . . 20
- 3.2 Orthogonality and projections . . . 21
- 3.3 Determinants . . . 23
- 3.4 Problems . . . 24

4 Linear regression project **26**
- 4.1 Introduction . . . 26
- 4.2 Theory Questions . . . 26
- 4.3 Practice Questions . . . 29

5 Row operations **32**
- 5.1 Introduction . . . 32
- 5.2 Applications . . . 34
- 5.3 Elementary matrices . . . 35

6 Eigenstuff **35**
6.1 Change of basis . . . 35
6.2 Eigenvalues and eigenvectors . . . 37
6.3 Spectral Theorem . . . 39
6.4 Properties of eigenvalues . . . 39
6.5 Test review . . . 41

7 Multilinear algebra **42**
7.1 Quadratic forms . . . 42
7.2 k-forms . . . 42

1 Vector spaces

1.1 Properties

Example 1.1. (a) These are linear combinations in $\mathbb{R}^3$:

$$3\begin{bmatrix}1\\1\\0\end{bmatrix} - \sqrt{2}\begin{bmatrix}\frac{1}{2}\\\frac{1}{3}\\\frac{1}{4}\end{bmatrix} + \sqrt{2}\begin{bmatrix}\sin 4\\\cos 7\\e^{12}\end{bmatrix}, \quad -1\begin{bmatrix}1\\1\\0\end{bmatrix} + 12\begin{bmatrix}\frac{1}{2}\\\frac{1}{3}\\\frac{1}{4}\end{bmatrix} - 0\begin{bmatrix}\sin 4\\\cos 7\\e^{12}\end{bmatrix}$$

(b) Some spans:

$$\text{span}\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right) \text{ is the } xy\text{-plane}, \quad \text{span}\left(\begin{bmatrix}1\\1\end{bmatrix}\right) \text{ is the line } x = y$$

(c) The vectors $\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\}$ are *linearly independent* because if

$$\alpha\begin{bmatrix}1\\1\end{bmatrix} + \beta\begin{bmatrix}1\\-1\end{bmatrix} = \mathbf{0}$$

then $\alpha + \beta = \alpha - \beta = 0$ which implies $\alpha = \beta = 0$.

(d) The vectors $\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}\right\}$ are not linearly independent (= are linearly dependent) because

$$-5\begin{bmatrix}1\\1\end{bmatrix} + 1\begin{bmatrix}1\\-1\end{bmatrix} + 2\begin{bmatrix}2\\3\end{bmatrix} = \mathbf{0}$$

We'll now start talking about *vector spaces*, which follow some properties of $\mathbb{R}^n$ (to be exact, properties that can be proven without relying on the properties of $\mathbb{R}$.)

Definition 1.2.
Any set V which satisfies the following axioms is called a *vector space*:

(i) $(V, +)$ is an abelian group, i.e. V has an associative and commutative addition $(+)$ with identity element $\mathbf{0}$ and additive inverses.

(ii) Elements of V can be scaled by scalars (which are members of a field F; in this class we will only consider $F = \mathbb{R}, \mathbb{C}$).

(iii) For any $\alpha, \beta \in F, \mathbf{v}, \mathbf{w} \in V$ the following properties must hold:

1. $(\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v})$
2. $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$
3. $\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}$

Example 1.3. These are all vector spaces over $\mathbb{R}$:

(a) $\mathbb{C}$, and so is $\mathbb{C}^n$.

(b) The set $\mathcal{P}_n(\mathbb{R})$ of polynomials in $\mathbb{R}[x]$ with degree at most n.

(c) The set of $m \times n$ matrices with real entries.

(d) $\text{span}(\sin x, \cos x) = \{\alpha \sin(\beta + x) \mid \alpha, \beta \in \mathbb{R}\}$.

(e) $\mathbb{R}^\infty$: the set of all real sequences

(f) $C^n(\mathbb{R})$: the set of all n-continuously-differentiable functions on $\mathbb{R}$.

(g) $C^\infty(\mathbb{R})$: analytic functions on $\mathbb{R}$.

(h) Set of step functions (= piecewise constant) on $\mathbb{R}$.

Exercise 1.4. A collection of vectors in a vector space V containing $\mathbf{0}$ is linearly dependent.

Solution. $1 \cdot \mathbf{0} + 0 \cdot \text{everything else} = \mathbf{0}$. ■

Exercise 1.5. A collection of vectors in a vector space V containing a duplicate is linearly dependent.

Solution. $1 \cdot \mathbf{v} + (-1) \cdot \mathbf{v} + 0 \cdot \text{everything else} = \mathbf{0}$. ■

Exercise 1.6. A collection $(\mathbf{v_1}, \cdots, \mathbf{v_k})$ of vectors is linearly independent *iff* any $\mathbf{w} \in \text{span}(\mathbf{v_1}, \cdots, \mathbf{v_k})$ is a unique linear combination of $(\mathbf{v_1}, \cdots, \mathbf{v_k})$.

Solution. ($\Rightarrow$) If there are two ways, subtract them to get a way to represent $\mathbf{0}$ as a nontrivial linear combination of $(\mathbf{v_1}, \cdots, \mathbf{v_k})$. ($\Leftarrow$) If $(\mathbf{v_1}, \cdots, \mathbf{v_k})$ is linearly dependent then there are > 1 ways to write $\mathbf{0}$ as a linear combination of $(\mathbf{v_1}, \cdots, \mathbf{v_k})$. ■

Exercise 1.7. A collection $(\mathbf{v_1}, \cdots, \mathbf{v_k})$ of vectors is linearly independent *iff* no $\mathbf{v_j}$ can be written as a linear combination of the others.

Solution. ($\Rightarrow$) If $\mathbf{v_j} = \sum_{i \neq j} \alpha_i \cdot \mathbf{v_i}$ then $-1 \cdot \mathbf{v_j} + \sum_{i \neq j} \alpha_i \cdot \mathbf{v_i} = \mathbf{0}$. ($\Leftarrow$) If $(\mathbf{v_1}, \cdots, \mathbf{v_k})$ is linearly dependent then there is a way to write $\sum \alpha_i \cdot \mathbf{v_i} = \mathbf{0}$ without all α_i's being zero. Pick $\alpha_j \neq 0$ and write $\mathbf{v_j} = \sum_{i \neq j} \frac{-\alpha_i}{\alpha_j} \cdot \mathbf{v_i}$. ■

1.2 Subspaces

Definition 1.8.
Let V be a vector space and $W \subseteq V$. Then, W is called a *subspace* of V if

(i) W is closed under vector addition, and

(ii) W is closed under scalar multiplication.

Example 1.9. If $\mathbf{v_1}, \ldots, \mathbf{v_k} \in V$, then $\text{span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ is a subspace of V

Solution. Let $\mathbf{w_1}, \mathbf{w_2} \in \text{span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$, so there are scalars α_i, β_i's such that $\mathbf{w_1} = \sum \alpha_i \mathbf{v_i}$ and $\mathbf{w_2} = \sum \beta_i \mathbf{v_i}$. Therefore, $\mathbf{w_1} + \mathbf{w_2} = \sum (\alpha_i + \beta_i) \mathbf{v_i} \in \text{span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ and $\gamma \mathbf{w_1} = \sum (\alpha_i \gamma) \mathbf{v_i} \in \text{span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$. ■

Exercise 1.10. Let $V = \mathbb{R}^2$ and $a, b \in \mathbb{R}$. Show that

$$W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \middle| ax + by = 0 \right\}$$

is a subspace of $\mathbb{R}^2$

Solution. Let $\mathbf{w_1} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \mathbf{w_2} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \in W$. (i) $\mathbf{w_1} + \mathbf{w_2} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}$ satisfies

$$a(x_1 + x_2) + b(y_1 + y_2) = (ax_1 + by_1) + (ax_2 + by_2) = 0$$

so $\mathbf{w_1} + \mathbf{w_2} \in W$. (ii) for any real α, $\alpha\mathbf{w_1} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \end{bmatrix}$ satisfies $a\alpha x_1 + b\alpha y_1 = \alpha(ax_1 + by_1) = 0$ so $\alpha\mathbf{w_1} \in W$. ■

Exercise 1.11. Show that

$$W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \middle| y = x + 1 \right\}$$

is *not* a subspace of $\mathbb{R}^2$

Solution. Just note that $\mathbf{w} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in W$ but $2\mathbf{w} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \notin W$. ■

Exercise 1.12. If W is a subspace of V, then $\mathbf{o} \in W$, and for any $\mathbf{w} \in W$, $-\mathbf{w} \in W$.

Solution. We have $0 \cdot \mathbf{w} = \mathbf{o} \in W$ and $(-1) \cdot \mathbf{w} = -\mathbf{w} \in W$. ■

Corollary 1.13. If $W \subseteq V$ does not contain $\mathbf{o}$, W cannot be a subspace of V.

Exercise 1.14. If U, W are subspaces of V, then

(a) $U \cap W$ is a subspace,

(b) $U \cup W$ may not be a subspace, and

(c) $U + W = \{u + w \mid u \in U, w \in W\}$ is a subspace.

Solution. (a) Let $\mathbf{v_1}, \mathbf{v_2} \in U \cap W$ and α be a scalar. Since U and W are subspaces, $\mathbf{v_1} + \mathbf{v_2}$ and $\alpha\mathbf{v_1}$ are members of both U and W, and thus are both in $U \cap W$.

(b) Choose $V = \mathbb{R}^2$, $U = \left\{ \begin{bmatrix} x \\ 0 \end{bmatrix} \in \mathbb{R}^2 \right\}$, $W = \left\{ \begin{bmatrix} 0 \\ y \end{bmatrix} \in \mathbb{R}^2 \right\}$ (which are clearly subspaces), then $\mathbf{v_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{v_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in U \cup W$ but $\mathbf{v_1} + \mathbf{v_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \notin U \cup W$.

(c) Let $\mathbf{v_1}, \mathbf{v_2} \in U + W$. Write $\mathbf{v_i} = \mathbf{u_i} + \mathbf{w_i}$ with $\mathbf{u_i} \in U$ and $\mathbf{w_i} \in W$. Now we can check that $\mathbf{v_1} + \mathbf{v_2} = (\mathbf{u_1} + \mathbf{u_2}) + (\mathbf{w_1} + \mathbf{w_2}) \in U + W$ and that for any scalar α, $\alpha\mathbf{v_1} = \alpha\mathbf{u_1} + \alpha\mathbf{w_1} \in U + W$. ■

We start with examples of how to write proofs for the previous exercise. The main idea is to just don't forget details (in particular, be more detailed than my proofs above).

Exercise 1.15. Let $V = \mathbb{R}^2$.

(a) Find a subset $X \subseteq V$ that is closed under addition but not scalar multiplication

(b) Find a subset $Y \subseteq V$ that is closed under scalar multiplication but not addition

Solution. (a) $\mathbb{Z}^2, \mathbb{Z}, \mathbb{Q}, \mathbb{Q}[\sqrt{5}], \ldots$

(b) Union of the axes: $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| xy = 0 \right\}$ ■

1.3 Bases and dimensions

Definition 1.16.
Let V be a vector space and $B = (\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n})$ a system of vectors that is linearly independent **and** spans V then B is called a *basis* of V.

Observation 1.17. We will see that every basis of a vector space has the same number of elements. This number is called the *dimension* of V, denoted by $\dim V$. Note that this is not a trivial fact; proof will come later.

Example 1.18. Standard basis of $\mathbb{R}^n$: $(\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n})$

$$\mathbf{e_i} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where 1 is in the i-th position.

Exercise 1.19. Show that $\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right)$ is a basis of $\mathbb{R}^2$.

Solution. See Exercise 2.1 (c) for linear independence and note that for any $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$,

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{a+b}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{a-b}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \blacksquare$$

We start with a discussion of Exercise 4.1. A more general example for 4.1(a) is lattices: sets of the form $L = \{m\mathbf{v} + n\mathbf{w} \mid m, n \in \mathbb{Z}\}$. Now we will tackle a famous result/foundation of Linear Algebra. **drumroll**

Theorem 1.20 (Exchange theorem)**.** *Let* $(\mathbf{v_1}, \ldots, \mathbf{v_k})$ *be linearly independent in vector space* V*, and let* $(\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_\ell})$ *span* V*. Then* $k \leqslant \ell$*.*

This has a very important corollary:

Corollary 1.21 (Dimension)**.** Let B_1 and B_2 be bases of V then $|B_1| = |B_2|$, and this value is called the *dimension* of V.

Proof. Use the Exchange theorem twice to get $|B_1| \leqslant |B_2|$ and $|B_2| \leqslant |B_1|$. $\square$

Now let's prove the Exchange theorem.

Proof (of Theorem 5.1). Start with the spanning system $(\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_\ell})$. We can then repeatedly *exchange* a $\mathbf{v_i}$ with a $\mathbf{w_j}$ as follows:

(i) add a new $\mathbf{v_i}$ in, making the system linearly dependent

(ii) write one of the terms as a linear combination of others, *and* we can force this to be a $\mathbf{w_j}$ because $\mathbf{v_i}$'s are linearly independent.

(iii) remove that $\mathbf{w_j}$, and still have the system be spanning.

If $k > \ell$ we can continue this until all $\mathbf{w_j}$'s are replaced by $\mathbf{v_i}$'s, *and* still add another v_i in, making $\mathbf{v_i}$'s linearly dependent which is the desired contradiction. ■

Exercise 1.22. Show that $\left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \middle| x + y - z = 0 \right\}$ is a subspace of $\mathbb{R}^3$ and find a basis and its dimension.

Solution. Choosing $\left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right)$ works because if $x + y - z = 0$ then

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \text{ and if } \alpha \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \mathbf{o}$$

then $\alpha = \beta = 0$. (Note that subspace-ness is implicitly shown since the set is a span of two vectors.) Therefore the dimension is 2. ■

Exercise 1.23. Let $P_2(\mathbb{R})$ be the set of polynomials of degree $\leqslant 2$. Find a basis and $\dim(P_2(\mathbb{R}))$.

Solution. Choose $(1, x, x^2)$. This spans $P_2(\mathbb{R})$ because $ax^2 + bx + c$ is equal to itself. It is also linearly independent because if $0 = ax^2 + bx + c$ (as a polynomial) but a, b, c are not all then we have a contradiction with the Fundamental Theorem of Algebra. Hence the dimension is 3. ■

A very important note: in the proof above we are treating polynomials as *functions* on $\mathbb{R}$. If we treat them as *formal objects* where $P = Q$ means every coefficient of P is equal to every coefficient of Q instead, we can just compare coefficients to get $a = b = c = 0$. However, in this class we will stick with the *function* point of view because we will consider more functions such as trigonometric and exponential later.

Theorem 1.24. *Let V be a vector space.*

(a) If $(\mathbf{v_1}, \dots, \mathbf{v_k})$ is linearly independent, then it can be extended to a basis $(\mathbf{v_1}, \dots, \mathbf{v_k}, \mathbf{w_1}, \dots, \mathbf{w_\ell})$ for some $\mathbf{w_1}, \dots, \mathbf{w_k} \in V$.

(b) If $(\mathbf{v_1}, \dots, \mathbf{v_k})$ spans V, then it contains a basis of V.

We will again follow the format of theorem statement $\to$ corollary statement $\to$ proof of corollary $\to$ proof of theorem.

Corollary 1.25. Every vector space has a basis

Proof. Let $\mathbf{o} \neq \mathbf{v} \in V$. By (a) we can extend $\mathbf{v}$ to a basis of V. □

Proof (of Theorem 6.1). (a) If $S = (\mathbf{v_1}, \dots, \mathbf{v_k})$ is a basis, we are done. If not, choose $\mathbf{w} \notin$ span S and put $\mathbf{w}$ in S. Rinse and repeat until done.

Note: this is a bit iffy when it comes to infinite-dimensional stuff–that's probably because it can involve the Axiom of Choice. (in fact a lookup on Wikipedia suggest that this is *equivalent* to the Axiom of Choice.) However, in this course we will focus on finite-dimensional vector spaces, so we will just let this slide...

