Math 121A — Linear Algebra

Neil Donaldson

Winter 2022

1 Vector Spaces

1.1 Introduction: What is Linear Algebra and why should we care?

Linear algebra is the study of *vector spaces* and *linear maps* between them. We'll formally define these concepts later, though they should be familiar from a previous class.
A function, or map, $\mathrm{T} : V \to W$ between vector spaces is *linear* if for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ and all scalars λ, we have the properties:

(a) $\mathrm{T}(\mathbf{v}_1 + \mathbf{v}_2) = \mathrm{T}(\mathbf{v}_1) + \mathrm{T}(\mathbf{v}_2)$

(b) $\mathrm{T}(\lambda \mathbf{v}_1) = \lambda \mathrm{T}(\mathbf{v}_1)$

Examples 1.1. You have seen many examples of these in your mathematical career.

1. $\mathrm{T}(x) = 3x$ defines a linear map $\mathrm{T} : \mathbb{R} \to \mathbb{R}$.

   More generally, $V = \mathbb{R}^m$, $W = \mathbb{R}^n$, with T being multiplication by a real $n \times m$ matrix.

2. Differentiation: Let $\mathrm{T} = \frac{\mathrm{d}}{\mathrm{d}x}$ be the usual differential operator and V the vector space of differentiable functions $f : \mathbb{R} \to \mathbb{R}$. More generally, T could be a *linear differential operator* such as $\mathrm{T} = \frac{\mathrm{d}^2}{\mathrm{d}x^2} + 2x\frac{\mathrm{d}}{\mathrm{d}x} + x^2 + 1$ whence

   $$\mathrm{T}(y) = y'' + 2xy' + (x^2 + 1)y$$

   The standard methods for solving *linear differential equations* seen in a lower-division class are based on linear algebra.

3. Integration: let V be a vector space of integrable functions then $\mathrm{T}(f) = \int_a^x f(t)\,\mathrm{d}t$ defines a linear map to a vector space of continuous functions.

The ubiquity of linear structures is one reason to study linear algebra. Another is that *linear problems* often admit systematic techniques that give us at least a fighting chance of finding a solution. By contrast, *non-linear* problems are typically much more difficult: if such can be solved, it is often due to some trickery or luck.

What makes linear problems easy? The core idea is to use simple solutions as building blocks to construct more complex solutions. Here is a, hopefully familiar, example.

Example 1.2. The power law tells us how to integrate monomials. The linearity of integration allows us to combine these building blocks to compute the integral of any polynomial:

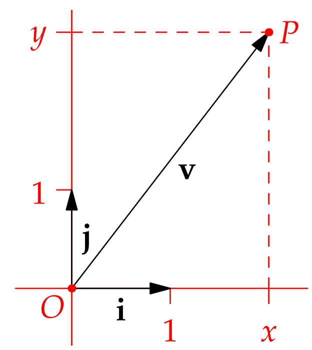
$$\int x^2 + 5x^3 \,\mathrm{d}x = \int x^2 \,\mathrm{d}x + 5\int x^3 \,\mathrm{d}x \qquad \text{(linearity)}$$
$$= \frac{1}{3}x^3 + \frac{5}{4}x^4 + c \qquad \text{(power law)}$$

By contrast, the integration of *products* is a non-linear problem. The fact that

$$\int e^x \sin x \,\mathrm{d}x \neq \left[\int e^x \,\mathrm{d}x\right]\left[\int \sin x \,\mathrm{d}x\right]$$

and the resulting need for integration-by-parts is a major source of difficulty in freshman calculus.

A Brief Review of $\mathbb{R}^2$ and $\mathbb{R}^3$ In these standard spaces, we often visualize vectors as arrows. In the picture, the *vector* $\mathbf{v}$ points from the *origin* O with *co-ordinates* $(0,0)$ to the point $P = (x, y)$. Writing $\mathbf{i}$, $\mathbf{j}$ for the standard basis vectors, there are several common notations for $\mathbf{v}$:

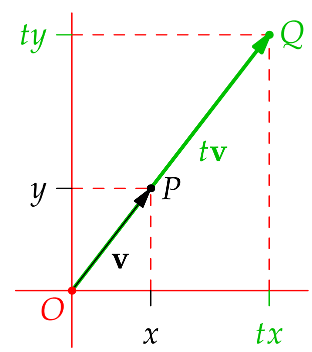


$$\mathbf{v} = \overrightarrow{OP} = \begin{pmatrix} x \\ y \end{pmatrix} = x\mathbf{i} + y\mathbf{j} = x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Column vector notation helps distinguish a *vector* from a *point* (x, y): we call x and y the *components* of the column vector $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$.

The *vector space* $\mathbb{R}^2$ is simply the set of all such vectors. There is no need for a vector to have its tail at the origin, only direction and magnitude matter. In $\mathbb{R}^3$ things are similar, a point has three co-ordinates and we need the three standard basis vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$.

Scalar multiplication involves lengthening or contracting a vector by a real multiple: the vector $t\mathbf{v}$ has components tx and ty and we write



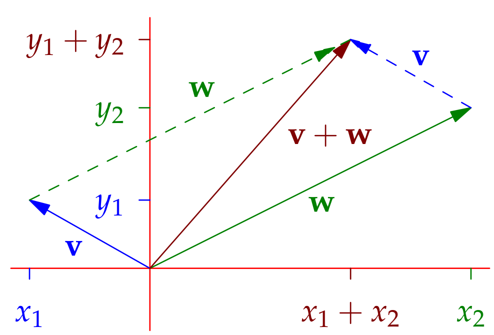
$$t\mathbf{v} = \overrightarrow{OQ} = \begin{pmatrix} tx \\ ty \end{pmatrix} = tx\mathbf{i} + ty\mathbf{j}$$

Note that if $t < 0$, then $t\mathbf{v}$ points in the opposite direction to $\mathbf{v}$.

Vector addition is defined by the *parallelogram law.* Simply add components: if $\mathbf{v}_1 = \left(\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right)$ and $\mathbf{v}_2 = \left(\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right)$, then

$$\mathbf{v}_1 + \mathbf{v}_2 := \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} = (x_1 + x_2)\mathbf{i} + (y_1 + y_2)\mathbf{j}$$

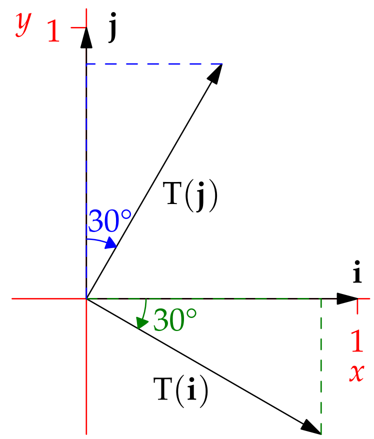
The intuitive *nose-to tail* interpretation of vector addition is immediate.



Example 1.3 (Rotation and the standard basis). We finish by reviewing an approach you should have seen in a previous course. By considering how to transform a basis, we obtain a complete formula for a linear map.

Consider the function $T : \mathbb{R}^2 \to \mathbb{R}^2$ which rotates a point 30° clockwise around the origin. You should believe, though a proof is tricky at the present, that T is indeed linear. To discover a formula for T it is enough to consider what it does to the standard basis $\{\mathbf{i}, \mathbf{j}\}$ of $\mathbb{R}^2$.

This is because if $\mathbf{v} = \binom{x}{y}$ is any vector, then, by linearity



$$\mathbf{v} = x\mathbf{i} + y\mathbf{j} \implies T(\mathbf{v}) = xT(\mathbf{i}) + yT(\mathbf{j})$$

Using the picture and a little trigonometry, you should be convinced that that

$$T(\mathbf{i}) = \begin{pmatrix} \cos 30° \\ -\sin 30° \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} \\ -\frac{1}{2} \end{pmatrix} \qquad T(\mathbf{j}) = \begin{pmatrix} \sin 30° \\ \cos 30° \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}$$

from which we obtain

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2}x + \frac{1}{2}y \\ -\frac{1}{2}x + \frac{\sqrt{3}}{2}y \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

where we've written the final expression as a matrix multiplication.[a]

[a] This is one of the advantages of column vector notation. Indeed one of the major goals of the course is to see that every linear map between finite-dimensional vector spaces can be represented in such a fashion.

Because of linearity, we were able to completely determine T merely by understanding how it acted on the basis vectors $\mathbf{i}$ and $\mathbf{j}$. This property is *not* shared by non-linear functions. For instance, if $|\mathbf{v}|$ returns the length of a vector $\mathbf{v} \in \mathbb{R}^2$, then the function

$$f : \mathbb{R}^2 \to \mathbb{R}^2 : \mathbf{v} \mapsto \left(|\mathbf{v}|^2 + 1\right)\mathbf{v} \qquad (*)$$

is non-linear. Simply knowing that $f(\mathbf{i}) = 2\mathbf{i}$ and $f(\mathbf{j}) = 2\mathbf{j}$ is insufficient to completely understand the function.

Exercises 1.1

1. Using the same approach as in Example 1.3, explicitly find a formula for the linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$ which *reflects* in the line $y = x$.
2. A linear map is not the same thing as a straight line! Explain why the function $f : x \mapsto 3x + 2$ is non-linear.
3. Give a reason why the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined in $(*)$ above is non-linear.
4. (a) Give an algebraic proof (use components!) of the distributive law $\lambda(\mathbf{v} + \mathbf{w}) = \lambda\mathbf{v} + \lambda\mathbf{w}$ in the vector space $\mathbb{R}^2$.
   (b) Give a *pictorial* argument for the distributive law?
   (*Hint: Consider similar triangles or parallelograms and channel your inner Euclid…*)

1.2 Vector Spaces: Basic Results, Examples and Subspaces

Vector spaces generalize the intuitive structure of $\mathbb{R}^2$, where identities such as commutativity

$$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$$

are geometrically obvious. The *axioms* of a vector space merely assert that such identities hold generally.