(b) If $S = \text{span}(\mathbf{v_1}, \ldots, \mathbf{v_k})$ is not linearly independent, just remove one v_j that is in the span of others, and repeat until we get a linearly independent set. ■

Definition 1.26.
Let V be a vector space with basis $B = (\mathbf{b_1}, \ldots, \mathbf{b_n})$. Let $\mathbf{x} = [x_i] \in \mathbb{R}^n$. Define $\Phi_B : \mathbb{R}^n \to V$ by $\Phi_B(\mathbf{x}) = \sum x_i \mathbf{b_i}$, and this is called the *concrete-to-abstract map.*

Theorem 1.27. *Φ_B is invertible (one-to-one and onto) and Φ_B^{-1} is called the* coordinate map *of V w.r.t. B.*

Since people were confused about being why invertible is equivalent to being one-to-one and being onto, here is an aside theorem.

Theorem 1.28. *A function $f : X \to Y$ is invertible iff f is bijective. Moreover, the inverse if unique and denoted by f^{-1}.*

Proof. ($\Rightarrow$) $f(x) = f(y)$ implies $x = f(g(x)) = f(g(y)) = y$ and $x = f(g(x))$ so f is both one-to-one and onto. ($\Leftarrow$) Since f is bijective, for each $x \in X$ there is a unique $y_x \in Y$ such that $f(y_x) = x$. Define $g(x) := y_x$. (uniqueness) If g_1, g_2 are inverses of f then $g_1 = g_1 \circ f \circ g_2 = g_2$ for all $y \in Y$. ■

Let's go back to what we wanted to prove.

Proof (of Theorem 6.4). Just use the span-ness and linear independence of B: if $\Phi_B(\mathbf{x}) = \Phi_B(\mathbf{y})$ then $\sum(x_i - y_i)\mathbf{b_i} = 0$ so $\mathbf{x} = \mathbf{y}$, and if $\mathbf{v} \in V$ then there is a unique (α_i) such that $\mathbf{v} = \sum \alpha_i \mathbf{b_i}$ so $\Phi_B(\alpha) = \mathbf{v}$. ■

This explains the term *coordinate map* for Φ_B^{-1}: for given $\mathbf{v}$, it gives a representation of $\mathbf{v} \in \mathbb{R}^n$ by giving the scalars that compose it.

Example 1.29. Let $V = P_2(\mathbb{R})$ and $B = (1, x, x^2)$. Then, $\Phi_B\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = a + bx + cx^2$ and $\Phi_B^{-1}(1 + 2x + 3x^2) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Exercise 1.30. Determine whether $S = \{\mathbf{v} \in \mathbb{R}^3 \mid v_1 v_2 v_3 = 0\}$ is a subspace of $\mathbb{R}^3$

Answer: No.

Solution. It suffices to note that $\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ are in S but $\mathbf{e_1} + \mathbf{e_2} + \mathbf{e_3} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is not in S. ■

Exercise 1.31. Suppose the U is a subspace of V. What is $U + U$?

Answer: $U + U$ is always U.

Solution. Any $\mathbf{v} \in U + U$ can be written as $\mathbf{v} = \mathbf{u_1} + \mathbf{u_2}$ where $\mathbf{u_1}, \mathbf{u_2} \in U$. Since U is a subspace, it is closed under addition, so $\mathbf{v} \in U$, which means $U + U \subseteq U$. On the other hand, since U is a subspace, it contains $\mathbf{o}$, so any $\mathbf{u} \in U$ can be written as $\mathbf{u} + \mathbf{o} \in U + U$ as well. Therefore $U \subseteq U + U$, and thus $U + U = U$. ■

Exercise 1.32. Prove or give a counterexample: if U_1, U_2, W are subspaces of V such that $U_1 + W = U_2 + W$, then $U_1 = U_2$.

Solution. A counterexample for any nontrivial V is $U_1 = \{\mathbf{o}\}, U_2 = W = V$. Clearly, $U_1 + W = V = U_2 + W$ (the second equality follows from the last exercise) but $U_1 \neq U_2$. ■

1.4 $\dim U + W$

Exercise 1.33. Suppose U and W are subspaces of $\mathbb{R}^8$ such that $\dim U = 3, \dim W = 5$, and $U + W = \mathbb{R}^8$. Prove that $U \cap W = \{\mathbf{o}\}$.

Solution. Suppose that $U \cap W$ is not $\{\mathbf{o}\}$, so $d := \dim U \cap W \geqslant 1$. Let $B = (\mathbf{b_1}, \ldots, \mathbf{b_d})$ be a basis of $U \cap W$. We can extend B to a basis $B_U = (\mathbf{b_1}, \ldots, \mathbf{b_d}, \mathbf{u_1}, \ldots, \mathbf{u_{3-d}})$ of U and a basis $B_W = (\mathbf{b_1}, \ldots, \mathbf{b_d}, \mathbf{w_1}, \ldots, \mathbf{w_{5-d}})$. Now we claim

$$S = (\mathbf{b_1}, \ldots, \mathbf{b_d}, \mathbf{u_1}, \ldots, \mathbf{u_{3-d}}, \mathbf{w_1}, \ldots, \mathbf{w_{5-d}})$$

spans $U + W = \mathbb{R}^8$. This is not hard: for each $\mathbf{v} \in U + W$, just write

$$\mathbf{v} = \mathbf{u}+\mathbf{w} = \left(\sum \alpha_i \mathbf{b_i} + \sum \alpha_i' \mathbf{u_i}\right) + \left(\sum \beta_i \mathbf{b_i} + \sum \alpha_i' \mathbf{w_i}\right) = \sum (\alpha_i + \beta_i)\mathbf{b_i} + \sum \alpha_i' \mathbf{u_i} + \sum \beta_i' \mathbf{w_i}.$$

Now since the standard basis of $\mathbb{R}^8$ has 8 elements and is linearly independent, by the Exchange theorem, we have $|S| \geqslant 8$. However,

$$|S| = d + (3 - d) + (5 - d) = 8 - d$$

which is a contradiction. ■

Exercise 1.34 (Homework)**.** Suppose that U and W are both five-dimensional subspaces of $\mathbb{R}^9$. Prove that $U \cap W \neq \{\mathbf{o}\}$

Solution. Pick bases $(\mathbf{u_1}, \ldots, \mathbf{u_5})$ of U and $(\mathbf{w_1}, \ldots, \mathbf{w_5})$ of W. Since $(\mathbf{u_1}, \ldots, \mathbf{u_5}, \mathbf{w_1}, \ldots, \mathbf{w_5})$ has $10 > 9$ vectors, it must be linearly dependent in $\mathbb{R}^9$–say $\sum \alpha_i \mathbf{u_i} + \sum \beta_i \mathbf{w_i} = \mathbf{o}$. This implies $\mathbf{o} \neq \sum \alpha_i \mathbf{u_i} \in U \cap W$. ■

[This is not in class, but I just want to note] In 6.10 we can actually show that S is linearly independent as well: suppose that

$$\sum \gamma_i \mathbf{b_i} + \sum \alpha_i \mathbf{u_i} + \sum \beta_i \mathbf{w_i} = \mathbf{o}$$

with the coefficients being not all zero. This implies $\sum \beta_i \mathbf{w_i} \in U \cap W$. Hence, $\sum \beta_i \mathbf{w_i} + \sum \gamma_i' \mathbf{b_i} = \mathbf{o}$ for some choice of γ_i''s, so by linear independence of B_W we have $\beta_i \equiv 0$. Therefore we have

$$\sum \alpha_i \mathbf{u_i} + \sum \beta_i \mathbf{w_i} = \mathbf{o}$$

with the coefficients being not all zero, contradicting the fact that B_U is a basis of U. From this we can derive the following theorem:

Theorem 1.35. *If U and W are subspaces of a vector space V then*

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W).$$

This gives an easy solution to Exercises 6.10 and 11:

Solution (to Exercise 6.10). Since $\dim(U+W) = 8$, we have $\dim(U \cap W) = 3+5-8 = 0$ so $U \cap W = \{\mathbf{o}\}$.

Solution (to Exercise 6.11). Since $U + W \subseteq \mathbb{R}^9$, $\dim U + W \leqslant 9$. Therefore, $\dim(U \cap W) \geqslant 5+5-9 = 1$, so $U \cap W$ is not $\{\mathbf{o}\}$. ■

[In-class stuff follow:] In class we discussed solutions to Exercises 6.10 and 6.11, and discussed some questions people have.

Exercise 1.36. Let V be a vector space, and U, W subspaces with bases B_U and B_W respectively. Then, $U + W = \text{span}(B_U \cup B_W)$.

Solution. For each $\mathbf{x} \in V$, we have $\mathbf{x} \in U + W$ *iff* there exists $\mathbf{u_x} \in U$ and $\mathbf{w_x} \in W$ with $\mathbf{x} = \mathbf{u_x} + \mathbf{w_x}$ *iff* there exists α_i, β_i with $\mathbf{x} = \sum_{u \in B_U} \alpha_i \mathbf{u} + \sum_{w \in B_W} \beta_i \mathbf{w}$ *iff* $x \in \text{span}(B_U \cup B_W)$. ■

Exercise 1.37. If a function $f : X \to Y$ is invertible the the inverse is unique

See proof in last part of Theorem 6.5. Also, for good problems, check out *Linear Algebra Done Right* by Sheldon Axler.

2 Linear transformations

2.1 Kernels and images

Definition 2.1.
Let $T : V \to W$ be a transformation between vector spaces. Then t is called *linear* or *vector space homomorphism* if is commutative with linear combinations, that is, for all $\mathbf{x}, \mathbf{y} \in V$, $\alpha, \beta \in \mathbb{R}$,

$$T(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha T(\mathbf{x}) + \beta T(\mathbf{y}).$$

Example 2.2. The map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3x + 4y \\ x - y \end{bmatrix}$$

is a linear transformation because if $\mathbf{v} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$, $\alpha, \beta \in \mathbb{R}$ then

$$\begin{aligned} T(\alpha\mathbf{v} + \beta\mathbf{w}) &= T\left(\begin{bmatrix} \alpha x_1 + \beta x_2 \\ \alpha y_1 + \beta y_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 3(\alpha x_1 + \beta x_2) + 4(\alpha y_1 + \beta y_2) \\ \alpha x_1 + \beta x_2 - \alpha y_1 - \alpha y_2 \end{bmatrix} \\ &= \alpha \begin{bmatrix} 3x_1 + 4y_1 \\ x_1 - y_1 \end{bmatrix} + \beta \begin{bmatrix} 3x_2 + 4y_2 \\ x_2 - y_2 \end{bmatrix} \\ &= \alpha T(\mathbf{v}) + \beta T(\mathbf{w}). \end{aligned}$$

Example 2.3. Let $\mathbf{a} \in \mathbb{R}^n$ and let $T_\mathbf{a}(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$. Then $T_\mathbf{a} : \mathbb{R}^n \to \mathbb{R}$ is a linear transformation. This can be shown using commutative / distributive properties.

Definition 2.4.
Let $\mathcal{L}(V, W)$ be the set of linear transformations between V and W.

Note [not covered in class]: If I recall correctly, $\mathcal{L}(V, W)$ is itself a vector space, and when $W = \mathbb{R}$, it is the dual space $V^\vee$ of V.

Exercise 2.5. Show that if $T \in \mathcal{L}(V, W)$ then $T(\mathbf{o}_V) = \mathbf{o}_W$.

Solution. $T(\mathbf{o}_V) = T(0\mathbf{o}_V + 0\mathbf{o}_V) = 0T(\mathbf{o}_V) + 0T(\mathbf{o}_V) = \mathbf{o}_W + \mathbf{o}_W = \mathbf{o}_W$. ■

Example 2.6. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 3x + 1$ is not a linear transformation because $f(0) \neq 0$. It is called *affine linear* which means linear added by a constant.

Definition 2.7.
Let $T \in \mathcal{L}(V, W)$. Then, the *kernel* of T is defined by

$$\ker T = \{\mathbf{v} \in V | T(\mathbf{v}) = \mathbf{o}\}.$$

The *image* of T is defined as

$$\operatorname{im} T = T(V) = \{T(\mathbf{v}) | \mathbf{v} \in V\}.$$

The graph representation of T (not used) is

$$\Gamma^T = \{(\mathbf{v}, T(\mathbf{v})) | \mathbf{v} \in V\}.$$

Theorem 2.8. *(a)* $\ker T$ *is a subspace of* V

(b) $\operatorname{im} T$ *is a subspace of* W.

Proof. (a) If $\mathbf{v}, \mathbf{w} \in \ker T$ and $\alpha \in \mathbb{R}$ then $T(\mathbf{v}+\mathbf{w}) = T(\mathbf{v})+T(\mathbf{w}) = \mathbf{o}$ and $T(\alpha\mathbf{v}+0\mathbf{o}) = \alpha T(\mathbf{v}) + 0T(\mathbf{o}) = \mathbf{o}$.

(b) Pretty much the same thing. ■

Exercise 2.9. $T : \mathbb{R}^3 \to \mathbb{R}$ is defined by $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = 3x + 2y - z$. Find a basis of $\ker T$.

Solution. Let $\mathbf{b_x} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$ and $\mathbf{b_y} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$. Then, $(\mathbf{b_x}, \mathbf{b_y})$ is a basis of $\ker T$ because

$$\text{if } T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = 0 \text{ then } \begin{bmatrix} x \\ y \\ z \end{bmatrix} = x\mathbf{b_x} + y\mathbf{b_y},$$

and if $x\mathbf{b_x} + y\mathbf{b_y} = \mathbf{o}$ then comparing coefficients give $x = y = 0$. ■

2.2 Matrices

Notation 2.10.
The set of all $m \times n$ matrices with real entries is denoted by $\mathbb{R}^m_n$. If $A \in \mathbb{R}^m_n$ then individual elements of A are denoted by $A = (a_{ij})^{m,n}_{i=1,j=1}$.

Matrix operations (here $A, B \in \mathbb{R}^m_n$.)

- Addition: $A + B = [a_{ij} + b_{ij}]_{i=1,j=1}^{m,n}$.
- Transposing: $A^T = [a_{ji}]_{j=1,i=1}^{n,m}$.
- Scalar multiplication: for $\alpha \in \mathbb{R}$, $\alpha A = [\alpha a_{ij}]_{i=1,j=1}^{m,n}$.

Observation 2.11. $\mathbb{R}^m_n$ is a vector of dimension mn because of the standard basis

$$E_{ij} = (e_{k\ell})_{k=1,\ell=1}^{m,n} \text{ where } e_{k\ell} = \begin{cases} 1 & \text{if } k = i, \ell = j \\ 0 & \text{otherwise.} \end{cases}$$

- Matrix multiplication: For $A \in \mathbb{R}^m_n, B \in \mathbb{R}^n_p$,

$$AB = \left[\sum_{k=1}^{n} a_{ik} b_{kp} \right]_{i=1,j=1}^{m,p}.$$

  This can be described as multiplying each row with each column.

Notation 2.12.

- Matrices in column notation:

$$A = \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_n} \end{bmatrix}.$$

- Matrices in row notation:

$$A = \begin{bmatrix} \mathbf{a_1}^T \\ \mathbf{a_2}^T \\ \vdots \\ \mathbf{a_m}^T \end{bmatrix}.$$

  Using this notation we can also write matrix multiplication as

$$AB = \begin{bmatrix} \mathbf{a_1}^T \\ \mathbf{a_2}^T \\ \vdots \\ \mathbf{a_m}^T \end{bmatrix} \begin{bmatrix} \mathbf{b_1} & \mathbf{b_2} & \cdots & \mathbf{b_p} \end{bmatrix} = (\mathbf{a_i}^T \cdot \mathbf{b_j})_{ij}.$$

A matrix $I \in \mathbb{R}^m_n$ is called an *identity matrix* if for all $A \in \mathbb{R}^m_n$, $AI = IA = A$. This works only if $m = n$.

Convince yourself that

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{e_1} & \mathbf{e_2} & \cdots & \mathbf{e_n} \end{bmatrix} = [a_{ij}]_{ij} \text{ where } a_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

For $A \in \mathbb{R}^n_n$, $B \in \mathbb{R}^n_n$ is called an *inverse* of A if $AB = BA = I$. It can again be shown that B is unique, justifying the notation $A^{-1} = B$.

Example 2.13 (Inverse of 2×2 matrices)**.** The inverse of a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ where $ad - bc \neq 0$ is

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Definition 2.14.
A matrix $A \in \mathbb{R}_n^n$ is called

- *symmetric* if $A = A^T$,
- *antisymmetric* if $A = -A^T$, and
- *diagonal* if $a_{ij} = 0$ for all $i \neq j$.

Exercise 2.15. Show that $(AB)^T = B^T A^T$.

Solution. Using column and row notation and the fact that $(M^T)^T = M$ for any matrix M,

$$(AB)^T = [\mathbf{a_i}^T \cdot \mathbf{b_j}]_{ij}^T = [\mathbf{a_i}^T \cdot \mathbf{b_j}]_{ji}^T = [\mathbf{b_j}^T \cdot \mathbf{a_i}]_{ji} = B^T A^T. \quad \blacksquare$$

Exercise 2.16. For any matrix A, $A^T A$ is symmetric.

Solution. Let $A \in \mathbb{R}_n^m$, so $A^T \in \mathbb{R}_m^n$, so $A^T A$ is defined. Now,

$$(A^T A)^T = A^T (A^T)^T = A^T A$$

so $A^T A$ is symmetric. $\blacksquare$

Theorem 2.17. *Let $A \in \mathbb{R}_n^m$. Then A induces a transformation*

$$T_A : \mathbb{R}^n \to \mathbb{R}^m$$

by $T_A(\mathbf{v}) = A\mathbf{v}$ and $T_A \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$.

Proof. Let $\alpha, \beta \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. We have

$$T_A(\alpha\mathbf{v} + \beta\mathbf{w}) = A(\alpha\mathbf{v} + \beta\mathbf{w}) = A(\alpha\mathbf{v}) + A(\beta\mathbf{w}) = \alpha(A\mathbf{v}) + \beta(A\mathbf{w}) = \alpha T_A(\mathbf{v}) + \beta T_A(\mathbf{w})$$

so T_A is a linear transformation. $\blacksquare$

2.3 Matrices of transformations

Theorem 2.18. *For each transformation $T \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$, there exists a matrix $[T] \in \mathbb{R}_n^m$ such that*

$$[T]\mathbf{v} = T(\mathbf{v})$$

for all $\mathbf{v} \in \mathbb{R}^m$. $[T]$ is called the matrix of T w.r.t. the standard basis.

Proof. We claim that

$$[T] = \begin{bmatrix} T(\mathbf{e_1}) & T(\mathbf{e_2}) & \cdots & T(\mathbf{e_n}) \end{bmatrix}$$

works. Let $\mathbf{v} \in \mathbb{R}^m$. There is a unique way of writing $\mathbf{v}$ as $\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{e_i}$. Then,

$$T(\mathbf{v}) = \sum_{i=1}^n \alpha_i T(\mathbf{e_i}) = [T]\mathbf{v}. \blacksquare$$

Exercise 2.19 (1.3.4 in book). (a) Let T be a linear transformation such that $T \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 2v_1 \\ v_2 \\ v_3 \end{bmatrix}$. What is its matrix?

(b) Repeat part (a) for $T\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix} = \begin{bmatrix}v_2\\v_1+2v_2\\v_3+v_1\end{bmatrix}$.

Answer: Just use theorem 10.3.

(a) $[T] = \begin{bmatrix}2&0&0\\0&1&0\\0&0&1\end{bmatrix}$. (b) $[T] = \begin{bmatrix}0&1&0\\1&2&0\\1&0&1\end{bmatrix}$.

Exercise 2.20. The rotation of $\mathbb{R}^2$ about the origin by angle θ is a linear transformation $\rho_\theta : \mathbb{R}^2 \to \mathbb{R}^2$. Find $[\rho_\theta]$.

Solution. $\rho_\theta\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}\cos\theta\\\sin\theta\end{bmatrix}$ and $\rho_\theta\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}-\sin\theta\\\cos\theta\end{bmatrix}$, so $[\rho_\theta] = \begin{bmatrix}\cos\theta & -\sin\theta\\\sin\theta & \cos\theta\end{bmatrix}$. ■

Exercise 2.21. Let $R : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection over the line $y = x$. Find $[R]$.

Solution. $R\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix}$ and $R\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}$, so $[R] = \begin{bmatrix}0&1\\1&0\end{bmatrix}$. ■

Exercise 2.22 (1.3.10 in book)**.** Is there a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ such that

$$T\begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}3\\0\\1\end{bmatrix},\ T\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}4\\2\\4\end{bmatrix},\ T\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}2\\3\\3\end{bmatrix}?$$

If so, what is its matrix?

Solution. We claim that the linear transformation defined by the matrix

$$[T] = \begin{bmatrix}3&1&-2\\0&2&1\\1&3&-1\end{bmatrix}$$

works. This is easy to verify and easy to write by hand, but hard to typeset so I'll skip that. ■

Note: Since the three vectors given as arguments of T are linearly independent, we know that the answer is yes (without the need to show a $[T]$.) To find a specific T, we need to solve for $T(\mathbf{e_i})$.

Theorem 2.23. *Let $T \in \mathcal{L}(V, W)$. Then T is defined uniquely by its values on a basis.*

Proof. Let $B = (\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_n})$ be a basis and $\mathbf{v} \in V$. There is a unique way to write $\mathbf{v} = \sum \alpha_i \mathbf{b_i}$. Then,

$$T(\mathbf{v}) = T\left(\sum \alpha_i \mathbf{b_i}\right) = \sum \alpha_i T(\mathbf{b_i}). \quad ■$$

Corollary 2.24. Let $T : V \to W$ and B a basis of V. Then there exists exactly one $S \in \mathcal{L}(V, W)$ such that $S|_B = T|_B$ (that is, S assumes the same values as T on the basis.)

Definition 2.25.
If $B = (\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_n})$ is a basis of $\mathbb{R}^n$ and $T \in \mathcal{L}(\mathbb{R}^n)$, then

$$[T]_B = \begin{bmatrix}T(\mathbf{b_1}) & T(\mathbf{b_2}) & \cdots & T(\mathbf{b_n})\end{bmatrix}$$

is the *matrix of T with respect to B*. It has the property that

$$T(\mathbf{v})_B = [T]_B[\mathbf{v}]_B.$$

Example 2.26. From the previous exercise the matrix $X = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 2 & 3 \\ 1 & 4 & 3 \end{bmatrix}$ sends an input w.r.t. B to an output w.r.t. standard matrix. This is not $[T]_B$. To get $[T]_B$, we need to write the columns in terms of B as well. Anyway, we have $\mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}_B$ so $X[\mathbf{v}]_B = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

2.4 Isomorphisms

Definition 2.27.
Let $T \in \mathcal{L}(V, W)$. Then T is called an *isomorphism* if T is a bijection. V is *isomorphic* to W, denoted $V \cong W$ if there exists an isomorphism between them.

Loosely speaking, isomorphic vector spaces are algebraically indistinguishable: they only differ by what their vectors represent.

Example 2.28. As vector spaces, $\mathbb{C} \cong \mathbb{R}^2$ with the isomorphism $a + bi \mapsto (a, b)$. More rigorously:

(i) define: Let $\varphi : \mathbb{C} \to \mathbb{R}^2$ be defined by $\varphi(z) = \begin{bmatrix} x \\ y \end{bmatrix}$ if $z = x + iy$.

(ii) linear: Then, if $z = x + iy, w = u + iv \in \mathbb{C}, \alpha, \beta \in \mathbb{R}$, then $\varphi(\alpha z + \beta w) = \varphi(\alpha x + \beta u + i(\alpha y + \beta v)) = \begin{bmatrix} \alpha x + \beta u \\ \alpha y + \beta v \end{bmatrix} = \alpha\varphi(z) + \beta\varphi(w)$.

(iii) 1-1: If $\varphi(z) = \varphi(w)$ then $\Re(z) = \Re(w)$ and $\Im(z) = \Im(w)$, so $z = w$.

(iv) onto: For any $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ we have $\varphi(x + iy) = \mathbf{v}$.

Exercise 2.29. $T \in \mathcal{L}(V, W)$ is 1-1 iff $\ker T = \{\mathbf{0}\}$.

Solution. ($\Rightarrow$) If T is 1-1 then for any $\mathbf{v} \in \ker T$, $T(\mathbf{0}) = T(\mathbf{v})$ so $\mathbf{0} = \mathbf{v}$, so $\ker T = \{\mathbf{0}\}$. ($\Leftarrow$) If $\ker T = \{\mathbf{0}\}$ then for any $\mathbf{v_1}, \mathbf{v_2}$ with $T(\mathbf{v_1}) = T(\mathbf{v_2})$, $T(\mathbf{v_1} - \mathbf{v_2}) = \mathbf{0}$. Therefore $\mathbf{v_1} - \mathbf{v_2} \in \ker T$ so it is $\mathbf{0}$, hence $\mathbf{v_1} = \mathbf{v_2}$. ∎

2.5 Dimension formula

Theorem 2.30 (Dimension formula)**.** *Let $T \in \mathcal{L}(V, W)$. Then,*

$$\dim V = \dim \ker T + \dim \operatorname{im} T.$$

Proof. Let $B_1 = (\mathbf{u_1}, \cdots, \mathbf{u_k})$ be a basis of $\ker T$. Extend B_1 to a basis $B = (\mathbf{u_1}, \cdots \mathbf{u_1}, \mathbf{w_1}, \cdots, \mathbf{w}_\ell$ of V. Then, $\dim V = \dim \ker T + \ell$ so we need to show

We claim that $B_2 = (T(\mathbf{w_1}), \cdots, T(\mathbf{w}_\ell))$ is a basis of $\operatorname{im} T$. If $\sum \alpha_i T(\mathbf{w_i}) = \mathbf{0}$ then $T\left(\sum \alpha_i \mathbf{w_i}\right) = \mathbf{0}$, so $\sum \alpha_i \mathbf{w_i} \in \ker T$. Therefore we can write

$$\sum \alpha_i \mathbf{w_i} - \sum \beta_j \mathbf{u_j} = \mathbf{0}$$

for some β_j's. Since B is a basis, it follows that $\alpha_i = \beta_j = 0$ for all i, j, so B_2 is a linearly independent system.

Now for any $\mathbf{y} \in \operatorname{im} T$, there exists a $\mathbf{x} \in V$ with $T(\mathbf{x}) = \mathbf{y}$. Now we write $\mathbf{x} = \sum \alpha_i \mathbf{w_i} + \sum \beta_j \mathbf{u_j}$ to get $T\left(\sum \alpha_i \mathbf{w_i} + \sum \beta_j \mathbf{u_j}\right) = \mathbf{y}$. This is

$$\sum \alpha_i T(\mathbf{w_i}) + \sum \beta_j T(\mathbf{u_j}) = \mathbf{y},$$

but since $\mathbf{u_j} \in \ker T$, the second sum vanish so $\mathbf{y} \in \operatorname{span} B_2$, so $\operatorname{im} T \subseteq \operatorname{span} B_2$. On the other hand, it is clear that $\operatorname{span} B_2 \subset \operatorname{im} T$ because we can take linear combinations of images of T in B_2 so $\operatorname{span} B_2 = \operatorname{im} T$.

Combining all of the above, B_2 is a basis of $\operatorname{im} T$, so $\dim \operatorname{im} T = \ell$. ■

There are also special terms for $\dim \ker T$ and $\dim \operatorname{im} T$: *nullity* and *rank* of T, respectively.

Exercise 2.31. If $T \in \mathcal{L}(V)$ then T is 1-1 iff T is onto.

Solution. ($\Rightarrow$) If T is 1-1 then $\ker T = \{0\}$, so $\dim \ker T = 0$. By the Dimension formula, $\dim \operatorname{im} T = \dim V$. Now we claim $\operatorname{im} T = V$. Suppose not, then there is an element $\mathbf{v} \in V - \operatorname{im} T$. We can add this to a basis $(\mathbf{b_1}, \cdots \mathbf{b_n})$ of $\operatorname{im} T$, getting a linearly independent set of $\dim V + 1$ vectors in V, which is impossible. Therefore, $\operatorname{im} T = V$, so T is onto.

($\Leftarrow$) If T is onto then $\operatorname{im} T = V$, so $\dim \operatorname{im} T = \dim V$. By the Dimension Formula, $\dim \ker T = 0$ so $\ker T = \{\mathbf{o}\}$, hence T is 1-1. ■

Exercise 2.32. Give an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that

$$f(a\mathbf{v}) = af(\mathbf{v})$$

for all $a \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^2$ but f is not linear.

Solution. We claim that the function f defined as follows works:

$$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{cases} 0 & \text{if } y \neq 0 \\ x & \text{if } y = 0 \end{cases}$$

For any $\alpha \in \mathbb{R}$ and $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$, if $y = 0$ then $a\mathbf{v} = \begin{bmatrix} ax \\ 0 \end{bmatrix}$ so $f(a\mathbf{v}) = ax = af(\mathbf{v})$. If $y \neq 0$ then $f(a\mathbf{v}) = 0 = af(\mathbf{v})$, so f preserves scalar multiplication.

However, we have $f\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = 0 \neq 1 = f\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + f\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$, so f is not linear. ■

Exercise 2.33. Suppose that V is finite dimensional. Prove that any linear map on a subspace of V can be extended to a linear map on V. In other words, show that if U is a subspace of V and $S \in \mathcal{L}(U, W)$, then there exists $T \in \mathcal{L}(V, W)$ such that $T(\mathbf{u}) = S(\mathbf{u})$ for all $u \in U$.

Solution. Extend a basis $(\mathbf{b_1}, \ldots, \mathbf{b_n})$ of U to a basis $(\mathbf{b_1}, \ldots, \mathbf{b_n}, \mathbf{v_1}, \cdots, \mathbf{v_k})$ of V. Define T by $T(\mathbf{b_i}) = S(\mathbf{b_i})$, and $T(\mathbf{v_i}) = \mathbf{o}$. Then, for any $\mathbf{u} \in U$, write $\mathbf{u} = \sum \alpha_i \mathbf{b_i}$, so

$$T(\mathbf{u}) = \sum \alpha_i T(\mathbf{b_i}) = \sum \alpha_i S(\mathbf{b_i}) = S(\mathbf{u}). \quad ■$$

2.6 Test review

Syllabus for test on linear maps (Thursday Oct 4)

- Definitions:
  - Linear transformations
  - kernel and image
  - dimension formula
  - matrix of a linear transformation
  - (1-1, onto)
- Computations:
  - kernel, image, and bases therefor
  - matrix of a linear transformation
- Proofs:
  - Show that a map is linear, 1-1, onto

Some review:

Exercise 2.34. Prove that if T is a linear map from $\mathbb{F}^4$ to $\mathbb{F}^2$ such that

$$\ker T = \{(x_1, x_2, x_3, x_4) \in \mathbb{F}^4 | x_1 = 5x_2 \text{ and } x_3 = 7x_4\}$$

then T is surjective.

Solution. Note that $((5,1,0,0),(0,0,7,1))$ is a basis of $\ker T$, so $\dim\ker T = 2$. By the Dimension formula, $\dim \operatorname{im} T = 2 = \dim F^2$. Since $\operatorname{im} T$ is a subspace of $\mathbb{F}^2$, it follows that $\operatorname{im} T = \mathbb{F}^2$, so T is surjective. ■

Exercise 2.35 (2.3 in book ch.2 review).
a. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Are the following statements true or false?
   1. If $\ker T = \{\mathbf{0}\}$ and $T(\mathbf{y}) = \mathbf{b}$, then $\mathbf{y}$ is the only solution to $T(\mathbf{x}) = \mathbf{b}$.
   2. If $\mathbf{y}$ is the only solution to $T(\mathbf{x}) = \mathbf{c}$, then for any $\mathbf{b} \in \mathbb{R}^m$, a solution exists to $T(\mathbf{x}) = \mathbf{b}$.
   3. If $\mathbf{y} \in \mathbb{R}^n$ is a solution to $T(\mathbf{x}) = \mathbf{b}$, it is the only solution.
   4. If for any $\mathbf{b} \in \mathbb{R}^m$ the equation $T(\mathbf{x}) = \mathbf{b}$ has a solution, then it is the only solution.

b. For any statements that are false, can one impose conditions on m and n that make them true?