Definition 1.4. Let V be a non-empty set (elements *vectors*) and $\mathbb{F}$ a *field* (elements *scalars*), and suppose we have two operations:

Vector Addition If $\mathbf{v}$ and $\mathbf{w}$ are vectors, we can form the *sum* $\mathbf{v} + \mathbf{w}$.

Scalar Multiplication If $\mathbf{v}$ is a vector and λ a scalar, we can form the *product* $\lambda\mathbf{v}$.

We say that V is a *vector space* over $\mathbb{F}$ if the following axioms are satisfied:[a]

G1:	Closure under addition	$\forall \mathbf{v}, \mathbf{w} \in V, \quad \mathbf{v} + \mathbf{w} \in V$
G2:	Associativity of addition	$\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V, \quad (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
G3:	Identity for addition	$\exists \mathbf{0} \in V, \quad \text{such that} \quad \forall \mathbf{v} \in V, \quad \mathbf{v} + \mathbf{0} = \mathbf{v}$
G4:	Inverse for addition	$\forall \mathbf{v} \in V, \exists -\mathbf{v} \in V, \quad \text{such that} \quad \mathbf{v} + (-\mathbf{v}) = \mathbf{0}$
G5:	Commutativity of addition	$\forall \mathbf{v}, \mathbf{w} \in V, \quad \mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$
A1:	Closure under scalar multiplication	$\forall \mathbf{v} \in V, \lambda \in \mathbb{F}, \quad \lambda\mathbf{v} \in V$
A2:	Identity for scalar multiplication	$\forall \mathbf{v} \in V, \quad 1\mathbf{v} = \mathbf{v}$
A3:	Action of scalar multiplication	$\forall \lambda, \mu \in \mathbb{F}, \mathbf{v} \in V, \quad \lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v}$
D1:	Distributivity I	$\forall \mathbf{v}, \mathbf{w} \in V, \lambda \in \mathbb{F}, \quad \lambda(\mathbf{v} + \mathbf{w}) = \lambda\mathbf{v} + \lambda\mathbf{w}$
D2:	Distributivity II	$\forall \mathbf{v} \in V, \lambda, \mu \in \mathbb{F}, \quad (\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$

[a] This is easier to remember if you've studied group theory: the 'G' axioms say that $(V, +)$ is an Abelian group, the 'A' axioms say that the field $\mathbf{F}$ has a *left action* on V. The distributivity axioms explain how the two operations interact.

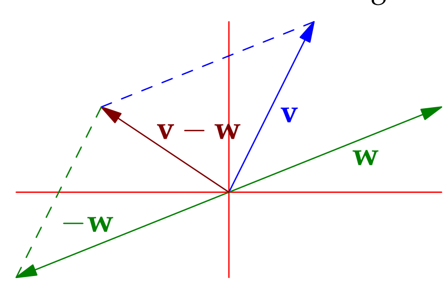
Notation You can use another notation (e.g. $\vec{v}$ or $\underline{v}$) for abstract vectors, but use something: distinguishing vectors and scalars helps avoid common mistakes like dividing by a vector. This notation might not be appropriate in certain examples, (e.g. polynomials, matrices) so take extra care.

Fields A *field* $\mathbb{F}$ is a set which behaves very like the real numbers under addition and multiplication. In almost all examples, $\mathbb{F}$ will be either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$. The symbols 0 and 1 (e.g. axiom A2) refer to the additive and multiplicative identities in $\mathbb{F}$. Be careful to distinguish the scalar $0 \in \mathbb{F}$ from the *zero vector* $\mathbf{0} \in V$.

Inverses and subtraction Subtraction of vectors is taken to mean *addition of the inverse*, namely

$$\mathbf{v} - \mathbf{w} := \mathbf{v} + (-\mathbf{w})$$

In $\mathbb{R}^2$ this can be viewed pictorially.



Essentially every example we will encounter falls into one of two classes.

Theorem 1.5 (Matrices & Sets of Functions). *Let $\mathbb{F}$ be a field.*

1. *The set $M_{m\times n}(\mathbb{F})$ of $m \times n$ matrices with entries in $\mathbb{F}$*

$$M_{m\times n}(\mathbb{F}) = \left\{ \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} : a_{ij} \in \mathbb{F} \right\}$$

forms a vector space over $\mathbb{F}$ under component-wise addition and scalar multiplication: given matrices $A = (a_{ij})$ and $B = (b_{ij})$ and $\lambda \in \mathbb{F}$, the ij^{th} entries of the matrices $A + B$ and λA are

$$(A+B)_{ij} := a_{ij} + b_{ij}, \qquad (\lambda A)_{ij} := \lambda a_{ij} \qquad (*)$$

2. *Let D be a set and V a vector space over $\mathbb{F}$. The set of functions*

$$\mathcal{F}(D, V) = \{f : D \to V\}$$

forms a vector space over $\mathbb{F}$ with addition and scalar multiplication defined by[a]

$$(f+g)(x) := f(x) + g(x), \qquad (\lambda f)(x) := \lambda(f(x))$$

[a] The *function* $f + g \in \mathcal{F}(D, V)$ is defined by what it does to an element $x \in D$. In particular, $f(x) \in V$ is *not* a function. However, it is acceptable to write 'the function $f(x)$', just make sure you know that this is an abuse of notation.

To prove the theorem, each axiom (G1–5, A1–3, D1,2) should be checked explicitly for each part of the theorem: this is tedious! For instance, axiom D2 may be verified for matrices as follows:

$$\big((\lambda+\mu)A\big)_{ij} = (\lambda+\mu)a_{ij} = \lambda a_{ij} + \mu a_{ij} = (\lambda A)_{ij} + (\mu A)_{ij}$$

Definitions $(*)$ provide the red equalities, while the blue is distributivity in the field $\mathbb{F}$.

Examples 1.6. 1. The *column vectors* (*n-tuples*) are a special case: $\mathbb{F}^n := M_{n\times 1}(\mathbb{F})$. E.g., in $\mathbb{R}^3$

$$2\begin{pmatrix}1\\0\\-4\end{pmatrix} + 7\begin{pmatrix}-1\\2\\1\end{pmatrix} = \begin{pmatrix}2\\0\\-8\end{pmatrix} + \begin{pmatrix}-7\\14\\7\end{pmatrix} = \begin{pmatrix}-5\\14\\-1\end{pmatrix}$$

2. In $M_{2\times 3}(\mathbb{C})$, we have

$$\begin{pmatrix}1 & i & 0\\ -3 & 1-i & 2+3i\end{pmatrix} + i\begin{pmatrix}2 & -3 & 1\\ 3-i & 0 & 2\end{pmatrix} = \begin{pmatrix}1+2i & -2i & i\\ -2+3i & 1-i & 2+5i\end{pmatrix}$$

3. A field is a vector space over itself! In particular, $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is a vector space. For instance, if f, g are defined by $f(x) = x^2$ and $g(x) = \sin x$, then $4f - 2g$ is the function given by

$$(4f - 2g)(x) = 4x^2 - 2\sin x$$

We'll shortly restrict to certain types of functions (e.g. *continuous functions*, *differentiable functions*, *polynomials*) and see that these also form vector spaces.

Basic Results Here we gather several basic facts about vector spaces that you'll use without thinking. Since these are not axioms, they do require *proof*.

Lemma 1.7.

1. *Cancellation law:* $\mathbf{x}+\mathbf{z}=\mathbf{y}+\mathbf{z} \implies \mathbf{x}=\mathbf{y}$
2. *Uniqueness of Identity: The zero vector* $\mathbf{0}$ *posited in axiom G3 is unique.*
3. *Uniqueness of Inverse: Given* $\mathbf{v}\in V$*, the vector* $-\mathbf{v}$ *posited in axiom G4 is unique.*
4. *Scalar multiplication by zero:* $\forall \mathbf{v}\in V$*, we have* $0\mathbf{v}=\mathbf{0}$*.*
5. *Action of negatives:* $\forall \mathbf{v}\in V,\ \lambda\in\mathbb{F}$*, we have* $(-\lambda)\mathbf{v}=-(\lambda\mathbf{v})$*.*
6. *Action on zero vector:* $\forall\lambda\in\mathbb{F}$*, we have* $\lambda\mathbf{0}=\mathbf{0}$*.*

Proof. We prove number 4, leaving the remainder as exercises: they are easiest if tackled in order! Since $0=0+0$ in $\mathbb{F}$, apply axioms D2, G3, G5 and the cancellation law to see that

$$\begin{aligned}
&0\mathbf{v}=(0+0)\mathbf{v}=0\mathbf{v}+0\mathbf{v} && \text{(Distributivity D2)}\\
\implies\ &\mathbf{0}+0\mathbf{v}=0\mathbf{v}+0\mathbf{v} && \text{(Identity G3 and Commutativity G5)}\\
\implies\ &\mathbf{0}=0\mathbf{v} && \text{(Cancellation law)}
\end{aligned}$$

■

Subspaces As in other areas of algebra (subgroup, subring, subfield, etc.) the prefix *sub* means that an object is a *subset*, while retaining the algebraic structure of the original set.

Definition 1.8. Let V be a vector space over a field $\mathbb{F}$.

A non-empty subset $W\subseteq V$ is a *subspace* of V (written $W\le V$) if it is a vector space over the *same* field $\mathbb{F}$ with respect to the *same* addition and scalar multiplication operations as V.