Solution.
a. 1. True. If $T(\mathbf{y}') = \mathbf{b}$ then $T(\mathbf{y}' - \mathbf{y}) = \mathbf{0}$ so $\mathbf{y}' - \mathbf{y} = \mathbf{0}$ so $\mathbf{y}' = \mathbf{y}$.
   2. False. If $m > n$ then $\dim \operatorname{im} T \leqslant n < m$ so $\operatorname{im} T \neq \mathbb{R}^m$.
   3. False. For example take the zero map $T(\mathbf{x}) \equiv \mathbf{0}$ which is clearly not injective.
   4. False in both interpretations.

b. 2. The statement is true for (m, n) iff $m \leqslant n$ (vacuously so if $m < n$). The equation $T(\mathbf{x}) = \mathbf{c}$ has a unique solution iff $\ker T = \{\mathbf{0}\}$ iff $\dim \operatorname{im} T = n$.
   3. No; the zero map is a counterexample for all (m, n).
   4. Depends on the interpretation; the statement "If (for any $\mathbf{b} \in \mathbb{R}^m$ the equation $T(\mathbf{x}) = \mathbf{b}$ has a solution), then it is the only solution." is true for (m, n) iff $m = n$. The statement "For any $\mathbf{b} \in \mathbb{R}^m$, if the equation $T(\mathbf{x}) = \mathbf{b}$ has a solution, then it is the only solution." is false for all (m, n). ■

Exercise 2.36. Prove that if there exists a linear map on V whose kernel and image are both finite dimensional, then V is finite dimensional.

Solution. Dimension formula, duh. ■

Exercise 2.37. Suppose that V and W are both finite dimensional. Prove that there exists a surjective linear map from V onto W if and only if $\dim W \leqslant \dim V$.

Solution. If there exists a surjective linear map T from V onto W then

$$\dim V = \dim\ker T + \dim\operatorname{im} T = \dim\ker T + \dim W \geqslant \dim V.$$

On the other hand, if $\dim V \geqslant \dim W$, let $B_V = \{\mathbf{v_1}, \dots, \mathbf{v}_{\dim V}\}$ be the basis of V, $B_W = \{\mathbf{w_1}, \dots, \mathbf{w}_{\dim W}\}$ be the basis of W. Then, the transformation $T \in \mathcal{L}(V, W)$ defined by $T(\mathbf{v_i}) = \mathbf{w_i}$ for $i = 1, \dots, \dim W$ and $T(\mathbf{v_j}) = \mathbf{o}$ for $j > \dim W$ is onto. ■

Exercise 2.38. Suppose that V and W are finite dimensional and that U is a subspace of V. Prove that there exists $T \in \mathcal{L}(V, W)$ such that $\ker T = U$ if and only if $\dim U \geqslant \dim V - \dim W$.

Solution. If there exists such a T then, since $\operatorname{im} T \subseteq W$, by the dimension formula,

$$\dim U = \dim\ker T = \dim V - \dim\operatorname{im} T \geqslant \dim V - \dim W.$$

On the other hand, if $\dim U > \dim V - \dim W$, let B_U be a basis of U, and B_V be a basis of V that is an extension of B_U. Also let B_W be a basis of W. The following T works: for all $\mathbf{u} \in B_U$, let $T(\mathbf{u}) = 0$, and for each $\mathbf{v} \in B_V \setminus B_U$, let $T(\mathbf{v})$ be different members of B_W. (This is possible as $\dim W \geqslant \dim V - \dim U$.) ■

Exercise 2.39. Suppose that U and V are finite-dimensional vector spaces and that $S \in \mathcal{L}(V, W), T \in \mathcal{L}(U, V)$. Prove that

$$\dim\ker ST \leqslant \dim\ker S + \dim\ker T.$$

Solution. I'm opting for a more rigorous solution than the one we got in class. Let B_T be a basis of $\ker T$. Since $\ker S$ and $\operatorname{im} T$ are both subspaces of V, $V' = \ker S \cap \operatorname{im} T$ is a subspace of V as well. Choose a basis $B_{V'} = \{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v}_\ell\}$ of V', and pick $\mathbf{b_1}, \dots, \mathbf{b}_\ell \in U$ such that $T\mathbf{b_i} = \mathbf{v_i}$ for $i = 1, \dots, \ell$.

Now, suppose that $\mathbf{u} \in U$ makes $ST\mathbf{u} = \mathbf{o}$. Then, $T\mathbf{u} \in V'$, so we can write $T\mathbf{u} = \sum_{i=1}^{\ell} \alpha_i \mathbf{v_i}$. Now note that for $\mathbf{u'} = \sum_{i=1}^{\ell} \alpha_i \mathbf{b_i}$, $T\mathbf{u'} = T\mathbf{u}$ so $\mathbf{u'} - \mathbf{u} \in \ker T$ so

$$\mathbf{u} \in \operatorname{span}\left(B_T \cup \{\mathbf{b_1}, \dots, \mathbf{b}_\ell\}\right)$$

Hence, $\ker ST \subseteq \operatorname{span}\left(B_T \cup \{\mathbf{b_1}, \dots, \mathbf{b}_\ell\}\right)$ so

$$\dim\ker ST \leqslant \dim\ker T + \dim(\ker S \cap \operatorname{im} T) \leqslant \dim\ker T + \dim\ker S. \quad ■$$

Exercise 2.40. Suppose that V is finite dimensional and $S, T \in \mathcal{L}(V)$. Prove that ST is invertible if and only if both S and T are invertible.

Solution. First, note that for any $R \in \mathcal{L}(V)$, by the Dimension Formula, $\ker R = \{\mathbf{o}\}$ iff $\operatorname{im} R = V$, so both of these are equivalent to R being invertible.

($\Rightarrow$) If ST is invertible then $\operatorname{im} ST = V$ and $\ker ST = \{\mathbf{o}\}$. Since $\operatorname{im} ST \subseteq \operatorname{im} S$, $\operatorname{im} S = V$ so S is invertible. Since $\ker T \subseteq \ker ST$, $\ker T = \{\mathbf{o}\}$ so T is invertible.

($\Leftarrow$) If both S and T are invertible then consider $T^{-1}S^{-1}$. Since $(T^{-1}S^{-1})(ST) = (ST)(T^{-1}S^{-1})$, ST is invertible with $T^{-1}S^{-1}$ as its inverse. ■

Exercise 2.41. Suppose that V is finite dimensional and $S, T \in \mathcal{L}(V)$. Prove that $ST = I$ if and only if $TS = I$.

Solution. Since $ST = I$ is invertible, T is invertible. Then, $TS = TSTT^{-1} = TT^{-1} = I$.

Exercise 2.42. Suppose that T is a linear map from V to $\mathbb{R}$. Prove that if $\mathbf{u} \in V$ is not in $\ker T$ then

$$V = \ker T \oplus \{a\mathbf{u} | a \in \mathbb{R}\}$$

Notation: $V = U \oplus W$ *iff* $V = U + W$ *and* $U \cap W = \{\mathbf{o}\}$

Solution. Let $n = \dim V$. Since $\operatorname{im} T \subseteq \mathbb{R}$, $\dim \operatorname{im} T \leqslant 1$. Since $\mathbf{u} \notin \ker T$, $\ker T \neq V$ so $\dim \ker T \leqslant n - 1$. However, by the Dimension Formula,

$$n = \dim V = \dim \ker T + \dim \operatorname{im} T \leqslant (n-1) + 1 = n$$

so $\dim \ker T = n - 1$ and $\dim \operatorname{im} T = 1$. Since $\mathbf{u} \notin \ker T$, neither is $a\mathbf{u}$ for any $a \neq 0 \in \mathbb{R}$, so $\ker T \cap \{a\mathbf{u} | a \in \mathbb{R}\} = \{\mathbf{o}\}$. Let B be a basis of $\ker T$. Then, $B \cup \{\mathbf{u}\}$ is a basis of $\ker T + \{a\mathbf{u} | a \in \mathbb{R}\}$ so $\dim \ker T + \{a\mathbf{u} | a \in \mathbb{R}\} = n = \dim V$ so $\ker T + \{a\mathbf{u} | a \in \mathbb{R}\} = V$. ■

3 Geometry in $\mathbb{R}^n$

3.1 The dot product

We introduce measures for lengths and angles in $\mathbb{R}^n$. In particular, we are interested in defining orthogonality (right angles).

Definition 3.1.
Let $\mathbf{v} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Then, the *dot product* $\mathbf{v} \cdot \mathbf{w}$ is defined as

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} x_i y_i.$$

The dot product can easily be seen to be commutative and distributive.

Definition 3.2.
The *norm*, which is a measure of length, is defined as $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$. The alternate notation $|\mathbf{v}|$ can be used as well, but this notation is normally reserved for $\mathbb{R}$ and $\mathbb{C}$.

Definition 3.3.
The *distance* between two vectors $\mathbf{v}, \mathbf{w}$ is the norm of their difference: $d(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$.

Definition 3.4.
The *angle* between two vectors $\mathbf{v}, \mathbf{w}$ $\angle(\mathbf{v}, \mathbf{w})$ is defined by

$$\cos \angle(\mathbf{v}, \mathbf{w}) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|}.$$

Now note that this definition would only work if $\frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|} \in [-1, 1]$, and this is indeed the case–we will prove it later.

First let's try see it makes sense in $\mathbb{R}^2$. Suppose we have two vectors $\mathbf{v}, \mathbf{w}$ making an angle θ. By the cosine law, we have

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

which expands to

$$2\mathbf{v} \cdot \mathbf{w} = 2\|v\|\|w\| \cos \theta$$

as we wanted.

Let's prove that $\frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|}$. In fact this theorem has a familiar name:

Theorem 3.5 (Cauchy-Schwarz). *For any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$,*

$$|\mathbf{v} \cdot \mathbf{w}| \leqslant \|\mathbf{v}\|\|\mathbf{w}\|.$$

Equality holds iff one of $\mathbf{v}$ or $\mathbf{w}$ is a scalar multiple of the other.

Proof. If either $\mathbf{v}$ or $\mathbf{w}$ is $\mathbf{0}$, the statement is obvious, so suppose both are nonzero. Consider the function $\|\mathbf{v} + t\mathbf{w}\|^2$ as a function of t. It is a second-degree polynomial of the form $at^2 + bt + c$:

$$\|t\mathbf{w} + \mathbf{v}\|^2 = \|\mathbf{w}\|^2 t^2 + 2\mathbf{v} \cdot \mathbf{w}t + \|\mathbf{v}\|.$$

This polynomial evidently has at most one root, so its discriminant is ≤ 0, which means

$$4(\mathbf{v}\cdot\mathbf{w})^2 - 4\|\mathbf{v}\|^2\|\mathbf{w}\|^2 \leqslant 0$$

, which is what we want! To prove that equality holds only when one of v or w is a scalar multiple of the other, note that the discriminant equality holds iff the quadratic polynomial has exactly one root, which means $\|t\mathbf{w}+\mathbf{v}\|$ is zero for exactly one value of t, which means $t\mathbf{w}+\mathbf{v}=0$. ■

3.2 Orthogonality and projections

Definition 3.6.
Two vectors $\mathbf{v},\mathbf{w}\in\mathbb{R}^n$ are *orthogonal*, denoted by $\mathbf{v}\perp\mathbf{w}$, if $\mathbf{v}\cdot\mathbf{w}=0$.

Definition 3.7.
A basis of $\mathbb{R}^n$ in which any two vectors are orthogonal is called an *orthogonal basis*. If, in addition, each vector has norm equal to 1, the basis is said to be *orthonormal*.

Orthogonality and orthonormality is very useful because of the following:

Theorem 3.8. *Let $\mathbf{v}\in\mathbb{R}^n$ and $(\mathbf{b_1},\mathbf{b_2},\cdots,\mathbf{b_n})$ be an o.n.b. (short for orthonormal basis). Then,*

$$\mathbf{v}=\sum_{i=1}^{n}(\mathbf{v}\cdot\mathbf{b_i})\mathbf{b_i}.$$

Proof. Let $\mathbf{v}=\sum\alpha_i\mathbf{b_i}$ for $\alpha_i\in\mathbb{R}$. We have

$$\mathbf{v}\cdot\mathbf{b_j}=\sum_{i=1}^{n}\alpha_i(\mathbf{b_i}\cdot\mathbf{b_j})=\alpha_i$$

since $\mathbf{b_i}\cdot\mathbf{b_j}=1$ if $i=j$ and 0 otherwise. ■

Definition 3.9.
Let $P\in\mathcal{L}(V)$ with $P^2=P$. Then, P is called a *projection*. If, furthermore, $P(\mathbf{v})\perp\big(P(\mathbf{v})-\mathbf{v}\big)$, then P is called an *orthogonal projection*.

Example 3.10. Some familiar projections include projecting a force onto the direction of displacement when calculating work in physics.

If a projection $P\in\mathcal{L}(V)$ is 1-1 / onto (on V) then from $P(P(\mathbf{v}))=P(\mathbf{v})$ so $P(\mathbf{v})=\mathbf{v}$ so P is the identity.

Exercise 3.11 (1.4.24 in book)**.** Let $\mathbf{v}\in\mathbb{R}^n$ be a nonzero vector, and denote by $\mathbf{v}^\perp\subset\mathbb{R}^n$ the set of vectors $\mathbf{w}\in\mathbb{R}^n$ such that $\mathbf{v}\cdot\mathbf{w}=0$.

(a) Show that $\mathbf{v}^\perp$ is a subspace of $\mathbb{R}^n$.

(b) Given any vector $\mathbf{a}\in\mathbb{R}^n$, show that $\mathbf{a}-\frac{\mathbf{a}\cdot\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v}$ is an element of $\mathbf{v}^\perp$.

(c) Define the projection of $\mathbf{a}$ onto $\mathbf{v}^\perp$ by the formula

$$P_{\mathbf{v}^\perp}(\mathbf{a})=\mathbf{a}-\frac{\mathbf{a}\cdot\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v}.$$

Show that there is a unique number $t(\mathbf{a})$ such that $\big(\mathbf{a}+t(\mathbf{a})\mathbf{v}\big)\in\mathbf{v}^\perp$. Show that

$$\mathbf{a}+t(\mathbf{a})\mathbf{v}=P_{\mathbf{v}^\perp}(\mathbf{a}).$$

Solution. (a) i. For any $\mathbf{w_1}, \mathbf{w_2} \in \mathbf{v}^\perp$, we have $\mathbf{v} \cdot (\mathbf{w_1} + \mathbf{w_2}) = \mathbf{v} \cdot \mathbf{w_1} + \mathbf{v} \cdot \mathbf{w_2} = 0$, so $w_1 + w_2 \in \mathbf{v}^\perp$.

ii. For any $\mathbf{w} \in \mathbf{v}^\perp$ and $\alpha \in \mathbb{R}$, we have $\mathbf{v} \cdot (\alpha \mathbf{w}) = \alpha(\mathbf{v} \cdot \mathbf{w}) = 0$ so $\alpha \mathbf{w} \in \mathbf{v}^\perp$.

Therefore $\mathbf{v}^\perp$ is a subspace of $\mathbb{R}^n$.

(b) Since

$$\mathbf{v} \cdot \left(\mathbf{a} - \frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v} \right) = \mathbf{v} \cdot \mathbf{a} - \left(\frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \right) \mathbf{v} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{a} - \left(\frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \right) \|\mathbf{v}\|^2 = 0,$$

it follows that $\mathbf{a} - \frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v}$ is in $\mathbf{v}^\perp$.

(c) i. First, $t(a) = -\frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{v}\|^2}$ works. Now suppose t_1, t_2 are numbers such that both $\mathbf{a} + t_1 \mathbf{v}$ and $\mathbf{a} + t_2 \mathbf{v}$ are in $\mathbf{v}^\perp$. Then, since $\mathbf{v}^\perp$ is a subspace, $(\mathbf{a} + t_1 \mathbf{v}) - (\mathbf{a} + t_2 \mathbf{v}) = (t_1 - t_2)\mathbf{v} \in \mathbf{v}^\perp$. This means $0 = \mathbf{v} \cdot (t_1 - t_2)\mathbf{v} = (t_1 - t_2)\|\mathbf{v}\|^2$, but $\mathbf{v}$ is nonzero, so $t_1 = t_2$.

ii. Since $P_{\mathbf{v}^\perp}(\mathbf{a})$ is in $\mathbf{v}^\perp$, and it can be written in the form $\mathbf{a} + t\mathbf{v}$ with $t \in \mathbb{R}$, from i. we have $t = t(a)$. ∎

How is the formula for $P_{\mathbf{v}^\perp}(\mathbf{a})$ derived? Since we want an orthogonal projection onto $\mathbf{v}^\perp$, we have $\mathbf{a} - P_{\mathbf{v}^\perp}(\mathbf{a}) = t\mathbf{v}$ for some $t \in \mathbb{R}$. Now we want $(\mathbf{a} - t\mathbf{v}) \cdot \mathbf{v} = 0$, and solving this yields $t = \frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{v}\|^2}$.

Exercise 3.12. Show that $P_{\mathbf{v}^\perp}$ is indeed a projection.

Solution. Let $\mathbf{a} \in V$. Since $P_{\mathbf{v}^\perp}(\mathbf{a}) \in \mathbf{v}^\perp$, $\mathbf{v} \cdot P_{\mathbf{v}^\perp}(\mathbf{a}) = 0$, so

$$P_{\mathbf{v}^\perp}(P_{\mathbf{v}^\perp}(\mathbf{a})) = P_{\mathbf{v}^\perp}(\mathbf{a}) - \frac{P_{\mathbf{v}^\perp}(\mathbf{a}) \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v} = P_{\mathbf{v}^\perp}(\mathbf{a}). \quad ∎$$

Now let's see how matrices of projections look like.

Example 3.13. Suppose we have $\mathbf{v} \in \mathbb{R}^3$ with $\|\mathbf{v}\| = 1$. The matrix of $P_{\mathbf{v}^\perp}$ is given by $[P_{\mathbf{v}^\perp}(\mathbf{e_i})]_{i=1}^3$.

Let's say $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$. Then, $P_{\mathbf{v}^\perp}(\mathbf{e_1}) = \mathbf{e_1} - (\mathbf{e_1} \cdot \mathbf{v})\mathbf{v} = \begin{bmatrix} 1 - v_1^2 \\ -v_1 v_2 \\ -v_1 v_3 \end{bmatrix}$. This, and similar results for $\mathbf{e_2}$ and $\mathbf{e_3}$, gives

$$[P_{\mathbf{v}^\perp}] = \begin{bmatrix} 1 - v_1^2 & -v_1 v_2 & -v_1 v_3 \\ -v_1 v_2 & 1 - v_2^2 & -v_2 v_3 \\ -v_1 v_3 & -v_2 v_3 & 1 - v_3^2 \end{bmatrix} = I_3 - \mathbf{v}\mathbf{v}^T.$$

Clearly this generalizes to $\mathbb{R}^n$ as well.

Exercise 3.14 (1.4.27 in Book)**.** Let $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ be a unit vector in $\mathbb{R}^3$, so $a^2 + b^2 + c^2 = 1$.

(a) Show that the transformation $T_\mathbf{a}$ defined by $T_\mathbf{a}(\mathbf{v}) = \mathbf{v} - 2(\mathbf{a} \cdot \mathbf{v})\mathbf{a}$ is a linear transformation $\mathbb{R}^3 \to \mathbb{R}^3$.

(b) What is $T_\mathbf{a}(\mathbf{a})$? If $\mathbf{v}$ is orthogonal to $\mathbf{a}$, what is $T_\mathbf{a}(\mathbf{v})$? Can you give a name to $T_\mathbf{a}$?

(c) Write the matrix M of $T_\mathbf{a}$ (in terms of a, b, c, of course). What can you say of M^2?

3.3 Determinants

Definition 3.15.
Let $\mathbf{a_1}, \ldots, \mathbf{a_n} \in \mathbb{R}^n$. Then, the *determinant* $\det \begin{bmatrix} \mathbf{a_1} & \cdots & \mathbf{a_n} \end{bmatrix}$ is defined inductively as follows:

- For $n = 1$, $\det[a] = a$.
- For $A = \det \begin{bmatrix} \mathbf{a_1} & \cdots & \mathbf{a_n} \end{bmatrix}$, define $\tilde{A_{ij}}$ as the matrix obtained from A by deleting row i and column j. Then

$$\det A = a_{11} \det \tilde{A_{11}} - a_{21} \det \tilde{A_{21}} + \cdots + (-1)^{n+1} a_{n1} \det \tilde{A_{n1}}$$

Example 3.16. $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = a \cdot d - b \cdot c.$

Example 3.17. $\det \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & -2 \\ 2 & 0 & 1 \end{bmatrix} = 1 \cdot \det \begin{bmatrix} 3 & -2 \\ 0 & 1 \end{bmatrix} - 0 \cdot \det \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix} + 2 \cdot \det \begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix} =$
$1 \cdot 3 - 0 \cdot 2 + 2 \cdot (-1) = 1$

Note: you can expand the determinant into any row or column using the above principle. The signs are determined by the “checkerboard rule”:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Example 3.18. From the last example, using the second column instead of the first also gives 1. Verification left to the reader.

Some more properties of determinants, given without proof:

- Let $P(\mathbf{v_1}, \ldots, \mathbf{v_n})$ be the parallelogram in $\mathbb{R}^n$ spanned by the vectors $\mathbf{v_1}, \ldots, \mathbf{v_n}$. Then the volume of $P(\mathbf{v_1}, \ldots, \mathbf{v_n})$ is equal to $|\det[\mathbf{v_1}, \ldots, \mathbf{v_n}]|$
- $\det[\mathbf{v_1}, \ldots, \mathbf{v_n}]$ is a multilinear and antisymmetric map $(\mathbb{R}^n)^n \to \mathbb{R}$. In more understandable terms, it is linear in each of the arguments $\mathbf{v_1}, \ldots, \mathbf{v_n}$, and if you switch any two arguments, the determinant will be multiplied by -1.
- It is invariant under transposition, that is, $\det A = \det A^T$ for any square matrix A.
- It is multiplicative, that is, $\det AB = \det A \cdot \det B$.

Preview: a multilinear, anti-symmetric map $\phi : (\mathbb{R}^n)^k \to \mathbb{R}$ is called a *k-form* on $\mathbb{R}^n$.

Let us go back to the fact that the determinant is *multilinear and antisymmetric* again. What this means is that if any column $\mathbf{v_i}$ is repeated in the determinant, the determinant becomes zero as we can swap the $\mathbf{v_i}$'s.

Finally, as I feel that it is very useful, even though we didn't cover it in class, it is worth knowing the Leibniz form of the determinant:

$$\det A = \sum_\sigma \operatorname{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)}.$$

3.4 Problems

Let's go back to Exercise 1.4.27 in the book.

Solution (to 1.4.27). (a) Follows from the linearity of dot products.

(b) $T_{\mathbf{a}}(\mathbf{a}) = -\mathbf{a}$; if $\mathbf{v}$ is orthogonal to $\mathbf{a}$, $T_{\mathbf{a}}(\mathbf{v}) = \mathbf{v}$; $T_{\mathbf{a}}$ is a reflection.

(c) $M = \begin{bmatrix} 1-2a^2 & -2ab & -2ac \\ -2ab & 1-2b^2 & -2bc \\ -2ac & -2bc & 1-2c^2 \end{bmatrix}$; $M^2 = I$; T^2 is the identity.

Exercise 3.19 (1.4.16a in Book)**.** What is the area of the parallelogram with vertices at $\binom{0}{0}, \binom{1}{2}, \binom{5}{1}, \binom{6}{3}$?

Answer: $\left|\det \begin{bmatrix} 1 & 5 \\ 2 & 1 \end{bmatrix}\right| = 9.$

Exercise 3.20 (1.4.5 in Book)**.** Calculate the angles between the following pairs of vectors:

a. $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

b. $\begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

Answer: a. $\arccos \frac{1}{\sqrt{3}} = 54.74°$

b. $\arccos 0 = 90°$

Exercise 3.21. Normalize the following vectors: $\begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} -3 \\ 7 \end{bmatrix}, \begin{bmatrix} \sqrt{2} \\ -2 \\ -5 \end{bmatrix}.$

Answer: $\begin{bmatrix} 0 \\ \frac{1}{\sqrt{17}} \\ \frac{4}{\sqrt{17}} \end{bmatrix}, \begin{bmatrix} -\frac{3}{\sqrt{58}} \\ \frac{7}{\sqrt{58}} \end{bmatrix}, \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{31}} \\ -\frac{2}{\sqrt{31}} \\ -\frac{5}{\sqrt{31}} \end{bmatrix}.$

Since $\det A^{-1} = (\det A)^{-1}$, matrices with determinant zero are non-invertible.

Exercise 3.22. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and let $\mathbf{w}$ be a vector such that $\mathbf{v} \cdot \mathbf{w} = 42$. What is the shortest $\mathbf{w}$ can be? The longest it can be?

Answer: $\|\mathbf{w}\|$ can take any value in $[3\sqrt{14}, \infty)$.

Solution. Cauchy-Schwarz for the lower bound, attained at $\mathbf{w} = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$. Take $\mathbf{w} = \begin{bmatrix} 2x \\ -x \\ 14 \end{bmatrix}$ with $x \to +\infty$ to get $\|\mathbf{w}\|$ as large as you want.

Exercise 3.23. Let P be a projection, $P \in \mathcal{L}(V)$. Show that

a) The identity map I is the only invertible projection.

b) $Q = I - P$ is a projection

c) $\ker Q = \operatorname{im} P$ and $\operatorname{im} Q = \ker P$.

Solution. a) Clearly, the identity map is a projection. If a projection P is invertible, then for any $\mathbf{v} \in V$, $P(\mathbf{v}) = P^{-1}(P(P(\mathbf{v}))) = P^{-1}(P(\mathbf{v})) = \mathbf{v}$ so $P = I$.

b) For any $\mathbf{v} \in V$,

$$Q(Q(\mathbf{v})) = I(I(\mathbf{v}) - P(\mathbf{v})) - P(I(\mathbf{v}) - P(\mathbf{v})) = \mathbf{v} - P(\mathbf{v}) - P(\mathbf{v} - P(\mathbf{v})).$$

Therefore, using linearity of P,

$$Q(Q(\mathbf{v})) = v - P(\mathbf{v}) - \cancel{P(\mathbf{v})} + \cancel{P(P(\mathbf{v}))} = v - P(\mathbf{v}) = Q(\mathbf{v}),$$

so Q is a projection.

c) Let P_0 be a projection. If $\mathbf{v} \in \operatorname{im} P_0$ then there is a $\mathbf{v}'$ such that $P_0(\mathbf{v}') = \mathbf{v}$, so $P_0(\mathbf{v}) = P_0(P_0(\mathbf{v}')) = P_0(\mathbf{v}') = \mathbf{v}$. On the other hand, if $P_0(\mathbf{v}) = \mathbf{v}$ then $\mathbf{v} \in \operatorname{im} P_0$, so $\mathbf{v} \in \operatorname{im} P_0 \iff P_0(\mathbf{v}) = \mathbf{v}$. Using the above, we have

$$v \in \ker Q \iff Q(\mathbf{v}) = \mathbf{o} \iff P(\mathbf{v}) = \mathbf{v} \iff v \in \operatorname{im} P$$

and

$$v \in \operatorname{im} Q \iff Q(\mathbf{v}) = \mathbf{v} \iff P(\mathbf{v}) = \mathbf{o} \iff v \in \ker P$$

so $\ker Q = \operatorname{im} P$ and $\operatorname{im} Q = \ker P$. ■

Exercise 3.24 (2.15 in book ch.2 review)**.**

Show that an orthogonal 2×2 matrix (this actually means orthonormal) is either a rotation or a reflection.

Solution. There are two possible forms to an orthogonal 2×2 matrix:

$$\begin{bmatrix} a & b \\ b & -a \end{bmatrix}, \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

for $a^2 + b^2 = 1$. For the former form, since

$$\begin{bmatrix} a & b \\ b & -a \end{bmatrix}^2 = I \text{ and } \begin{bmatrix} a & b \\ b & -a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ bx - ay \end{bmatrix}$$

with $\sqrt{(ax + by)^2 + (bx - ay)^2} = \sqrt{x^2 + y^2}$, the matrix describes an isometric involution, which is a reflection. For the latter form, since $a^2 + b^2 = 1$, there exists θ such that $a = \cos\theta, b = \sin\theta$, the matrix describes the rotation by θ. ■

4 Linear regression project

For Hotchkiss class MA661 Linear Algebra. Using the guideline questions, we develop the theory of linear regression and apply it to create a model predicting body height based on hand span, foot length, and femur length.

4.1 Introduction

Suppose that we want to model a quantity Y from observed quantities $X_1, X_2, \ldots, X_p$. Clearly, there are as many ways to do this as there are functions $f : \mathbb{R}^p \to \mathbb{R}$. But what if we restrict the function to be linear? *Linear regression* is exactly that.

Let's say we collected N data sets $(X_{1i}, X_{2i}, \ldots, X_{pi}, Y_i)_{i=1}^N$ as *training data* for our model to work on. We want to find coefficients $\beta_0, \beta_1, \ldots, \beta_p$ such that

$$Y_i \approx \hat{y}_i = \beta_0 + \sum_{j=i}^{p} \beta_j X_{ji} \tag{1}$$

for each $i = 1, 2, \ldots, N$. Here $\hat{y}_i$ is the our model's *predicted value*. But what exactly does the $\approx$ symbol mean here? How do we say one approximation (for all N data sets) is better or worse than another? The answer is that we use the *least squares method*: we want the sum of squares of errors

$$\sum_{i=1}^{N} (Y_i - \hat{y}_i)^2 \tag{2}$$

to be minimized. Looking at what $\hat{y}_i$ expands to, this would be pretty ugly, except that we have linear algebra to the rescue!

Let $\mathbf{y} = (Y_i)_{i=1}^N$ and $\hat{\mathbf{y}} = (\hat{y}_i)_{i=1}^N$ be vectors in $\mathbb{R}^N$, $X = (\mathbf{1}, \mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_p})$ an $N \times (p+1)$ matrix (where $\mathbf{1}$ is the vector with all 1's), and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ be a vector of unknown coefficients in $\mathbb{R}^{p+1}$. Using this notation, (1) can be written as

$$\hat{\mathbf{y}} = X\beta$$

and the sum of squares in (2) becomes just

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \|\mathbf{y} - X\beta\|^2 .$$

With this, we are ready to delve into the theory!

4.2 Theory Questions

1. *Explain why the* orthogonal *projection of the training data y into the column space of X is a sensible choice for minimizing the total error $\|\mathbf{y} - \hat{\mathbf{y}}\|$.*

Geometrical intuition tells us that the distance from a point A to another point B on a line/plane/n-dimensional space is minimized when AB is perpendicular to the line/plane/space, that is, when B is the projection of A onto the line/plane/space.

Since the column space of X is, by definition, span $(\mathbf{1}, \mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_p})$, which is precisely all the values $\hat{\mathbf{y}} = X\beta$ can take. Since we want our choice of $\hat{\mathbf{y}}$ to be closest to $\mathbf{y}$, intuitively, we want $\hat{\mathbf{y}}$ to be the projection of $\mathbf{y}$ onto the column space of X.

2. *Show that the rank of X (or $\dim \operatorname{im} X$) is equal to the number of linearly independent columns of X.*

Since for a vector $\mathbf{v} = (v_0, v_1, \dots, v_p)^T$, $X\mathbf{v}$ is simply $\sum_{j=0}^{p} v_j \mathbf{X_j}$ (taking $\mathbf{X_0} = \mathbf{1}$), $\operatorname{im} X$ is precisely the column space of X. Suppose that the number of linearly independent columns of X is m, that is, there is a set W of m columns of X that are linearly independent, but not such a set of $m+1$ columns. Therefore, any column $\mathbf{X_k} \notin W$ can be written as a linear combination of vectors in W:

$$\mathbf{X_k} = \sum_{X_j \in W} \alpha_{kj} \mathbf{X_j}.$$

(else $W \cup \{\mathbf{X_k}\}$ sould be a linearly independent set of $m+1$ columns.) Using this we can write each $\mathbf{u} = \sum_{j=0}^{p} \gamma_j \mathbf{X_j} \in \operatorname{span}(\mathbf{X_0}, \mathbf{X_1}, \dots, \mathbf{X_p})$ as

$$\mathbf{u} = \sum_{X_j \in W} \gamma_j \mathbf{X_j} + \sum_{X_i \notin W} \gamma_i \left(\sum_{X_j \in W} \alpha_{ij} \mathbf{X_j} \right),$$

so W spans the column space of X. Therefore W is a basis of $\operatorname{im} X$, which means that $\dim \operatorname{im} X = m$.

3. *Argue why it is reasonable to assume that realistic examples of X have full rank (We say that X has full rank if all columns of X are linearly independent)*

If X does not have full rank, there has to be a linear combination

$$\sum_{j=0}^{p} \alpha_j \mathbf{X_j} = \mathbf{0}$$

with coefficients α_j not all zero. This means that

$$\sum_{j=0}^{p} \alpha_j X_{ji} = 0$$

perfectly for all $i = 1, \dots, N$. In realistic situations where we collect enough data (N is big enough), there is likely to be at least some error/variance from measurement, so it is unlikely that there would be a linear combination that is zero for all $i = 1, \dots, N$.