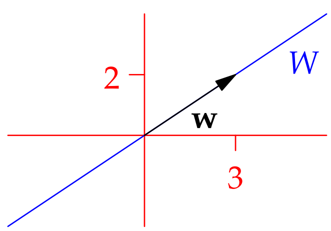
A subspace is *proper* if it is a proper subset (i.e., $W\neq V$).

The *trivial subspace* of V is the point set $\{\mathbf{0}\}$.

The subset approach allows us to quickly construct many more examples.

Example 1.9. Consider the line containing $\mathbf{w}=\binom{3}{2}\in\mathbb{R}^2$:

$$W:=\left\{a\mathbf{w}=\begin{pmatrix}3a\\2a\end{pmatrix}:a\in\mathbb{R}\right\}$$



It is almost trivial, if tedious, to check that W satisfies the axioms and is therefore a subspace of $\mathbb{R}^2$. For instance:

$$\text{G1:}\quad \begin{pmatrix}3a\\2a\end{pmatrix}+\begin{pmatrix}3b\\2b\end{pmatrix}=\begin{pmatrix}3(a+b)\\2(a+b)\end{pmatrix}\in W$$

by appealing to the distributivity laws in $\mathbb{R}$.

Thankfully, as the next theorem shows, there is no need to check all the axioms to determine when we have a subspace.

Theorem 1.10. *Suppose W is a non-empty subset of a vector space V over $\mathbb{F}$. Then W is a subspace of V if and only if it is closed under addition and scalar multiplication:*[a]

G1: $\forall \mathbf{w}_1, \mathbf{w}_2 \in W$, *we have* $\mathbf{w}_1 + \mathbf{w}_2 \in W$.

A1: $\forall \mathbf{w} \in W,\ \lambda \in \mathbb{F}$, *we have* $\lambda \mathbf{w} \in W$.

[a] That $\mathbf{w}_1 + \mathbf{w}_2$ and $\lambda \mathbf{w}$ lie in W and not just V is what makes these genuine conditions.

There are two common variants of this result: feel free to use these in examples if you prefer.

- Explicitly verify axiom G3 ($\mathbf{0}_V \in W$) instead of checking the non-emptiness of W. You still need to check that W is a subset of V!
- Combine the closure axioms into a single statement: $\forall \mathbf{x}, \mathbf{y} \in W,\ \lambda \in \mathbb{F}$, we have $\lambda \mathbf{x} + \mathbf{y} \in W$.

Proof. If W is a subspace, then it is a vector space and so axioms G1 and A1 hold as written.

Conversely, assume that W is a non-empty subset of V satisfying G1 and A1. Reread the axioms (Definition 1.4) and observe that all except perhaps G1, G3, G4 and A1 hold on any subset of V. Under our assumptions therefore, it remains only to verify G3 and G4.

G3: Choose any $\mathbf{w} \in W$. By Lemma 1.7 (part 4) and axiom A1 (for W!), we see that the zero vector of V satisfies

$$\mathbf{0}_V = 0\mathbf{w} \in W$$

Since $\mathbf{0}_V$ satisfies axiom G3 for V, it necessarily does on any subset: we therefore have $\mathbf{0}_W = \mathbf{0}_V$.

G4: Given $\mathbf{w} \in W$, let $-\mathbf{w} \in V$ be its additive inverse in V. Now observe that

$$-\mathbf{w} = (-1)\mathbf{w} \in W$$

by Lemma 1.7 (part 5) and axiom A1. ■

Examples 1.11. 1. Returning to Example 1.9, recall that we already checked axiom G1. Moreover,

- $\mathbf{0} = 0\mathbf{w} \in W$ so that W is non-empty.
- A1: $\lambda(a\mathbf{w}) = (\lambda a)\mathbf{w} \in W$ by axiom A3.

so that W is a subspace of $\mathbb{R}^2$. Alternatively, if $\lambda \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in W$, then $\exists a, b \in \mathbb{R}$ for which

$$\lambda \mathbf{x} + \mathbf{y} = \lambda a \mathbf{w} + b\mathbf{w} = (\lambda a + b)\mathbf{w} \in W$$

2. For any field $\mathbb{F}$, let $P_n(\mathbb{F}) = \{a_0 + a_1 x + \cdots + a_n x^n : a_i \in \mathbb{F}\}$ be the set of polynomials of degree $\leq n$ with coefficients in $\mathbb{F}$. By considering axioms G1 and A1, this is plainly a subspace of the space of functions $\mathcal{F}(\mathbb{F}, \mathbb{F})$:

$$\lambda(a_0 + a_1 x + \cdots + a_n x^n) + (b_0 + b_1 x + \cdots + b_n x^n) = (a_0 + b_0) + \cdots + (a_n + b_n)x^n$$

Non-emptiness is guaranteed by considering the *zero polynomial* $0(x) = 0 + 0x + \cdots + 0x^n$.

More generally, if $n \leq m$, then $P_n(\mathbb{F}) \leq P_m(\mathbb{F}) \leq P(\mathbb{F})$, where the last denotes the space of *all* polynomials of any degree.

3. If $U \subseteq \mathbb{R}$ is an interval, then $V = \mathcal{F}(U, \mathbb{R})$ is a vector space over $\mathbb{R}$. The subset $C(U, \mathbb{R})$ of continuous functions is a subspace of V. Indeed, as is verified in any analysis course,

   If $f, g : I \to \mathbb{R}$ are continuous and $\lambda \in \mathbb{R}$, then $\lambda f + g : I \to \mathbb{R}$ is continuous.

   This also extends to sets of differentiable functions, etc.

4. The *trace* $\operatorname{tr} : M_n(\mathbb{F}) \to \mathbb{F}$ of an $n \times n$ matrix is defined by summing the main diagonal:

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}$$

   The subset of *trace-free* matrices is denoted

$$\mathfrak{sl}_n(\mathbb{F}) = \{A \in M_n(\mathbb{F}) : \operatorname{tr} A = 0\}$$

   It is easy to check that $\mathfrak{sl}_n(\mathbb{F}) \leq M_n(\mathbb{F})$:

$$\operatorname{tr}(\lambda A + B) = \sum_{i=1}^{n} \lambda a_{ii} + b_{ii} = \lambda \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = \lambda \operatorname{tr} A + \operatorname{tr} B = 0$$

Intersections and Direct Sums Since vector spaces are sets, we may take intersections...

Theorem 1.12. *If V and W are subspaces of some vector space U, then their* intersection *$V \cap W$ is a subspace of both V and W.*

Proof. Since V and W are subspaces of U, they both contain $\mathbf{0}$ and so $V \cap W$ is non-empty.
Now suppose $\mathbf{x}, \mathbf{y} \in V \cap W$ and $\lambda \in \mathbb{F}$. Since V and W are both vector spaces, they are closed under addition and scalar multiplication (in U!): in particular,

$$\mathbf{x} + \mathbf{y} \in V, \qquad \mathbf{x} + \mathbf{y} \in W, \qquad \lambda \mathbf{x} \in V, \qquad \lambda \mathbf{x} \in W$$

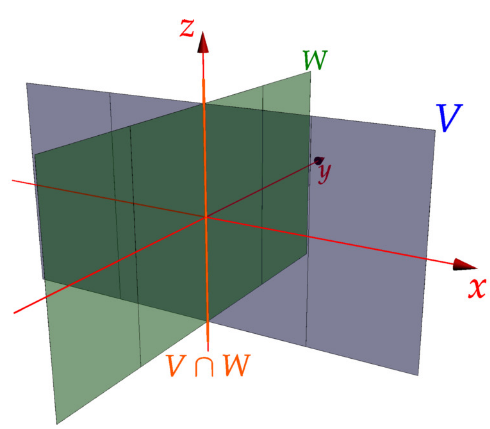
But then $\mathbf{x} + \mathbf{y} \in V \cap W$ and $\lambda \mathbf{x} \in V \cap W$, whence $V \cap W$ is closed and thus a subspace. ■

Example 1.13. Suppose that

$$V = \{x\mathbf{i} + z\mathbf{k} : x, z \in \mathbb{R}\}$$
$$W = \{y\mathbf{j} + z\mathbf{k} : y, z \in \mathbb{R}\}$$

are the xz- and yz-planes respectively. Plainly, V and W are subspaces of $\mathbb{R}^3$ with intersection the z-axis

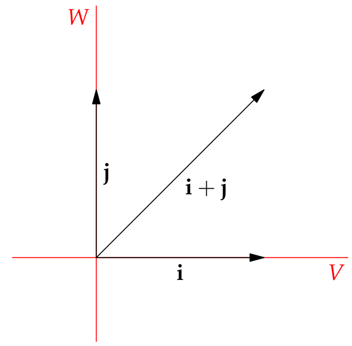
$$V \cap W = \{z\mathbf{k} : z \in \mathbb{R}\}$$



Attempting the same thing for unions results in a problem.
For a simple counterexample, let

$$V = \{x\mathbf{i} : x \in \mathbb{R}\} \qquad W = \{y\mathbf{j} : y \in \mathbb{R}\}$$

be the x- and y-axes in $\mathbb{R}^2$, whose intersection is the trivial subspace $V \cap W = \{\mathbf{0}\}$. Their union



$$V \cup W = \{x\mathbf{i}, y\mathbf{j} : x, y \in \mathbb{R}\}$$

is not a subspace of $\mathbb{R}^2$ since it is not closed under addition:

$$\mathbf{i} \in V \ \text{ and } \ \mathbf{j} \in W \ \text{ but } \ \mathbf{i} + \mathbf{j} \notin V \cup W$$

Instead we search for the *smallest* vector space containing $V \cup W$.