4. *For the case $p = 1$, but general N, show directly: If X has full rank, $X^T X$ is invertible*

Let $\mathbf{X_1} = (X_{11}, X_{12}, \dots, X_{1N})^T$. We have

$$X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{12} & \cdots & X_{1N} \end{bmatrix} \begin{bmatrix} 1 & X_{11} \\ 1 & X_{12} \\ \vdots & \vdots \\ 1 & X_{1N} \end{bmatrix} = \begin{bmatrix} N & \sum_{i=1}^{n} X_{1i} \\ \sum_{i=1}^{N} X_{1i} & \sum_{i=1}^{N} X_{1i}^2 \end{bmatrix}.$$

Let $S_1 = \sum_{i=1}^{N} X_{1i}$ and $S_2 = \sum_{i=1}^{n} X_{1i}^2$. Since X has full rank, the X_{1i}'s are not all equal (else, $X_{11}\mathbf{1} - \mathbf{X_1} = \mathbf{0}$ means X does not have full rank.) Hence,

$$NS_2 - S_1^2 = \sum_{1 \leqslant i < j \leqslant N} (X_{1i} - X_{1j})^2 \neq 0.$$

Therefore, by the formula for inverse of 2×2 matrices [1], $X^T X$ is invertible with inverse

$$\frac{1}{NS_2 - S_1^2} \begin{bmatrix} S_2 & -S_1 \\ -S_1 & N \end{bmatrix}.$$

5. *For general p, show: If X has full rank, $X^T X$ is invertible.*

Consider a vector $\mathbf{v} \in \ker X^T X$, and let $\mathbf{w} = X\mathbf{v}$. We have

$$\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w} = \mathbf{v}^T X^T X\mathbf{v} = 0,$$

so $\mathbf{w} = \mathbf{o}$, which means that $\mathbf{v} \in \ker X$ as well. Therefore $\ker X^T X \subset \ker X$. However, X has full rank, therefore all columns of X are linearly independent, thus $\ker X = \{\mathbf{o}\}$. Therefore, $\ker X^T X = \{\mathbf{o}\}$ as well.

Next we claim that $X^T X$ has full rank. Let $\mathbf{c_o}, \ldots, \mathbf{c_p}$ be the columns of $X^T X$. Suppose that there are real numbers $\alpha_0, \ldots, \alpha_p$ such that

$$\sum_{j=0}^{p} \alpha_i \mathbf{c_j} = \mathbf{o}.$$

Then, the vector $\alpha = (\alpha_0, \cdots, \alpha_p)^T$ satisfy $X^T X\alpha = \mathbf{o}$, so $\alpha \in \ker X^T X \Rightarrow \alpha = \mathbf{o}$. Therefore $X^T X$ has full rank.

Now consider the column space $\operatorname{im} X^T X$ of $X^T X$. Since $X^T X$ has full rank, all columns of $X^T X$ are linearly independent, so $\dim \operatorname{im} X^T X = p + 1$. Since $X^T X \subset \mathbb{R}^{p+1}$, it follows that $\operatorname{im} X^T X = \mathbb{R}^{p+1}$. Therefore, there are vectors $\mathbf{u_1}, \mathbf{u_{p+1}}$ such that

$$X^T X\mathbf{u_j} = \mathbf{e_j}$$

for each $j = 1, 2, \ldots, p + 1$. It then follows that the matrix

$$U = \begin{bmatrix} \mathbf{u_o} & \mathbf{u_1} & \cdots & \mathbf{u_{p+1}} \end{bmatrix}$$

makes $(X^T X)U = I$

Observe[2] that

$$(X^T X)U(X^T X) = X^T X \quad \text{implies} \quad (X^T X)(UX^T X - I) = 0.$$

Since $X^T X$ has full rank, $UX^T X - I$ must be the zero matrix, so $UX^T X = I$ as well, so $X^T X$ is invertible (with inverse U).

6. *Show: If X has full rank, the solution of $X^T(\mathbf{y} - X\beta) = \mathbf{o}$ is $\hat{\beta} = (X^T X)^{-1}X^T\mathbf{y}$.*

Since $X^T(\mathbf{y} - X\hat{\beta}) = X^T(\mathbf{y} - X(X^T X)^{-1}X^T\mathbf{y}) = X^T\mathbf{y} - \cancel{X^T X(X^T X)^{-1}}X^T\mathbf{y} = \mathbf{o}$, $\hat{\beta} = (X^T X)^{-1}X^T\mathbf{y}$ is a solution of $X^T(\mathbf{y} - X\beta) = \mathbf{o}$.

If there are two solutions $\beta = \beta_1, \beta_2$ to $X^T(\mathbf{y} - X\beta) = \mathbf{o}$ then

$$X^T X(\beta_1 - \beta_2) = X^T(\mathbf{y} - X\beta_1) - X^T(\mathbf{y} - X\beta_2) = \mathbf{o}.$$

Since $\ker X^T X = \{\mathbf{o}\}$, it follows that $\beta_1 - \beta_2 = \mathbf{o}$, that is, $\beta_1 = \beta_2$. Therefore $\hat{\beta} = (X^T X)^{-1}X^T\mathbf{y}$ is *the* solution of $X^T(\mathbf{y} - X\beta) = \mathbf{o}$.

[1] Shown in class: the inverse of a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ where $ad - bc \neq 0$ is $A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

[2] Based on Davidac897's answer in Math.SE at `https://math.stackexchange.com/a/3860`.

7. *Show that the best-fit model is $\hat{\mathbf{y}} = X(X^TX)^{-1}X^T\mathbf{y}$, and that $H := X(X^TX)^{-1}X^T$ is the matrix of an orthogonal projection.*

From

$$\mathbf{o} = X^T(\mathbf{y} - X\hat{\beta}) = \begin{bmatrix} \mathbf{X_0} \cdot (\mathbf{y} - X\hat{\beta}) \\ \mathbf{X_1} \cdot (\mathbf{y} - X\hat{\beta}) \\ \vdots \\ \mathbf{X_p} \cdot (\mathbf{y} - X\hat{\beta}) \end{bmatrix}, \tag{3}$$

for any $j = 0, \dots, p$ we have $\mathbf{X_j} \cdot (\mathbf{y} - X\hat{\beta}) = 0$ so $\mathbf{y} - X\hat{\beta}$ is orthogonal to every column of X.

Since $X\mathbf{v}$ can be written as a linear combination of columns of X,

$$(\mathbf{y} - X\hat{\beta}) \cdot X\mathbf{v} = 0$$

for all $\mathbf{v} \in \mathbb{R}^{p+1}$.

We are now ready to show that $\hat{\mathbf{y}} = X(X^TX)^{-1}X^T\mathbf{y} = X\hat{\beta}$ gives the best-fit model. Suppose that we have a linear model described by a vector β'. Then,

$$\begin{aligned}
\|\mathbf{y} - X\beta'\|^2 &= \|(\mathbf{y} - X\hat{\beta}) + \big(X(\hat{\beta} - \beta')\big)\|^2 \\
&= \Big((\mathbf{y} - X\hat{\beta}) + \big(X(\hat{\beta} - \beta')\big)\Big) \cdot \Big((\mathbf{y} - X\hat{\beta}) + \big(X(\hat{\beta} - \beta')\big)\Big) \\
&= (\mathbf{y} - X\hat{\beta}) \cdot (\mathbf{y} - X\hat{\beta}) + 2(\mathbf{y} - X\hat{\beta}) \cdot \big(X(\hat{\beta} - \beta')\big) + \big(X(\hat{\beta} - \beta')\big) \cdot \big(X(\hat{\beta} - \beta')\big) \\
&= \|\mathbf{y} - X\hat{\beta}\|^2 + 0 + \|X(\hat{\beta} - \beta')\|^2 \\
&\geqslant \|\mathbf{y} - X\hat{\beta}\|^2.
\end{aligned}$$

Therefore $\hat{\mathbf{y}} = X\hat{\beta}$ gives the best-fit model, and since $X(\hat{\beta} - \beta') \neq \mathbf{o}$ for any $\beta' \neq \hat{\beta}$, it is the unique best-fit model as well.

The last thing we need to show is that H is a matrix of an orthogonal projection. First, since

$$H^2 = X(X^TX)^{-1}\cancel{X^TX(X^TX)^{-1}}X^T = H,$$

H is a matrix of a projection. Now, using the same method as in question 6., for any vector $\mathbf{v} \in \mathbb{R}^N$,

$$X^T(H\mathbf{v} - \mathbf{v}) = -X^T(\mathbf{v} - X(X^TX)^{-1}X^T\mathbf{v}) = \mathbf{o}.$$

Therefore, in the same way as (3), $H\mathbf{v} - \mathbf{v}$ is orthogonal to every column of X. Since

$$H\mathbf{v} = X\left((X^TX)^{-1}X^T\mathbf{v}\right)$$

can be written as linear combination of vectors in X, so

$$H\mathbf{v} \cdot (H\mathbf{v} - \mathbf{v}) = 0$$

so H is a matrix of an orthogonal projection. This ends the theory part.

4.3 Practice Questions

In this part we will apply the theory to predict body height based on hand span, foot length, and femur length.

8. *In the Google spreadsheet I provided, anonymously, enter for yourself X_1 = "hand span", X_2 = "length of foot", X_3 = "body height"*

Done!

9. *Copy the collected data into a spreadsheet of your own.*

We will use Mathematica; this is done using the following.

```
In[5]:= raw = Import["/Users/student/Downloads/MA661-2018 Regression.xlsx"][[1]]
Out[5]= {{X1,X2,X3,Y},{Hand(cm),Foot(cm),Femur(cm),Height(cm)},{18.1,25.5,48.,173.},
 {18.9,23.4,45.,166.3},{19.2,22.,48.,168.},{19.1,22.3,48.2,166.},{20.3,29.,53.,174.},
 {20.4,25.,45.6,173.},{20.5,24.8,43.,179.},{18.7,22.5,54.5,178.},{20.5,23.9,53.5,176.},
 {19.5,26.,45.5,173.5},{20.3,26.,46.5,175.},{20.3,25.,53.,185.},{20.,22.,44.,156.},
 {20.,26.,48.,174.},{15.7,21.1,45.,170.},{20.5,25.5,47.5,178.5}}
```

10. *Using the above theory, create a linear best-fit model to preduct Y from X_1, X_2 and X_3. Use technology as appropriate. Which input provides the best predictor for Y?*

To remove headers, we remove the first two items in `raw`:

```
In[6]:= data = Delete[Delete[raw,2],1]
Out[6]= {{18.1,25.5,48.,173.},{18.9,23.4,45.,166.3},{19.2,22.,48.,168.},
 {19.1,22.3,48.2,166.},{20.3,29.,53.,174.},{20.4,25.,45.6,173.},{20.5,24.8,43.,179.},
 {18.7,22.5,54.5,178.},{20.5,23.9,53.5,176.},{19.5,26.,45.5,173.5},{20.3,26.,46.5,175.},
 {20.3,25.,53.,185.},{20.,22.,44.,156.},{20.,26.,48.,174.},{15.7,21.1,45.,170.},
 {20.5,25.5,47.5,178.5}}
```

Looks good! Next we find the vectors $\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}$, for hand span, foot length, and femur length respectively.

```
In[9]:= X1=data[[All,1]];
In[10]:= X2=data[[All,2]];
In[11]:= X3=data[[All,3]];
In[12]:= Y=data[[All,4]];
```

Then we generate the vector $\mathbf{X_0}$ with all elements being 1. We make sure that $\mathbf{X_0}$ have the same length (dimension) as $\mathbf{X_1}$ by

```
In[13]:= X0=ConstantArray[1,Length[X1]];
```

Next we combine `X0,X1,X2,X3` to create X.

```
In[15]:= X=Transpose[{X0, X1, X2, X3}];
```

Finally we find β using $\beta = (X^T X)^{-1} X^T$:

```
In[17]:= beta=Inverse[Transpose[X].X].Transpose[X].Y
Out[17]= {106.816,-0.056452,1.30874,0.733368}
```

Therefore, we get a best-fit linear model of

$$\hat{\mathbf{y}} = 106.816\mathbf{1} - 0.056452\mathbf{X_1} + 1.30874\mathbf{X_2} + 0.733368\mathbf{X_3},$$

that is, the height y is predicted from hand span x_1, foot length x_2, and femur length x_3 as

$$\hat{y} = 106.816 - 0.056452x_1 + 1.30874x_2 + 0.733368x_3.$$

Since the coefficient 1.30874 of foot length is largest, foot length is the best predictor of height.

Finally, here is a comparison of the predicted height and actual height (all units are cm).

```
In[27]:= comparisonTable=Transpose[{X1,X2,X3,Y,X.beta,Y-X.beta}];
 TableForm[Prepend[comparisonTable,{"Hand", "Foot", "Femur", "Height",
 "PredictedHeight","Residual"}]]
Out[28]=
```

Hand	Foot	Femur	Height	Predicted	Residual
18.1	25.5	48.	173.	174.369	-1.36886
18.9	23.4	45.	166.3	169.375	-3.07525
19.2	22.	48.	168.	169.726	-1.72619
19.1	22.3	48.2	166.	170.271	-4.27113
20.3	29.	53.	174.	182.492	-8.49208
20.4	25.	45.6	173.	171.825	1.17543
20.5	24.8	43.	179.	169.65	9.34958
18.7	22.5	54.5	178.	175.176	2.82433
20.5	23.9	53.5	176.	176.173	-0.172921
19.5	26.	45.5	173.5	173.111	0.389225
20.3	26.	46.5	175.	173.799	1.20102
20.3	25.	53.	185.	177.257	7.74286
20.	22.	44.	156.	166.748	-10.7476
20.	26.	48.	174.	174.916	-0.915968
15.7	21.1	45.	170.	166.546	3.45419
20.5	25.5	47.5	178.5	173.867	4.63331

The residual sum of squares (RSS) is 412.765, found using the following code:

```
errbeta=Y-X.beta;
Sum[errbeta[[i]]^2,{i,Length[X1]}]
Out[62]= 412.765
```

This completes the project.

5 Row operations

5.1 Introduction

Row reduction is an algorithm for solving systems of linear equations.

Any system of linear equations can be written in matrix form $A\mathbf{x} = \mathbf{b}$. To solve this system, only A and $\mathbf{b}$ are necessary, therefore it suffices to work with the augmented coefficient matrix $[A \mid \mathbf{b}]$.

There are three operations that can be performed on this matrix without changing the solution of the modified system, and can reduce the system to a normal form from which the solution is immediately apparent. The operations are:

Definition 5.1 (Row operations).
- Multiplying a row by a nonzero number
- Adding a multiple of a row onto another row
- Exchanging two rows

and the resulting normal form is called *row-reduced echelon form*, and it is unique to the original system.

Definition 5.2.
A matrix is in *echelon form* if

(i) In every row, the first nonzero entry is 1, called a *pivotal* 1.

(ii) The pivotal 1 of a lower row is always to the right of the pivotal 1 of a higher row

(iii) In every column that contains a pivotal 1, all other entries are 0.

(iv) Any rows consisting entirely of 0's are at the bottom.