Definition 1.14. Suppose V and W are subspaces of U. Their *sum* is the set

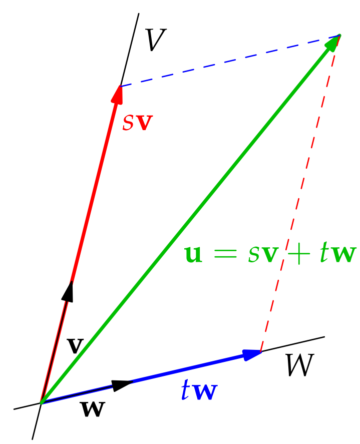
$$V + W := \{\mathbf{v} + \mathbf{w} : \mathbf{v} \in V, \mathbf{w} \in W\}$$

In addition, if $V \cap W = \{\mathbf{0}\}$, we call this the *direct sum* and write $V \oplus W$.

Examples 1.15. 1. The x- and y-axes, $V = \{x\mathbf{i} : x \in \mathbb{R}\}$ and $W = \{y\mathbf{j} : y \in \mathbb{R}\}$ are clearly subspaces of $\mathbb{R}^2$ with trivial intersection. It is immediate that

$$V \oplus W = \{x\mathbf{i} + y\mathbf{j} : x, y \in \mathbb{R}\} = \mathbb{R}^2$$

2. More generally, let $V = \{s\mathbf{v} : s \in \mathbb{R}\}$ and $W = \{t\mathbf{w} : t \in \mathbb{R}\}$ be distinct, non-trivial subspaces of $\mathbb{R}^2$ (i.e. $\mathbf{v}, \mathbf{w}$ are non-parallel). Observe:



- If $V \cap W \neq \{\mathbf{0}\}$, then $\exists s, t \neq 0$ such that $s\mathbf{v} = t\mathbf{w}$, whence $\mathbf{v}$ and $\mathbf{w}$ would be parallel: contradiction.
- Writing $\mathbf{v} = \binom{a}{b}$ and $\mathbf{w} = \binom{c}{d}$, we see that for any given $\mathbf{u} = \binom{x}{y} \in \mathbb{R}^2$,

$$\mathbf{u} = s\mathbf{v} + t\mathbf{w} \iff \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix}$$
$$\iff \begin{pmatrix} s \\ t \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -c \\ -b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Since $\mathbf{v}, \mathbf{w}$ are non-parallel, we are not dividing by zero: every $\mathbf{u}$ can be written in the form $s\mathbf{v} + t\mathbf{w}$ and so $\mathbb{R}^2 = V + W$.

Putting both parts together, we conclude that $\mathbb{R}^2 = V \oplus W$ is a direct sum.

Indeed we see that every $\mathbf{u} \in \mathbb{R}^2$ can be written *uniquely* in terms of the subspaces: as the next result shows, this is a defining property of direct sums.

Theorem 1.16. *Let V, W be subspaces of U with trivial intersection. Then:*

1. *$V \oplus W$ is a subspace of U.*
2. *V and W are subspaces of $V \oplus W$.*
3. *If X is a subspace of U such that both V and W are subspaces of X, then $V \oplus W \leq X$.*
4. *$U = V \oplus W \iff \forall \mathbf{u} \in U, \exists$unique $\mathbf{v} \in V$ and $\mathbf{w} \in W$ such that $\mathbf{u} = \mathbf{v} + \mathbf{w}$.*

The proofs are exercises. Note how the third property says that $V \oplus W$ is the smallest space containing V and W, while the fourth says that direct sums are synonymous with unique decompositions.

Exercises 1.2 1. Let $S = \{0, 1\}$ and $\mathbb{F} = \mathbb{R}$. In the vector space of functions $\mathcal{F}(S, \mathbb{R})$, let

$$f(t) = 2t + 1, \qquad g(t) = 1 + 4t - 2t^2, \qquad h(t) = 5^t + 1$$

Show that $f = g$ and $f + g = h$.

2. (a) If $p(x) = 2x + 3x^2$ and $q(x) = 4 - x^2$, compute the polynomial $p(x) + 3q(x)$.
   (b) Explain why the set of degree two polynomials with coefficients in $\mathbb{R}$ is *not* a vector space.
   (c) Prove explicitly that $P_1(\mathbb{R})$ is a subspace of $P_3(\mathbb{R})$.

3. Prove parts 2, 3 and 6 of Lemma 1.7.

4. Consider the vector space $\mathbb{C}^2 = \{\left(\begin{smallmatrix} w \\ z \end{smallmatrix}\right) : w, z \in \mathbb{C}\}$ over the field $\mathbb{C}$ of complex numbers.
   (a) Show that $\mathbf{v} = \left(\begin{smallmatrix} i \\ 2+3i \end{smallmatrix}\right)$ and $\mathbf{w} = \left(\begin{smallmatrix} 1+2i \\ 7+4i \end{smallmatrix}\right)$ are parallel.
   (b) $\mathbb{C}^2$ is automatically a vector space over $\mathbb{C}$. Prove that it is a vector space over $\mathbb{R}$.

5. Is $M_{m \times n}(\mathbb{R})$ a vector space over the rational numbers $\mathbb{Q}$? Explain.

6. Let $V = \{(a_1, a_2) : a_1, a_2 \in \mathbb{R}\}$. For $(a_1, a_2), (b_1, b_2) \in V$ and $\lambda \in \mathbb{R}$, define

$$(a_1, a_2) + (b_1, b_2) := (a_1 + 2b_1, a_2 + 3b_2) \quad \text{and} \quad \lambda(a_1, a_2) = (\lambda a_1, \lambda a_2)$$

   Is V a vector space over $\mathbb{R}$ with respect to these operations? Explain.

7. Let $V = \mathbb{R}^2$, define vector addition as usual and scalar multiplication (by $\lambda \in \mathbb{R}$) by

$$\lambda \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} \lambda x \\ \lambda^{-1} y \end{pmatrix} \quad \text{if } \lambda \neq 0 \quad \text{or} \quad \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{if } \lambda = 0$$

   Is V a vector space with respect to these operations? Why/why not?

8. Prove or disprove:
   (a) $V := \{\left(\begin{smallmatrix} 4a \\ -a \end{smallmatrix}\right) : a \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.
   (b) $W := \{\left(\begin{smallmatrix} 4a+1 \\ -a \end{smallmatrix}\right) : a \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.
   (c) $X := \left\{\left(\begin{smallmatrix} 4a+b \\ -a \\ 2a-b \end{smallmatrix}\right) : a, b \in \mathbb{R}\right\}$ is a subspace of $\mathbb{R}^3$.

9. Let V be the set of differentiable real-valued functions with domain $\mathbb{R}$. Prove that V is a subspace of the set of functions $\mathcal{F}(\mathbb{R}, \mathbb{R})$.
   (*You may quote anything you like from elementary calculus without proof*)

10. With reference to Theorem 1.10, prove that properties G1 & A1 are equivalent to the combined closure property: $\forall \mathbf{x}, \mathbf{y} \in W, \lambda \in \mathbb{F}$, we have $\lambda \mathbf{x} + \mathbf{y} \in W$.

11. (a) Let $V = \{x\mathbf{i} : x \in \mathbb{R}\}$ and $W = \{y\mathbf{j} + z\mathbf{k} : y, z \in \mathbb{R}\}$ be subspaces of $\mathbb{R}^3$. Prove that $V \oplus W = \mathbb{R}^3$.
    (b) Repeat part (a) but this time with $V = \{x(\mathbf{i} + \mathbf{j}) : x \in \mathbb{R}\}$.

12. (a) Prove all four parts of Theorem 1.16.
    (b) If we drop the assumption that V and W have trivial intersection, which parts of Theorem 1.16 must be true for the *sum* $V + W$.

13. A matrix A is *symmetric* if it equals its transpose: $A^T = A$. It is *skew-symmetric* if $A^T = -A$. Let S be the set of symmetric matrices and K the set of skew-symmetric matrices in $M_2(\mathbb{R})$.

    (a) Show that S and K are subspaces of $M_2(\mathbb{R})$.
    (b) Prove or disprove: $M_2(\mathbb{R}) = S \oplus K$.
    (c) Does your argument extends to $M_n(\mathbb{R})$ and, if you've studied fields, to $M_n(\mathbb{F})$?

14. Let $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$ together with addition and multiplication modulo 5.

    (a) Prove that every non-zero element of $\mathbb{Z}_5$ has a multiplicative inverse ($\mathbb{Z}_5$ is a *field*): for all $x \in \mathbb{Z}_5 \setminus \{0\}$, there exists $y \in \mathbb{Z}_5$ such that $xy = 1$.
    (b) By part (a), $\mathbb{Z}_5^n$ is a vector space. Evaluate the expression $4\begin{pmatrix} 3 \\ 2 \end{pmatrix} + 2\begin{pmatrix} 1 \\ 4 \end{pmatrix} \in \mathbb{Z}_5^2$. For any $n \in \mathbb{N}$, how many vectors are there in $\mathbb{Z}_5^n$? (What is the cardinality of $\mathbb{Z}_5^n$?)

15. Let $V \times W = \{(\mathbf{v}, \mathbf{w}) : \mathbf{v} \in V, \mathbf{w} \in W\}$ be the Cartesian product of spaces V, W over $\mathbb{F}$.

    (a) (Briefly!) Argue that $V \times W$ is a vector space over $\mathbb{F}$ with respect to the operations

    $$(\mathbf{v}_1, \mathbf{w}_1) + (\mathbf{v}_2, \mathbf{w}_2) := (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_1 + \mathbf{w}_2) \qquad \lambda(\mathbf{v}, \mathbf{w}) := (\lambda \mathbf{v}, \lambda \mathbf{w})$$

    (b) Verify that $\hat{V} := \{(\mathbf{v}, \mathbf{0}_W) : \mathbf{v} \in V\}$ and $\hat{W} := \{(\mathbf{0}_V, \mathbf{w}) : \mathbf{w} \in W\}$ are subspaces of $V \times W$ and that $\hat{V} \oplus \hat{W} = V \times W$.