Example 5.3. These matrices are in echelon form:

$$\begin{bmatrix} \underline{1} & 0 & 0 & -3 \\ 0 & \underline{1} & 0 & -2 \\ 0 & 0 & \underline{1} & 0 \end{bmatrix}, \begin{bmatrix} 0 & \underline{1} & 3 & 0 & 0 & 3 & 0 & -4 \\ 0 & 0 & 0 & \underline{1} & -2 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \underline{1} & 2 \end{bmatrix}.$$

These are not:

$$\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Example 5.4. The following is a series of row operations on a matrix, reducing it to an echelon form.

$$\begin{bmatrix} 1 & 2 & 3 & 1 \\ -1 & 1 & 0 & 2 \\ 1 & 0 & 1 & 2 \end{bmatrix} \xrightarrow[\text{R3}\to\text{R3}-\text{R1}]{\text{R2}\to\text{R2}+\text{R1}} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 3 & 3 & 3 \\ 0 & -2 & -2 & 1 \end{bmatrix}$$

$$\xrightarrow{\text{R2}\to\text{R2}/3} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & -2 & -2 & 1 \end{bmatrix}$$

$$\xrightarrow[\text{R3}\to\text{R3}+2\text{R2}]{\text{R1}\to\text{R1}-2\text{R2}} \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

$$\xrightarrow{\text{R3}\to\text{R3}/3} \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[\text{R2}\to\text{R2}-\text{R3}]{\text{R1}\to\text{R1}+2\text{R3}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Exercise 5.5. Reduce the following matrices to echelon form:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 & 5 \\ 2 & 3 & 0 & -1 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

Solution.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \xrightarrow{\text{R2}\to\text{R2}-4\text{R1}} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \end{bmatrix}$$

$$\xrightarrow{\text{R2}\to-\text{R2}/3} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix}$$

$$\xrightarrow{\text{R1}\to\text{R1}-2\text{R2}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & 5 \\ 2 & 3 & 0 & -1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \xrightarrow{\text{R2}\to\text{R2}-2\text{R1}} \begin{bmatrix} 1 & 2 & 3 & 5 \\ 0 & -1 & -6 & -11 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

$$\xrightarrow{\text{R2}\to-\text{R2}} \begin{bmatrix} 1 & 2 & 3 & 5 \\ 0 & 1 & 6 & 11 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

$$\xrightarrow[\text{R3}\to\text{R3}-\text{R2}]{\text{R1}\to\text{R1}-2\text{R2}} \begin{bmatrix} 1 & 0 & -9 & -17 \\ 0 & 1 & 6 & 11 \\ 0 & 0 & -4 & -8 \end{bmatrix}$$

$$\xrightarrow{\text{R3}\to-\text{R3}/4} \begin{bmatrix} 1 & 0 & -9 & -17 \\ 0 & 1 & 6 & 11 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

$$\xrightarrow[\text{R2}\to\text{R2}-6\text{R3}]{\text{R1}\to\text{R1}+9\text{R3}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

5.2 Applications

Applications of row reduction:

- Solve systems of linear equations, of course
- Checking linear independence of vectors $(\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k})$: row-reduce $[\mathbf{v_1} \cdots \mathbf{v_k}]$ and see if it reduces to a matrix giving the unique trivial solution. This can also be used to find bases for kernels.
- Finding inverses of a matrix A: by solving $AX = I$ for $X = [\mathbf{x_1} \cdots \mathbf{x_k}]$. Since only the coefficients matter when row-reducing, we can row-reduce $[A \mid I]$. If A is invertible, this will row-reduce to $[I \mid A^{-1}]$.

Exercise 5.6. Row-reduce to find the inverse of

$$\begin{bmatrix} 1 & 0 & -2 \\ 2 & -1 & 0 \\ 0 & 1 & 3 \end{bmatrix}.$$

Solution.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -2 & 1 & 0 & 0 \\ 2 & -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 3 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R2 \to R2-2R1} \left[\begin{array}{ccc|ccc} 1 & 0 & -2 & 1 & 0 & 0 \\ 0 & -1 & 4 & -2 & 1 & 0 \\ 0 & 1 & 3 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{R2 \to -R2} \left[\begin{array}{ccc|ccc} 1 & 0 & -2 & 1 & 0 & 0 \\ 0 & 1 & -4 & 2 & -1 & 0 \\ 0 & 1 & 3 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{R3 \to R3-R2} \left[\begin{array}{ccc|ccc} 1 & 0 & -2 & 1 & 0 & 0 \\ 0 & 1 & -4 & 2 & -1 & 0 \\ 0 & 0 & 7 & -2 & 1 & 1 \end{array}\right]$$

$$\xrightarrow{R3 \to R3/7} \left[\begin{array}{ccc|ccc} 1 & 0 & -2 & 1 & 0 & 0 \\ 0 & 1 & -4 & 2 & -1 & 0 \\ 0 & 0 & 1 & -\frac{2}{7} & \frac{1}{7} & \frac{1}{7} \end{array}\right]$$

$$\xrightarrow[R2 \to R2+4R3]{R1 \to R1+2R3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{7} & \frac{2}{7} & \frac{2}{7} \\ 0 & 1 & 0 & \frac{6}{7} & -\frac{3}{7} & \frac{4}{7} \\ 0 & 0 & 1 & -\frac{2}{7} & \frac{1}{7} & \frac{1}{7} \end{array}\right] \qquad \blacksquare$$

Row operations also have another application:

- Find bases for images:
  - Since row operations do not change the set of solutions, columns in the original matrix are linearly independent iff they are linearly independent in the echelon form.
  - To find a basis of im A, row-reduce A, determine all pivotal columns (containing a pivotal 1) and select the corresponding columns from A.

Here's a summary from the book. See more on p.198.

Theorem 5.7 (Linear independence and span; 2.4.5 in book)**.** *Let* $\mathbf{v_1}, \ldots, \mathbf{v_n}$ *be vectors in* $\mathbb{R}^n$*, and let* A *be the* $n \times k$ *matrix* $[\mathbf{v_1} \cdots \mathbf{v_k}]$*. Then*

(a) $\mathbf{v_1}, \ldots, \mathbf{v_n}$ *are linearly independent iff the r.r.e.f matrix* $\tilde{A}$ *has a pivotal* 1 *in every column.*

(b) $\mathbf{v_1}, \ldots, \mathbf{v_n}$ *spans* $\mathbb{R}^n$ *iff the r.r.e.f matrix* $\tilde{A}$ *has a pivotal* 1 *in every row.*

5.3 Elementary matrices

Each row operation can also be described as a multiplication by *elementary matrices* (Def. 2.3.5 in the book)

Definition 5.8.
There are three types of elementary matrices:

- *Type 1 elementary matrix* $E_1(i,x)$: the identity matrix I, except that the entry (i,i) is x. It corresponds to multiplying row i by x.
- *Type 2 elementary matrix* $E_2(i,j,x)$: the identity matrix I, except that the entry (i,j) is x. It corresponds to adding x times row j to row i
- *Type 3 elementary matrix* $E_3(i,j)$, $i \neq j$: the identity matrix I, except that the entries $(i,i),(j,j)$ are 0 and entries $(i,j),(j,i)$ are 1. It corresponds to swapping rows i and j.

Clearly, elementary matrices are invertible (and their inverses are obvious) Since elementary matrices correspond to row operations, every invertible matrix A can be written as a multiplication of elementary matrices. This is useful theoretically. For example, if we want to show $\det AB = \det A \det B$, it suffices to show that $\det EB = \det E \det B$ for elementary matrices E. Here, Mathematica can be useful in sounding the validity of conjectures to be proved (= i.e. it's easy to just try some cases in Mathematica)

6 Eigenstuff

6.1 Change of basis

Come to think of it, there is nothing special about the standard bases in $\mathbb{R}^n$. We can easily work on other bases as well, and we will use the notation $\mathbf{x} = \begin{bmatrix} \vdots \end{bmatrix}_B$ to denote $\mathbf{x}$ written in form of the basis B.

However, how do we know what happens when we change the basis? Turns out when we want to change from basis B_1 to B_2, there is a *change of basis matrix* $[\Phi_{B_1 \to B_2}]$ that takes any $[\mathbf{x}]_{B_1}$ and returns $[\mathbf{x}]_{B_2}$.

Basically, we are finding the matrix of $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ w.r.t. the two bases. To do so, take each vector from B_1 as input and express the result as linear combinations of vectors in B_2. The scalars will form the columns of $[\Phi_{B_1 \to B_2}]$. This sounds pretty confusing, so let's work out an example.

Example 6.1 (Changing basis in $\mathbb{R}^2$)**.** To change basis from $B_1 = (\mathbf{v_1}, \mathbf{v_2})$ to $B_2 = (\mathbf{w_1}, \mathbf{w_2})$, we solve

$$\begin{aligned} \mathbf{v_1} &= \alpha_{11}\mathbf{w_1} + \alpha_{12}\mathbf{w_2} \\ \mathbf{v_2} &= \alpha_{21}\mathbf{w_1} + \alpha_{22}\mathbf{w_2} \end{aligned}$$

to get

$$[\Phi_{B_1 \to B_2}] = \begin{bmatrix} \alpha_{11} & \alpha_{21} \\ \alpha_{12} & \alpha_{22} \end{bmatrix}.$$

This is equivalent to row-reducing

$$\left[B_2 \middle| B_1\right] \to \left[I \middle| [\Phi_{B_1 \to B_2}]\right].$$

The following is not covered in class, don't cite it in the test!

It is easy to see that this works for $\mathbb{R}^n$ as well, and also that it is equivalent to

$$\Phi_{B_1 \to B_2} = [B_1][B_2]^{-1}.$$

This might feel a bit counterintuitive (why not $[B_1]^{-1}[B_2]$?) but consider this: obviously, $\Phi_{B_1 \to I}$ is $[B_1]$ and *not* $[B_1]^{-1}$. Another view: if we change the basis from $[B_0]$ (standard basis, = I) to B_1, it should have *the opposite effect* of taking the transformation T defined by B_1 because doing both cancels out: if you change the basis to B_1 while transforming everything by the matrix B_1 as well, you just get the same thing. Back to class notes.

Given $T : \mathbb{R}^n \to \mathbb{R}^n$ and matrix $[T]$, how do we find $[T]_B$? Answer:

$$T_B = \Phi_{B_0 \to B} \circ T \circ \Phi_{B \to B_0}$$

so

$$[T]_B = [B]^{-1}[T][B]. \tag{4}$$

Explanation: if you do this on $[\mathbf{v}]_B$, the $\Phi_{B \to B_0}$ take it to B_0, where you can apply T, and then use $\Phi_{B_0 \to B}$ to take it back to B.

More explanation: the function T_B looks like this

```
func T_B (v_B: vector_B) -> vector_B {
    v = v_B as vector
    return T(v) as vector_B
}
```

and `vector_B` is not the same 'datatype' as `vector`.

Let's do a numerical example.

Example 6.2. Find the rotation matrix by $\frac{\pi}{6}$ w.r.t. basis $B_1 = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right)$. We have

$$[R_{\frac{\pi}{6}}]_{B_1} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}.$$

Whoops we got the same matrix. This makes sense though as B_1 is basically just a rotation and scaling of the standard basis.

Equation (6.1) works when we change between any two bases B_1, B_2 too:

$$[T]_{B_2} = [\Phi_{B_1 \to B_2}][T]_{B_1}[\Phi_{B_1 \to B_2}]^{-1}.$$

How do we classify the matrices that can represent the same transformation? There is a term for it.

Definition 6.3.
Two matrices S, T are said to be *similar* if there is an invertible matrix U such that

$$S = UTU^{-1}.$$

6.2 Eigenvalues and eigenvectors

Definition 6.4.
Let V be a vector space, U a subspace, $T \in (V)$. Then U is *invariant* under T is $TU \subseteq U$. In other words: if $\mathbf{v} \in U$ then $T(\mathbf{v}) \in U$.

Exercise 6.5. If U is a 1-dimensional invariant subspace of V, then there exists $\lambda \in \mathbb{R}$ such that $T_{|u}(\mathbf{v}) = \lambda\mathbf{v}$ for $\mathbf{v} \in U$.

Solution. Let $(\mathbf{b})$ be a basis of U and choose λ from $T(\mathbf{b}) = \lambda\mathbf{b}$. Then, for any $\mathbf{v} = \alpha\mathbf{b} \in U$, $T(\mathbf{v}) = T(\alpha\mathbf{b}) = \alpha T(\mathbf{b}) = \alpha\lambda\mathbf{b} = \lambda\mathbf{v}$. ∎

If we have a basis of $B = (\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_n})$ of V satisfying: for all $i = 1, \ldots, n$, there exists a scalar λ_i such that $T(\mathbf{b_i}) = \lambda_i\mathbf{b_i}$. What would $[T]_B$ look like?

$$[T]_B = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Given $T \in \mathcal{L}(V)$, is there a λ satisfying $T(\mathbf{v}) = \lambda\mathbf{v}$ for some $\mathbf{v} \neq \mathbf{o}$? Let's see. If $T(\mathbf{v}) = \lambda\mathbf{v}$ then

$$(T - \lambda I)\mathbf{v} = \mathbf{o},$$

so $T - \lambda I$ has a non-trivial kernel, and is not invertible, so

$$\det(T - \lambda I) = 0,$$

which gives an equation for λ (which is a polynomial of degree n.) This brings us to *the* definition:

Definition 6.6 (Eigenvalues and eigenvectors)**.**
If $T \in \mathcal{L}(V)$ and there exists $\mathbf{v} \neq \mathbf{o}$ such that $T(\mathbf{v}) = \lambda\mathbf{v}$ for some $\lambda \in \mathbb{R}$, then $\mathbf{v}$ is called an *eigenvector* to *eigenvalue* λ.

Note that each eigenvector spans a 1-dimensional invariant subspace for T.

Definition 6.7.
If V has a basis consisting of eigenvectors, this basis is called the *eigenbasis*.

As always, there is a numerical example.

Example 6.8. To find eigenvalues and eigenvectors of

$$T = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}.$$

We want

$$0 = \det(T - \lambda I) = \begin{vmatrix} 2-\lambda & 1 \\ 1 & -1-\lambda \end{vmatrix} = (2-\lambda)(-1-\lambda) - 1 = \lambda^2 - \lambda - 3,$$

so $\lambda = \dfrac{1 \pm \sqrt{13}}{2}$ with corresponding eigenvectors $\begin{bmatrix} 1 \\ \lambda - 2 \end{bmatrix}$ (these eigenvectors can be found by solving $(T - \lambda I)\mathbf{v} = 0$.)

Theorem 6.9 (Eigenvectors with distinct eigenvalues are linearly independent)**.** *If $A : V \to V$ is a linear transformation, and $\mathbf{v_1}, \ldots, \mathbf{v_k}$ are eigenvectors of A with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then $\mathbf{v_1}, \ldots, \mathbf{v_k}$ are linearly independent.*

Proof (by contradiction, from book). If $\mathbf{v_1}, \ldots, \mathbf{v_k}$ are not linearly independent, then there is a first vector $\mathbf{v_j}$ that is a linear combination of the earlier ones. Thus we can write

$$\mathbf{v_j} = a_1\mathbf{v_1} + \cdots + a_{j-1}\mathbf{v_{j-1}},$$

where at least one coefficient is not zero, say a_i. Apply $\lambda_j I - A$ to both sides to get

$$\mathbf{o} = (\lambda_j I - A)\mathbf{v_j} = (\lambda_j I - A)(a_1\mathbf{v_1} + \cdots + a_{j-1}\mathbf{v_{j-1}}) = \sum_{i=1}^{j-1} a_i(\lambda_j - \lambda_i)\mathbf{v_i}.$$

which is a linear combination of $\mathbf{v_1}, \ldots, \mathbf{v_{j-1}}$ without all coefficients being zero; this is the desired contradiction. ■

Complex and repeated eigenvalues are studied in the context of Jordan normal forms, which is beyond the range of this class.

We have been studying eigenvalues without even thinking if they actually exist or not. Fortunately, they do:

Theorem 6.10. *Eigenvalues always exist.*

Proof. First, note that we can apply polynomials to square matrices. For any nonzero $\mathbf{w} \in \mathbb{C}^n$, there is a smallest m such that $A^m\mathbf{w}$ is a linear combination of $\mathbf{w}, A\mathbf{w}, \ldots, A^{m-1}\mathbf{w}$. (Such an m has to exist because $\mathbb{C}^n$ has finite dimension n.) Write the linear combination as

$$a_0\mathbf{w} + a_1A\mathbf{w} + \cdots + a_{m-1}A^{m-1}\mathbf{w} + A^m\mathbf{w} = \mathbf{o}.$$

Then we can use the coefficients to define a polynomial

$$p(t) := a_0 + a_1t + \cdots + a_{m-1}t^{m-1} + t^m$$

which satisfies $p(A)\mathbf{w} = \mathbf{o}$ and is the lowest degree polynomial with this property. (Wow, this resembles minimal polynomials.)

By FTA, $p(t)$ has at least one root λ, so we can write $p(t) = (t - \lambda)q(t)$ for some q with $\deg q = m - 1$. Define $\mathbf{v} = q(A)\mathbf{w}$. Then,

$$(A - \lambda I)\mathbf{v} = p(A)\mathbf{w} = \mathbf{o},$$

so $A\mathbf{v} = \lambda\mathbf{v}$. Now we only have to check that $\mathbf{v} \neq 0$, but this is evident from $q \neq 0$ and $\deg q < \deg p$. ■

The next question is how to find an eigenvalue. That is easy. Just row reduce $\mathbf{w}, A\mathbf{w}, \ldots, A^n\mathbf{w}$. Use the first non-pivotal column to get a linear combination/polynomial, and use the above.