    (*$V \times W$ is an alternative definition of the direct sum $V \oplus W$, which should be familiar if you've seen* direct products *in group theory.*)

16. (Optional: should be familiar if you've studied group theory) Let W be a subspace of V over $\mathbb{F}$. For any $\mathbf{v} \in V$, define the *coset of W containing* $\mathbf{v}$ to be the set

    $$\mathbf{v} + W := \{\mathbf{v} + \mathbf{w} : \mathbf{w} \in W\}$$

    (a) If $V = \mathbb{R}^3$ and $W = \operatorname{Span}\{\mathbf{i}, \mathbf{j}\}$, describe the coset $\mathbf{k} + W$ in words.
    (b) Let V be a a general vector space. Prove that $\mathbf{v} + W$ is a subspace of V if and only if $\mathbf{v} \in W$.
    (c) Prove that $\mathbf{v}_1 + W = \mathbf{v}_2 + W \iff \mathbf{v}_1 - \mathbf{v}_2 \in W$.
    (d) Define the *quotient space* $V/W = \{\mathbf{v} + W : \mathbf{v} \in V\}$ to be the set of cosets of W in V together with the operations

    $$(\mathbf{v}_1 + W) + (\mathbf{v}_2 + W) := (\mathbf{v}_1 + \mathbf{v}_2) + W \qquad \lambda(\mathbf{v} + W) := \lambda \mathbf{v} + W$$

    Prove that addition and scalar multiplication are well-defined, and (briefly) convince yourself that V/W is a vector space over $\mathbb{F}$ under these operations.

1.3 Linear Combinations & Linear Independence

In this section we consider what vectors we can generate from a given collection using only the vector space operations af addition and scalar multiplication.

Definition 1.17. Let S be a non-empty subset of a vector space V over a field $\mathbb{F}$. A *linear combination* of vectors in S is any vector of the form

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \tag{$*$}$$

where each $\mathbf{v}_i \in S$ and each $a_i \in \mathbb{F}$. The *span* of S is the set of all linear combinations of vectors in S:

$$\operatorname{Span} S = \{a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n : n \in \mathbb{N}, a_1, \ldots, a_n \in \mathbb{F},\ \mathbf{v}_1, \ldots, \mathbf{v}_n \in S\}$$

By convention, $\operatorname{Span}\emptyset := \{\mathbf{0}\}$ is the trivial subspace.
Important: A linear combination contains *finitely many* terms—no infinite sums!

Our primary goal of this chapter is to identify the *smallest* possible spanning sets for a vector space: such a set will be called a *basis*. The full discussion is difficult and length; for the present, we consider a few simple examples of linear combinations and spanning sets.

Examples 1.18. 1. In $P_2(\mathbb{R})$, the vector $p(x) = 2 - 3x^2$ is a linear combination of the vectors $q(x) = 2x - x^2$ and $r(x) = 1 - x - x^2$, since

$$p = q + 2r$$

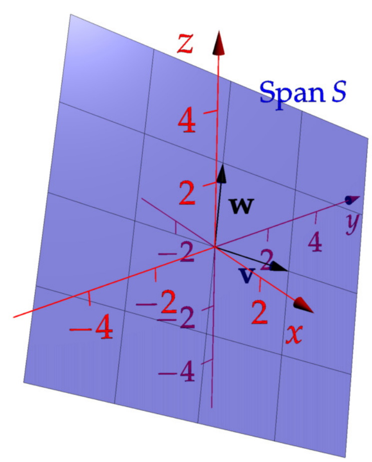
2. In Example 1.9, $W = \operatorname{Span}\{\mathbf{w}\}$. Since this is the span of a single vector, it is common to abuse notation and write $\operatorname{Span}\mathbf{w}$. In this notation, and following Definition 1.14, we see that

$$\operatorname{Span}\{\mathbf{v}, \mathbf{w}\} = \operatorname{Span}\mathbf{v} + \operatorname{Span}\mathbf{w}$$

3. Let $S = \{\mathbf{v}, \mathbf{w}\} \subseteq \mathbb{R}^3$ where $\mathbf{v} = \left(\begin{smallmatrix}1\\2\\-1\end{smallmatrix}\right)$ and $\mathbf{w} = \left(\begin{smallmatrix}-1\\1\\2\end{smallmatrix}\right)$. Then

$$\operatorname{Span} S = \left\{ a\begin{pmatrix}1\\2\\-1\end{pmatrix} + b\begin{pmatrix}-1\\1\\2\end{pmatrix} : a, b \in \mathbb{R} \right\}$$

This is the plane through the origin 'spanned by' $\mathbf{v}$ and $\mathbf{w}$: hence the use of the word *span*!



4. (a) Let $S = \{\mathbf{i}, \mathbf{k}\} \subseteq \mathbb{R}^3$. The span of S is the xz-plane

$$\operatorname{Span} S = \{a\mathbf{i} + b\mathbf{k} : a, b \in \mathbb{R}\}$$

(b) If $T = S \cup \{3\mathbf{i} - 2\mathbf{k}\} = \{\mathbf{i}, \mathbf{k}, 3\mathbf{i} - 2\mathbf{k}\} \subseteq \mathbb{R}^3$, then $\operatorname{Span} T$ remains the xz-plane. The third vector $3\mathbf{i} - 2\mathbf{k}$ is redundant since it is a linear combination of the first two. Indeed

$$a\mathbf{i} + b\mathbf{k} + c(3\mathbf{i} - 2\mathbf{k}) = (a + 3c)\mathbf{i} + (b - 2c)\mathbf{k} \in \operatorname{Span}\{\mathbf{i}, \mathbf{k}\}$$

Part of our concern in this chapter is to more carefully consider such redundancies.

The examples should suggest the following.

Lemma 1.19. *If S is a subset of a vector space V, then* $\operatorname{Span} S$ *is a subspace of V.*

Proof. This is trivial if $S = \emptyset$. Otherwise, we follow the criteria in Theorem 1.10. Let $\mathbf{x}, \mathbf{y} \in \operatorname{Span} S$ and $\lambda \in \mathbb{F}$. Then $\exists m, n \in \mathbb{N}$, $\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{y}_1, \ldots, \mathbf{y}_n \in S$, and $a_1, \ldots, a_m, b_1, \ldots, b_n \in \mathbb{F}$ such that

$$\mathbf{x} = a_1\mathbf{x}_1 + \cdots + a_m\mathbf{x}_m, \qquad \mathbf{y} = b_1\mathbf{y}_1 + \cdots + b_n\mathbf{y}_n$$

But then

$$\lambda\mathbf{x} + \mathbf{y} = \lambda a_1\mathbf{x}_1 + \cdots + \lambda a_m\mathbf{x}_m + b_1\mathbf{y}_1 + \cdots + b_n\mathbf{y}_n \in \operatorname{Span} S$$

■

Generating or Spanning Sets Part of our goal is to identify subsets S, particularly *small* subsets, of a vector space such that $\operatorname{Span} S$ is the entire space.

Definition 1.20. Let S be a subset of a vector space V. If $\operatorname{Span} S = V$, we say that S is a *spanning* or *generating set* for V. Alternatively, we say that S *spans* V or S *generates* V.

Examples 1.21.

1. $S = \{\mathbf{i}, \mathbf{j}\}$ generates $\mathbb{R}^2$, since every vector $\mathbf{v} \in \mathbb{R}$ is a linear combination $\mathbf{v} = x\mathbf{i} + y\mathbf{j}$ of vectors in S. Indeed $\mathbb{R}^2$ is essentially *defined* as $\operatorname{Span} S$!
2. Many vector spaces are defined via a spanning set: e.g. $P_3(\mathbb{R}) := \operatorname{Span}\{1, x, x^2, x^3\}$.
3. Consider $S = \{1 - 2x^2,\ 1 + x - x^2,\ 1 + 2x + x^3\} \subseteq P_3(\mathbb{R})$.
   (a) The polynomial x^3 lies in $\operatorname{Span} S$.
   For this, we need to find coefficients a, b, c such that
   $$a(1 - 2x^2) + b(1 + x - x^2) + c(1 + 2x + x^3) = x^3$$
   By equating the coefficients of $1, x, x^2$ and x^3 is it enough for us to solve a linear system
   $$\begin{cases} a + b + c = 0 \\ b + 2c = 0 \\ -2a - b = 0 \\ c = 1 \end{cases} \iff (a, b, c) = (1, -2, 1)$$
   (b) $1 + 3x + x^3 \notin \operatorname{Span} S$. It follows that S *does not generate* $P_3(\mathbb{R})$.
   This time, we need to show that there are no coefficients a, b, c such that
   $$a(1 - 2x^2) + b(1 + x - x^2) + c(1 + 2x + x^3) = 1 + 3x + x^3 \iff \begin{cases} a + b + c = 1 \\ b + 2c = 3 \\ -2a - b = 0 \\ c = 1 \end{cases}$$
   Substituting $c = 1$ in the second equation yields $b = 1$, however the remaining equations are now $a + 2 = 1$ and $-2a - 1 = 0$ which are inconsistent.

4. $S = \{1+x^2,\ 2-x^2,\ x,\ 1+4x\}$ generates the vector space $P_2(\mathbb{R})$.