Mathematica implementation:

```
a = RandomInteger[{-10, 10}, {7, 7}];
MatrixForm[a]
w = Table[0, {Length[a]}];
w[[1]] = 1;
```

```
paw = Transpose[Table[MatrixPower[a, k].w, {k, 0, Length[a]}]];
rrd = RowReduce[paw];
b = rrd[[All, -1]]
p[t_] := Sum[-b[[i]] t^(i - 1), {i, 1, Length[b]}] + t^Length[b]
p[t]
lambda1 = t /. NSolve[p[t] == 0, t, Reals][[1]]
Eigenvalues[1. a]
q[t_] := PolynomialQuotient[p[t], (t - lambda1), t];
q[t]
qcoeff = CoefficientList[q[t], t];
qa = Sum[qcoeff[[i]]*MatrixPower[a, i - 1], {i, 1, Length[qcoeff]}];
v = Normalize[qa.w];
MatrixForm[v]
Eigenvectors[1. a]
```

6.3 Spectral Theorem

Let A be a diagonalizable matrix, that is, a matrix similar to a diagonal matrix. It follows that A has an eigenbasis B where $D = [A]_B$ is a diagonal matrix. Therefore, for $\Phi = \Phi_{B_0 \to B}$,

$$\Phi A \Phi^{-1} = D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

Consider the k^{th} power of A:

$$A^k = (\Phi^{-1} D \Phi)^k = \Phi^{-1} D \cancel{\Phi\Phi^{-1}} D \cdots \cancel{\Phi\Phi^{-1}} D \Phi = \Phi^{-1} D^k \Phi = \Phi^{-1} \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k) \Phi.$$

What is e^A?

$$\begin{aligned} e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} &= \sum_{k=0}^{\infty} \Phi^{-1} \frac{\operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)}{k!} \Phi \\ &= \Phi^{-1} \left(\sum_{k=0}^{\infty} \frac{\operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)}{k!} \right) \Phi = \Phi^{-1} \operatorname{diag}(e_1^{\lambda}, \ldots, e_k^{\lambda}) \Phi. \end{aligned}$$

This leads us to the Spectral Theorem, named after the *spectral* of a matrix, which is the collection of its eigenvalues. The spectral itself is apparently named after literal spectrum in physics based on some quantum physics phenomenon.

Theorem 6.11 (Spectral Theorem)**.** *Let A be a symmetric $n \times n$ matrix. Then, A has an orthonormal eigenbasis, and all its eigenvalues are real.*

6.4 Properties of eigenvalues

Eigenvalues of matrices have several properties:

(i) If λ is an eigenvalue of A, then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1} provided $\lambda \neq 0$ and A^{-1} exists.

(ii) A is invertible iff 0 is not an eigenvalue of A.

(iii) If λ is an eigenvalue of A then it is an eigenvalue of A^T.

(iv) If A is diagonalizable with eigenvalues $(\lambda_1, \ldots, \lambda_n)$ then det $A = \prod_{i=1}^n \lambda_i$.

(v) Let $A = [a_{ij}]$ and define trace of A tr $(A) = \sum_{i=1}^n a_{ii}$. Then, tr $(A) = \sum_{i=1}^n \lambda_i$.

The last two properties imply that the trace and determinants are invariant under a change of basis.

Exercise 6.12. Show properties (i) to (iv) above for general matrices, and (v) for 2×2 matrices.

Solution. Here we assume A has size $n \times n$. We will prove (v) for general matrices anyway.

(i) Let $\mathbf{v}$ be an eigenvector to λ. We have $A\mathbf{v} = \lambda\mathbf{v}$, so

$$\frac{1}{\lambda}(\lambda\mathbf{v}) = A^{-1}A\mathbf{v} = A^{-1}(\lambda\mathbf{v}),$$

hence $\frac{1}{\lambda}$ is an eigenvalue of A^{-1}.

(ii) 0 is an eigenvalue of $A \iff$ there is a vector $\mathbf{v} \neq \mathbf{o}$ with $A\mathbf{v} = 0 \iff A$ has a nontrivial kernel $\iff A$ is not invertible (because A is a square matrix)

(iii) For any matrix A_0, λ is an eigenvalue of a matrix A_0 iff λ is a root of the polynomial $\det(A_0 - xI)$. Since $(A - xI)^T = A^T - xI$,

$$\det(A^T - xI) = \det(A - xI)^T = \det(A - xI)$$

as polynomials, so λ is an eigenvalue of A iff it is an eigenvalue of A^T.

(iv) Let U be the change of basis matrix from the standard basis to the eigenbasis of A. We have

$$\prod_{i=1}^n \lambda_i = \det \operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \det UAU^{-1} = \det U \det A \det U^{-1} = \det A.$$

(v) Consider $\det(A - xI)$ as a polynomial in x. In the Leibniz sum of the determinant, the diagonal term gives

$$\prod_{i=1}^n (a_{ii} - x) = (-1)^n x^n - (-1)^n \sum_{i=1}^n a_{ii} + p(x)$$

for some p with $\deg p \leqslant n-2$. As any other permutation misses the diagonal in at least two rows, and thus has degree $\leqslant n - 2$,

$$\det(A - xI) = (-1)^n x^n - (-1)^n \sum_{i=1}^n a_{ii} + q(x)$$

for some q with $\deg q \leqslant n - 2$. As $\lambda_1, \ldots, \lambda_n$ are the roots of $\det(A - xI)$, by Vieta,

$$\sum_{i=1}^n \lambda_i = -\frac{-(-1)^n \sum_{i=1}^n a_{ii}}{(-1)^n} = \sum_{i=1}^n a_{ii} = \operatorname{tr} A.$$

6.5 Test review

- Suppose we have bases $B_1 = (\mathbf{b_1}, \mathbf{b_2}), B_2 = (\mathbf{c_1}, \mathbf{c_2})$.
- To find the matrix $\Phi_{B_1 \to B_2}$, we use row reduction:

$$\left[B_2 \middle| B_1\right] \to \left[I \middle| \Phi_{B_1 \to B_2}\right].$$

This matrix sends a vector $\mathbf{v}$ written in B_1 to the vector $\mathbf{v}$ written in B_2, that is, if $\alpha_1 \mathbf{b_1} + \alpha_2 \mathbf{b_2} = \mathbf{v} = \beta_1 \mathbf{c_1} + \beta_2 \mathbf{c_2}$ then

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}_{B_2} = \Phi_{B_1 \to B_2} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}_{B_1}.$$

- The matrix for the inverse transformation is simply the inverse matrix:

$$\Phi_{B_2 \to B_1} = \Phi^{-1}_{B_1 \to B_2}.$$

- If we have a transformation $T \in \mathcal{L}(V)$ then

$$[T]_{B_2} = \Phi_{B_1 \to B_2} [T]_{B_1} \Phi_{B_2 \to B_1}.$$

- Eigenbasis: if B_E is an eigenbasis for T then

$$[T]_{B_E} = \text{diag}\,(\lambda_1, \ldots, \lambda_n)$$

where $T(\mathbf{v_i}) = \lambda_i \mathbf{v_i}$ for $\mathbf{v_i} \in B_E$.
- If T is symmetric then Spectral theorem guarantees an orthonormal eigenbasis.
- To actually find an eigenbasis, solve $\det(A - \lambda I) = 0$ for λ and find all linearly independent eigenvectors for all λ_i. If all eigenvalues are distinct, we have an eigenbasis. Also, eigenvectors to distinct eigenvalues are always linearly independent.
- If we have repeated eigenvalues, there may or may not be linearly independent eigenvectors. For example,

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

has 0 as an eigenvalue of multiplicity 3 with $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ as the only linearly independent eigenvector. However, I has 1 as an eigenvalue of multiplicity 3 but with $(\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3})$ as an eigenbasis.
- Suppose A has an eigenbasis. We can write $A = U\text{diag}\,(\lambda_1, \ldots, \lambda_n)U^{-1}$ for invertible U. Then,

$$A^k = U\text{diag}\,(\lambda_1^k, \ldots, \lambda_n^k)U^{-1}.$$

7 Multilinear algebra

7.1 Quadratic forms

A *quadratic form* is a homogeneous polynomial of degree 2. For example,

$$q(x,y,z) = c_{11}x^2 + c_{12}xy + c_{13}xz + c_{22}y^2 + c_{23}yz + c_{33}z^2$$

is a quadratic form in three variables.

We can relate quadratic forms to vectors and matrices as follows:

$$q(x,y,z) = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} c_{11} & \frac{1}{2}c_{12} & \frac{1}{2}c_{13} \\ \frac{1}{2}c_{12} & c_{22} & \frac{1}{2}c_{23} \\ \frac{1}{2}c_{13} & \frac{1}{2}c_{23} & c_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

In general, for a quadratic form q, there is a symmetric matrix M_q such that

$$q(\mathbf{x}) = \mathbf{x}^T M_q \mathbf{x}.$$

Definition 7.1 ((posi/negative) (in/semi)definite)**.**
Let $q(\mathbf{x})$ be a quadratic form. The q is called

- *positive definite* if $q(\mathbf{x}) \geqslant 0$ for all $\mathbf{x}$, and $q(\mathbf{x}) = 0$ iff $\mathbf{x} = 0$.
- *positive semidefinite* if $q(\mathbf{x}) \geqslant 0$ for all $\mathbf{x}$.
- *negative definite* if $q(\mathbf{x}) \leqslant 0$ for all $\mathbf{x}$, and $q(\mathbf{x}) = 0$ iff $\mathbf{x} = 0$.
- *negative semidefinite* if $q(\mathbf{x}) \leqslant 0$ for all $\mathbf{x}$.
- *indefinite* otherwise

Clearly, the definiteness of a (real) diagonal matrix M can be determined easily by the signs of its elements. Therefore, to determine the definiteness of each q, we want to write M_q in a basis B which makes it a diagonal matrix.

Since M_q is symmetric, the Spectral Theorem guarantees us an orthonomal eigenbasis B. We show that choosing B works. Since B is orthonormal, $\Phi = \Phi_{B_0 \to B}$ has orthonormal columns, so $\Phi^T \Phi = I$. Set $\mathbf{y} = \Phi\mathbf{x}$. Then,

$$q_A(\mathbf{y}) = (\Phi^{-1}\mathbf{y})^T (\Phi^{-1}\mathbf{y}) = \mathbf{y}^T \left[(\Phi^{-1})^T A \Phi^{-1}\right] \mathbf{y}$$

so we want $(\Phi^{-1})^T A \Phi^{-1}$ to be diagonal. However, $(\Phi^{-1})^T = (\Phi^T)^T = \Phi$, so

$$(\Phi^{-1})^T A \Phi^{-1} = \Phi A \Phi^{-1} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

is diagonal.

7.2 k-forms

Definition 7.2.
A *k-form* on $\mathbb{R}^n$ is an anti-symmetric multilinear function $(\mathbb{R}^n)^k$ taking in k vectors and returning a number. The set of k-forms is donated as $A_c^k(\mathbb{R}^n)$.

Example 7.3. The 2-form $dx_1 \wedge dx_2$ takes in two vectors and outputs the determinant of the square matrix formed by the first and second entries of the vectors.

$$dx_1 \wedge dx_2 \left(\begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 1 \\ 2 \end{bmatrix} \right) = \begin{vmatrix} 1 & 3 \\ 2 & -2 \end{vmatrix} = -8$$

Exercise 7.4 (6.1.3 in book)**.** Compute the following numbers:

(a) $dx_1 \wedge dx_4 \left(\begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ -1 \\ 2 \end{bmatrix} \right) = \begin{vmatrix} 1 & 1 \\ 2 & 2 \end{vmatrix} = 0.$

(d) $dx_1 \wedge dx_2 \wedge dx_2$ (something) $= 0$ because dx_2 is repeated.

Definition 7.5.
$dx_1 \wedge dx_2$ is an example of *elementary k-forms*: those of the form

$$dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}, \quad i_1 < i_2 < \cdots < i_k.$$

Example 7.6. There are 2^4 elementary k-forms in $\mathbb{R}^4$, corresponding to subsets of $\{1, 2, 3, 4\}$:

$$1, dx_1, \cdots, dx_4, dx_1 \wedge dx_2, \cdots, dx_3 \wedge dx_4, dx_1 \wedge dx_2 \wedge dx_3, \cdots, dx_2 \wedge dx_3 \wedge dx_4, dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4.$$

Since k-forms can be added together or multiplied by a scalar, they form a vector space. As the last example illustrates, the vector space of k-forms in $\mathbb{R}^n$ has dimension $\binom{n}{k}$.

Let x be a point, and $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$ be vectors in $\mathbb{R}^n$. Then, $P_x(\mathbf{v_1}, \ldots, \mathbf{v_k})$ is the parallelogram spanned by $(\mathbf{v_1}, \ldots, \mathbf{v_k})$ attached to point x.

A *k-form field* in $\mathbb{R}^n$ is a set of k-forms in which the scalars depend on $(x_1, \ldots, x_n)$. Essentially, it takes a parallelogram in.

Example 7.7. $\phi = 3dx_1 \wedge dx_3$ is a 2-form; $\omega = e^{x+y} dx \wedge dy$ is a 2-form field.

Example 7.8. $\cos(xz) dx \wedge dy$ is a 2-form field on $\mathbb{R}^3$. As an example of evaluation,

$$\cos(xz) dx \wedge dy \left(P_{\begin{pmatrix} 1 \\ 2 \\ \pi \end{pmatrix}} \left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} \right) \right) = \cos(1 \cdot \pi) \begin{vmatrix} 1 & 2 \\ 0 & 2 \end{vmatrix} = -2.$$

We used the *wedge* $\wedge$ symbol in our notation for k-forms. This represents a wedge product, which has the following formal definition, which I guess comes from the need to make sure that $\wedge$ preserves antisymmetry as well as multilinearity.

Definition 7.9.
The *wedge product* of the forms $\phi \in A_c^k(\mathbb{R}^n)$ and $\omega \in A_c^\ell(\mathbb{R}^n)$ is the element $\phi \wedge \omega \in A_c^{k+\ell}(\mathbb{R}^n)$ defined by

$$(\phi \wedge \omega)(\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{k+l}}) = \sum_{\substack{\sigma \text{ permutes } \{1,2,\ldots,k+\ell\} \\ \sigma(1) < \cdots < \sigma(k) \\ \sigma(k+1) < \cdots < \sigma(k+\ell)}} \operatorname{sgn}(\sigma) \phi(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(k)}) \omega(\mathbf{v}_{\sigma(k+1)}, \ldots, \mathbf{v}_{\sigma(k+\ell)}).$$

Example 7.10. The wedge product of $\phi \in A_c^2(\mathbb{R}^n)$ and $\omega \in A_c(\mathbb{R}^n)$ is

$$\phi \wedge \omega(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}) = \phi(\mathbf{v_1}, \mathbf{v_2})\omega(\mathbf{v_3}) - \phi(\mathbf{v_1}, \mathbf{v_3})\omega(\mathbf{v_2}) + \phi(\mathbf{v_2}, \mathbf{v_3})\omega(\mathbf{v_1}).$$

Exercise 7.11 (6.1.11 in book)**.** Let ϕ and ψ be 2-forms. Write out $\phi \wedge \psi(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4})$.

Answer:

$$\begin{aligned}\phi \wedge \psi(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}) = {} & \phi(\mathbf{v_1}, \mathbf{v_2})\psi(\mathbf{v_3}, \mathbf{v_4}) - \phi(\mathbf{v_1}, \mathbf{v_3})\psi(\mathbf{v_2}, \mathbf{v_4}) + \phi(\mathbf{v_1}, \mathbf{v_4})\psi(\mathbf{v_2}, \mathbf{v_3}) \\ & + \phi(\mathbf{v_2}, \mathbf{v_3})\psi(\mathbf{v_1}, \mathbf{v_4}) - \phi(\mathbf{v_2}, \mathbf{v_4})\psi(\mathbf{v_1}, \mathbf{v_3}) + \phi(\mathbf{v_3}, \mathbf{v_4})\psi(\mathbf{v_1}, \mathbf{v_1})\end{aligned}$$