Given any $a+bx+cx^2 \in P_2(\mathbb{R})$ we need to see that there exist $g,h,j,k \in \mathbb{R}$ such that

$$\begin{aligned} a+bx+cx^2 &= g(1+x^2)+h(2-x^2)+jx+k(1+4x) \\ &= g+2h+k+(j+4k)x+(g-h)x^2 \end{aligned}$$

By equating coefficients, this amounts to finding a solution (g,h,j,k) (as functions of a,b,c) to the underdetermined linear system

$$\begin{cases} g+2h+k=a \\ j+4k=b \\ g-h=c \end{cases}$$

Only one solution is required, and $k=0$, $j=b$, $g=\frac{1}{3}(a+2c)$, $h=\frac{1}{3}(a-c)$ does the trick.

Alternatively, you could try to explicitly construct the elements $1, x, x^2$ from those of S: in this situation it is fairly easy to do by inspection, e.g.,

$$1 = (1+4x)-4x, \qquad x = x, \qquad x^2 = \frac{2}{3}(1+x^2)-\frac{1}{3}(2-x^2)$$

It follows that $\{1,x,x^2\} \subseteq \operatorname{Span} S$ and so $P_2(\mathbb{R}) = \operatorname{Span}\{1,x,x^2\} \subseteq \operatorname{Span} S$. Since, plainly, $\operatorname{Span} S \subseteq P_2(\mathbb{R})$ we have equality: $\operatorname{Span} S = P_2(\mathbb{R})$.

Aside: row operations review It should be revision, but the solution to the above linear system would likely have been found very slowly in a previous class. Here are some of the details. The required system can be put in augmented matrix form:

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 4 \\ 1 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} g \\ h \\ j \\ k \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \rightsquigarrow \left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & a \\ 0 & 0 & 1 & 4 & b \\ 1 & -1 & 0 & 0 & c \end{array}\right)$$

Applying row operations, we can put this in (reduced) row echelon form:

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & \frac{1}{3} & \frac{1}{3}(a+2c) \\ 0 & 1 & 0 & \frac{1}{3} & \frac{1}{3}(a-c) \\ 0 & 0 & 1 & 4 & b \end{array}\right)$$

There is a free variable (k) here, but all solutions can easily be read off:

$$g = \frac{1}{3}(a+2c)-\frac{1}{3}k, \qquad h = \frac{1}{3}(a-c)-\frac{1}{3}k, \qquad j = b-4k$$

Choosing $k=0$ gives the solution referenced above.

Linear systems can always be tackled using augmented matrices, but it is encouraged to avoid them if you can: see, e.g., the alternative method for the last example.

Linear Dependence & Independence: when is a spanning set larger than necessary?

If $\mathbf{w} = 2\mathbf{v}$, then $\operatorname{Span}\{\mathbf{v}, \mathbf{w}\} = \operatorname{Span} \mathbf{v}$; for the purpose of spanning a subspace, the vector $\mathbf{w}$ is therefore redundant. To generalize this idea, we essentially have to extend the notion of *parallel.*

Definition 1.22. A finite non-empty subset $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of a vector space is *linearly dependent* if

$$\exists a_i \in \mathbb{F} \text{ not all zero, for which } a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

Such an equation is a *linear dependence.*[a]

An infinite set is linearly dependent if it has as least one non-empty linearly dependent subset.

[a] The *not all zero* condition is crucial! You can always write $\mathbf{0} = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n$ (a *trivial representation* of $\mathbf{0}$), but this tells you nothing about the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. A linear dependence is a *non-trivial* representation of the zero vector!

Examples 1.23. 1. Vectors $\mathbf{v}, \mathbf{w}$ are linearly dependent (i.e. $\{\mathbf{v}, \mathbf{w}\}$ *is* linearly dependent) if and only if they are *parallel.*

2. $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ and $\mathbf{v}_3 = \begin{pmatrix} 7 \\ 5 \\ 6 \end{pmatrix}$ are linearly dependent since

$$2\mathbf{v}_1 + 3\mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}$$

3. The infinite set $S = \{\begin{pmatrix} x \\ y \end{pmatrix} : y > 1\}$ is linearly dependent in $\mathbb{R}^2$. For instance $\{\begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix}\}$ is a finite linearly dependent subset of S, since $3\begin{pmatrix} 0 \\ 2 \end{pmatrix} - 2\begin{pmatrix} 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

We now state the negation of the definition.

Definition (1.22 cont.). A finite subset $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of a vector space is *linearly independent* if

$$\forall a_i \in \mathbb{F}, \quad a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0} \implies a_1 = \cdots = a_n = 0$$

An infinite set is linearly independent if *all* of its finite subsets are linearly independent.

Examples 1.24. 1. The set $S = \{\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -5 \end{pmatrix}\}$ is linearly independent in $\mathbb{R}^2$ since

$$a\begin{pmatrix} 2 \\ 1 \end{pmatrix} + b\begin{pmatrix} 3 \\ -5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies \begin{cases} 2a + 3b = 0 \\ a - 5b = 0 \end{cases} \implies a = b = 0$$

2. The empty set $\emptyset$ is trivially linearly independent since there is no condition to check.

3. Consider the set $S = \{1 - x^2, -x + 2x^2, 1 + 2x - x^2\}$ in $P_2(\mathbb{R})$. Attempting to find a linear dependence is equivalent to finding a non-trivial solution (a, b, c) to a system of linear equations

$$a(1 - x^2) + b(-x + 2x^2) + c(1 + 2x - x^2) = 0 \iff \begin{cases} a + c = 0 \\ -b + 2c = 0 \\ -a + 2b - c = 0 \end{cases}$$

Since the only solution is trivial $(a, b, c) = (0, 0, 0)$, the set S is linearly independent.

4. $S = \{1, x, x^2, x^3, \ldots\}$ is a linearly independent subset of $P(\mathbb{R})$: we leave this as an exercise.

We consider shrinking or enlarging certain sets of vectors. Prove the next result yourself.

Lemma 1.25. *Suppose $S_1 \subseteq S_2$ are subsets of a vector space. If S_1 is linearly dependent,[a] so is S_2.*

[a] Equivalently (the contrapositive): if S_2 is linearly independent, so is S_1.

Now turn the lemma on its head: if S_2 is linearly independent, when it is possible to find a *larger* linearly independent set $S_2 \supseteq S_1$? What follows is one of the most important results in the course.

Theorem 1.26. *Suppose that S is a linearly independent subset of V and that $\mathbf{v} \notin S$ is given. Then*

$$S \cup \{\mathbf{v}\} \text{ is linearly independent} \iff \mathbf{v} \notin \operatorname{Span} S$$

Be careful reading the proof: we use the contrapositive and prove both directions simultaneously!

Proof. By definition, $S \cup \{\mathbf{v}\}$ is linearly *dependent* if and only if there exist *finitely many* vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$ and scalars $a, a_1, \ldots, a_n$ (not all zero), such that

$$a\mathbf{v} + a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0} \qquad (*)$$

Plainly $a \neq 0$, for otherwise S would be linearly dependent. By dividing through all coefficients by $-a$ we therefore see that $(*)$ is equivalent to

$$\mathbf{v} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n \iff \mathbf{v} \in \operatorname{Span} S$$

■

Examples 1.27. 1. Let $S = \{\mathbf{i}\} = \{\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)\}$ and $\mathbf{v} = \left(\begin{smallmatrix}a\\b\end{smallmatrix}\right)$. Then

$$\{\mathbf{i}, \mathbf{v}\} \text{ linearly independent} \iff \mathbf{v} \notin \operatorname{Span}\{\mathbf{i}\} \iff \mathbf{v} \text{ not parallel to } \mathbf{i} \iff b \neq 0$$

2. Plainly $S = \{\mathbf{v}, \mathbf{w}\} = \left\{\left(\begin{smallmatrix}1\\2\\-1\end{smallmatrix}\right), \left(\begin{smallmatrix}-1\\1\\2\end{smallmatrix}\right)\right\}$ is linearly independent (recall Example 1.18.3).

   (a) Let $\mathbf{u} = \left(\begin{smallmatrix}1\\-3\\0\end{smallmatrix}\right)$: we check that $\mathbf{u} \notin \operatorname{Span} S$. If it were, there would exist $a, b \in \mathbb{R}$ such that

$$\mathbf{u} = a\begin{pmatrix}1\\2\\-1\end{pmatrix} + b\begin{pmatrix}-1\\1\\2\end{pmatrix} \implies \begin{pmatrix}1\\-3\\0\end{pmatrix} = \begin{pmatrix}1 & -1\\2 & 1\\-1 & 2\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}$$

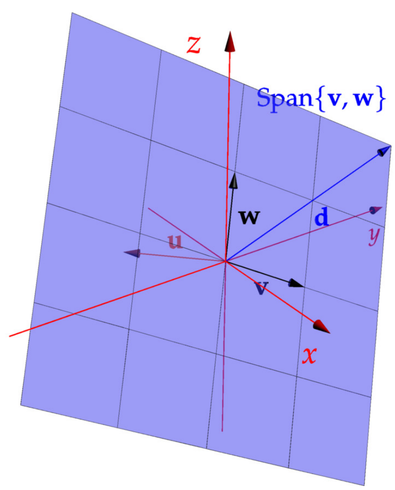
   which has no solutions. It follows that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent. Indeed $\operatorname{Span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\} = \mathbb{R}^3$.

   (b) If we let $\mathbf{d} = \left(\begin{smallmatrix}0\\6\\2\end{smallmatrix}\right)$, then

$$\mathbf{d} = 2\mathbf{v} + 2\mathbf{w}$$

   whence $\mathbf{d} \in \operatorname{Span}\{\mathbf{v}, \mathbf{w}\}$ and so $\{\mathbf{d}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent.

   In the picture, $\mathbf{d}$ lies in the plane spanned by $\mathbf{v}, \mathbf{w}$ while $\mathbf{u}$ does not.



The Theorem should be intuitive in $\mathbb{R}^2$ and $\mathbb{R}^3$ where planes and lines are easy to visualize. Analogues abound elsewhere: indeed the following are reasonable statements according to the RGB (additive) theory of colors,

$$\text{purple} \in \operatorname{Span}\{\text{red}, \text{blue}\}, \qquad \text{brown} \notin \operatorname{Span}\{\text{red}, \text{blue}\}$$

For instance purple, red and blue are not independent colors.

Exercises 1.3

1. In $\mathbb{R}^3$, let $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 1 \\ 7 \\ 5 \end{pmatrix}$. Show that $\mathbf{x} \in \operatorname{Span}\{\mathbf{v}, \mathbf{w}\}$.

2. For the following lists of polynomials in $P_3(\mathbb{R})$, determine whether f can be expressed as a linear combination of g and h.
   (a) $f = 4x^3 + 2x^2 - 6, \quad g = x^3 - 2x^2 + 4x + 1, \quad h = 3x^3 - 6x^2 + x + 4$
   (b) $f = x^3 - 8x^2 + 4x, \quad g = x^3 - 2x^2 + 3x - 1, \quad h = x^3 - 2x + 3$

3. Determine whether the vectors $A, B \in M_2(\mathbb{R})$ lie in the span of S:
$$A = \begin{pmatrix} 1 & 2 \\ -3 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad S = \left\{ \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \right\}$$

4. Determine whether the set $S = \{1 + x + x^2,\ x - x^2,\ 2 + 3x^2\}$ generates $P_2(\mathbb{R})$.

5. Which of the following sets are linearly independent? Prove your assertions.
   (a) $\{2, 3 - x, 1 - 2x^2\}$ in $P_2(\mathbb{R})$
   (b) $\{1, x^2 - x, x^2 + x, x^2\}$ in $P_2(\mathbb{R})$
   (c) $\{\sin(x), \cos(x), \tan(x)\}$ in $C\big((-\frac{\pi}{2}, \frac{\pi}{2}), \mathbb{R}\big)$ (recall Example 1.11.3 for the notation)
   (d) $\{\cos^2(x), \sin^2(x), \cos(2x)\}$ in $C(\mathbb{R}, \mathbb{R})$

6. Suppose that $S = \{\mathbf{v}\}$ is a linearly *dependent* set. What is $\mathbf{v}$?

7. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be linearly independent vectors in V. Prove that $\operatorname{Span}\{\mathbf{v}_2, \ldots, \mathbf{v}_n\} \neq V$.

8. Explicitly verify the claim on page 16 that $\operatorname{Span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\} = \mathbb{R}^3$.

9. Show that the functions f, g defined by $f(x) = 2x$ and $g(x) = |x|$ are linearly independent in the vector space $C([-1, 1], \mathbb{R})$, but linearly dependent in $C([0, 1], \mathbb{R})$.

10. Suppose that c is a constant, and consider the continuous functions $f, g \in C(\mathbb{R}, \mathbb{R})$ defined by
$$f(x) = \cos(x + c), \qquad g(x) = 2 \sin x$$
For what values of c are the functions linearly independent? Draw a picture of what happens.

11. Prove that $\operatorname{Span} S$ is the intersection of all subspaces of V which contain S.

12. Justify Example 1.24.4: the infinite set $\{1, x, x^2, \ldots\}$ is linearly independent.

13. Let X, Y, Z be subsets of a vector space V. Prove that:
   (a) $\operatorname{Span}(X \cup Y) = \operatorname{Span} X + \operatorname{Span} Y$;
   (b) $\operatorname{Span}(X \cup Y) = \operatorname{Span} X \iff Y \subseteq \operatorname{Span} X$
   (c) If $\mathbf{y} \in Z$, then $\operatorname{Span}(Z \setminus \{\mathbf{y}\}) = \operatorname{Span} Z \iff \mathbf{y} \in \operatorname{Span}(Z \setminus \{\mathbf{y}\})$

1.4 Bases and Dimension

We now come, arguably, to the most important definition of the course.

Definition 1.28. A *basis* of a vector space is a linearly independent spanning set.

Our main goals are to see that every vector space has a basis and that all bases of the same space have the same number of elements, what we'll call the dimension.

Standard Bases Many vector spaces have commonly used *standard* bases.

Vector Space V	Standard Basis β
$\mathbb{R}^2$	$\{\mathbf{i}, \mathbf{j}\}$
$\mathbb{R}^3$	$\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$
$\mathbb{F}^n$	$\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ where $\mathbf{e}_i$ has i^{th} entry 1 and the rest 0
$M_{m\times n}(\mathbb{F})$	$\{E^{ij} : 1 \le i \le m, 1 \le j \le n\}$ where E^{ij} has ij^{th} entry 1 and remaining entries 0
$P_n(\mathbb{F})$	$\{1, x, x^2, \ldots, x^n\}$
$P(\mathbb{F})$	$\{1, x, x^2, x^3, \ldots\}$

Examples 1.29. 1. The standard bases of $P_3(\mathbb{R})$ and $M_2(\mathbb{R})$ are, respectively,

$$\{1, x, x^2, x^3\} \qquad \text{and} \qquad \{E^{11}, E^{12}, E^{21}, E^{22}\} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

2. The above spaces have other bases! For example $\beta = \{x, x-1, 1+x^2\}$ is a basis of $P_2(\mathbb{R})$: verification of the following should be straightforward:

   - β is linearly independent: $ax + b(x-1) + c(1+x^2) = 0 \implies a = b = c = 0$
   - β spans $P_2(\mathbb{R})$: $\forall s, t, u,\ \exists a, b, c$ such that $s + tx + ux^2 = ax + b(x-1) + c(1+x^2)$

   What matters is that you are comfortable transforming the definitions into algebra!

3. Following the convention that $\operatorname{Span}\emptyset = \{\mathbf{0}\}$, the empty set is a basis of the trivial space $\{\mathbf{0}\}$.

The Unique Co-ordinate Representation We first discuss one of the primary uses of a basis: the representation of vectors in terms of *co-ordinates*.

Definition 1.30. Let $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V over $\mathbb{F}$ and suppose $\mathbf{v} \in V$ is given. The *co-ordinate representation of* $\mathbf{v}$ *with respect to* β is the column vector

$$[\mathbf{v}]_\beta := \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{F}^n \quad \text{where} \quad \mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$$

To check that this is well-defined, we need to make sure that each vector $\mathbf{v}$ has exactly one co-ordinate representation. We'll deal with this in Theorem 1.32, after seeing an example.

Example 1.31. In $P_2(\mathbb{R})$, consider the bases $\alpha = \{1, x, x^2\}$ and $\beta = \{x, x-1, 1+x^2\}$, and the vector

$$p(x) = 3x - 2(x-1) + 5(1+x^2) = 7 + x + 5x^2$$

The co-ordinate representations with respect to the two bases are then:

$$[p]_\alpha = \begin{pmatrix} 7 \\ 1 \\ 5 \end{pmatrix}, \qquad [p]_\beta = \begin{pmatrix} 3 \\ -2 \\ 5 \end{pmatrix}$$

The advantage of co-ordinates is that we can easily invoke matrix methods. The challenge is to keep in mind the *basis* used in the conversion, so we can properly convert back once we're done!

We now verify the uniqueness of co-ordinate representations. Amazingly, this property essentially characterises the concept of a basis.

Theorem 1.32. *Let $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a non-empty finite subset of a vector space V. Then β is a basis if and only if each $\mathbf{v} \in V$ can be written as a unique linear combination*

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \tag{$*$}$$

Compare this to Theorem 1.16: we are really saying that $V = \operatorname{Span}\mathbf{v}_1 \oplus \cdots \oplus \operatorname{Span}\mathbf{v}_n$.

Proof. ($\Rightarrow$) If β is a basis, then $V = \operatorname{Span}\beta$ and so every vector can be expressed in the form $(*)$. Now suppose $\exists\mathbf{v} \in V$ with at least two *distinct* representations:

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$$

It follows that

$$(a_1 - b_1)\mathbf{v}_1 + \cdots + (a_n - b_n)\mathbf{v}_n = \mathbf{0}$$

is a linear dependence on β. Contradiction.
($\Leftarrow$) Conversely, suppose β is not a basis. There are two possibilities:

(a) β does not generate V. In this case, $\exists\mathbf{v} \notin \operatorname{Span}\beta$ with no representation.

(b) β generates V but is linearly dependent. In this case there exists a linear dependence

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

from which the zero vector[a] has at least two representations!

Either way, there exists some $\mathbf{v} \in V$ without a unique representation. ■

[a] $\mathbf{0} = 0\mathbf{v}_1 + \cdots 0\mathbf{v}_n$. Indeed any $\mathbf{v} \in V$ will have multiple representations in this case.

We don't typically refer to *co-ordinates* with respect to infinite bases, but the Theorem can be rephrased so that the uniqueness of representation holds. We will return to co-ordinate representations and their relationship to linear maps and matrix multiplication in the next chapter.

Existence of Finite Bases

If we attempt to enlarge a basis β of V by adding a new vector $\mathbf{v} \notin \beta$, a quick appeal to Theorem 1.26 (let $S = \beta$) shows that

$$\mathbf{v} \in \operatorname{Span}\beta\,(= V) \implies \beta \cup \{\mathbf{v}\} \text{ is linearly dependent}$$

Otherwise said, a basis is a *maximal linearly independent set*. This suggests a simple looped algorithm:

1. Start with a linearly independent set X (even $X = \emptyset$ will do!).
2. Does there exist a vector $\mathbf{s}$ such that $X \cup \{\mathbf{s}\}$ is linearly independent?

   Yes: Repeat step 2 with X replaced with $X \cup \{\mathbf{s}\}$.

   No: Stop. We have a basis.

The algorithm has two problems: how do we find a suitable $\mathbf{s}$, and how do we know that the algorithm will terminate? Both these problems can be addressed by restricting to vector spaces spanned by a *finite* set.

Definition 1.33. A vector space V is *finite-dimensional* if it has a finite spanning set: if there exists a *finite* subset $S \subseteq V$ such that $\operatorname{Span} S = V$.

Theorem 1.34 (Existence of a Basis). *Every finite-dimensional vector space has a basis.*
More specifically, suppose X and S are subsets of V such that:

- *S is a finite spanning set for V;*
- *X is a linearly independent subset of S.*

Then there exists a basis β of V such that $X \subseteq \beta \subseteq S$: in particular β is a finite *set.*

Proof. Suppose V is non-trivial, for otherwise $\emptyset$ is a basis ($X = \emptyset$ and $S = \emptyset$ or $\{\mathbf{0}\}$).
Let $m = |X|$ and $n = |S|$ be the cardinalities so that $m \leq n$, and label $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$. Then

$$X \subseteq S \implies \operatorname{Span} X \subseteq \operatorname{Span} S = V$$

Loop: If $\operatorname{Span} X = \operatorname{Span} S = V$, we are done: X is a basis.
Otherwise, $\exists \mathbf{s}_{m+1} \in S$ such that $\mathbf{s}_{m+1} \notin \operatorname{Span} X$, whence (Theorem 1.26)

$$X \cup \{\mathbf{s}_{m+1}\} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{s}_{m+1}\} \text{ is linearly independent.}$$

Now repeat the loop with $X \cup \{\mathbf{s}_{m+1}\}$ in place of X (induction).

The process must terminate with a basis in at most $n - m$ steps since S is a *finite* spanning set.
To establish the primary claim, simply choose any $\mathbf{x} \in S$ and let $X = \{\mathbf{x}\}$. ∎

The finite spanning set S was crucial in resolving the problems with our looped algorithm: it provided a finite list of vectors from which to choose, and guaranteed that only finitely many loops were possible thus forcing the algorithm to terminate.[1] The existence of bases for *infinite-dimensional* spaces (no finite spanning sets) is more technical and will be outlined in the next section.

Example 1.35. We follow the algorithm in $\mathbb{R}^3$, omitting explicit calculations for brevity.

$$S = \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4, \mathbf{s}_5, \mathbf{s}_6\} = \left\{ \begin{pmatrix}1\\2\\0\end{pmatrix}, \begin{pmatrix}2\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\end{pmatrix}, \begin{pmatrix}5\\4\\2\end{pmatrix}, \begin{pmatrix}0\\1\\3\end{pmatrix}, \begin{pmatrix}1\\1\\4\end{pmatrix} \right\}$$

1. Let $X = \{\mathbf{s}_1\}$. Since $\mathbf{s}_2 \notin \operatorname{Span} X$ we conclude that $\{\mathbf{s}_1, \mathbf{s}_2\}$ is linearly independent.
2. $\{\mathbf{s}_1, \mathbf{s}_2\}$ does not span $\mathbb{R}^3$ so we need another vector.
   - $\mathbf{s}_3 = \mathbf{s}_2 - \mathbf{s}_1 \in \operatorname{Span}\{\mathbf{s}_1, \mathbf{s}_2\}$ so we reject $\mathbf{s}_3$
   - $\mathbf{s}_4 = \mathbf{s}_1 + 2\mathbf{s}_2 \in \operatorname{Span}\{\mathbf{s}_1, \mathbf{s}_2\}$ so we also reject $\mathbf{s}_4$
   - We accept $\mathbf{s}_5$ since $\mathbf{s}_5 \notin \operatorname{Span}\{\mathbf{s}_1, \mathbf{s}_2\}$.
3. $\beta := \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_5\}$ is linearly independent and spans $\mathbb{R}^3$: it is a suitable basis.

The Exchange Theorem and its Consequences

Our next goal is *comparison* of the cardinalities of spanning sets and linearly independent set (and thus bases), the key to which is the *Exchange* (or *Replacement*) *Theorem.* Take your time, since this is the trickiest result of the course so far.

Theorem 1.36 (Exchange Theorem). *Let V be a finite-dimensional vector space. If S is a finite spanning set and X is a linearly independent subset of V, then $|X| \le |S|$. More specifically,*

$$\exists T \subseteq S \text{ such that } |T| = |X| \text{ and } \operatorname{Span}(X \cup (S \setminus T)) = V$$

A few observations before we see the proof.

- The hypotheses are the same as for the Existence Theorem (1.34), except that X need not be a subset of S. The result therefore allows us to compare *unrelated* subsets.
- The result shows that *every* linearly independent set is no larger than *every* finite spanning set. In particular, we obtain the important fact that *every linearly independent subset and thus basis is finite*!
- The subset T is sometimes called the *exchange,* since the theorem essentially exchanges T with X while preserving the span.
- Since the proof depends on a tricky induction, it is strongly recommended to work through an example (say Example 1.37) while reading. The exchange is often easy to compute when X and S are small. If you really want to understand the proof, make up more examples! Alternatively, simply skip the proof and come back later; while important, it is technical and hard to use directly in examples.

[1] Imagine applying the algorithm $X = \{1\}$ and $S = \{1, x, x^2, \ldots\}$ in the space of polynomials $P(\mathbb{R})$: what happens?

Proof. Denote $n = |S|$ and $m = \min\{n, |X|\}$: our eventual goal is to see that $|X| = m$, but at present we don't know whether X is finite!

Consider a subset $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subseteq X$. We prove the following claim by induction:

$$\forall k \in \{0, 1, \ldots, m\},\ \exists \mathbf{s}_1, \ldots, \mathbf{s}_k \in S \text{ such that } \operatorname{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{s}_{k+1}, \ldots, \mathbf{s}_n\} = V \tag{†}$$

Base case If $k = 0$ then the claim is trivial for S spans V.

Induction step Suppose the claim holds for some $k < m$. Since $\{\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{s}_{k+1}, \ldots, \mathbf{s}_n\}$ spans V, there exist coefficients a_i, b_j such that

$$\mathbf{x}_{k+1} = a_1\mathbf{x}_1 + \cdots + a_k\mathbf{x}_k + b_{k+1}\mathbf{s}_{k+1} + \cdots + b_n\mathbf{s}_n \tag{$*$}$$

Since $(*)$ is a linear dependence, the *independence* of X shows that at least one $b_j \neq 0$: WLOG assume $b_{k+1} \neq 0$. Since

$$\mathbf{s}_{k+1} = -b_{k+1}^{-1}(a_1\mathbf{x}_1 + \cdots + a_k\mathbf{x}_k - \mathbf{x}_{k+1} + b_{k+2}\mathbf{s}_{k+2} + \cdots + b_n\mathbf{s}_n)$$

we may eliminate $\mathbf{s}_{k+1}$ from any linear combination describing an element of V at the cost of including $\mathbf{x}_{k+1}$: we conclude that

$$V = \operatorname{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_{k+1}, \mathbf{s}_{k+2}, \ldots, \mathbf{s}_n\}$$

By induction, the claim is proved. Taking $k = m$ and setting $T = \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ we see that

$$V = \operatorname{Span}\Big(\{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \cup (S \setminus T)\Big)$$

Now suppose, for contradiction, that $|X| > m$. Then $n = m$ and $\exists \mathbf{x}_{m+1} \in X$. Since (†) holds for $k = m = n$, we see that

$$\mathbf{x}_{m+1} \in V = \operatorname{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$$

which contradicts the linear independence of X.

We complete the proof by observing that $|X| = m \leq n = |S|$. ∎

Example 1.37. Let $V = \mathbb{R}^3$, $S = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, $X = \left\{ \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}, \begin{pmatrix} 6 \\ 9 \\ 12 \end{pmatrix} \right\}$ and apply the induction step twice:

1. Since $\mathbf{x}_1 = 2\mathbf{i} + 3\mathbf{j} + 5\mathbf{k}$ and the coefficient of $\mathbf{i}$ is non-zero, we choose $\mathbf{s}_1 = \mathbf{i}$. Observe that $\operatorname{Span}\{\mathbf{x}_1, \mathbf{j}, \mathbf{k}\} = \operatorname{Span} S = \mathbb{R}^3$.

2. Now find the coefficients of $\mathbf{x}_2$ with respect to $\{\mathbf{x}_1, \mathbf{j}, \mathbf{k}\}$: this might need a little augmented matrix work, but we see that

$$\mathbf{x}_2 = \begin{pmatrix} 6 \\ 9 \\ 12 \end{pmatrix} = 3\mathbf{x}_1 + 0\mathbf{j} - 3\mathbf{k}$$

so we choose $\mathbf{s}_2 = \mathbf{k}$. Again we have $\operatorname{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{j}\} = \mathbb{R}^3$. The exchange is therefore $T = \{\mathbf{i}, \mathbf{k}\}$.

Armed with the Exchange Theorem, the key facts come quickly and easily:

Corollary 1.38. *Given a finite-dimensional vector space:*

1. *(Extension Theorem) Any linearly independent subset may be extended to a basis.*
2. *(Well-definition of Dimension) Any two bases have the same cardinality.*

Proof. If the space is trivial then both statements are immediate. Otherwise:

1. Suppose S is a finite spanning set for V and that X a linearly independent subset. The Exchange Theorem says there exists a finite spanning set $X \cup (S \setminus T)$ containing X. By the Existence theorem, there exists a basis β such that $X \subseteq \beta \subseteq X \cup (S \setminus T)$.
2. If β and γ are bases, take $X = \beta$ and $S = \gamma$ in the Exchange Theorem to see that $|\beta| \leq |\gamma|$. Now repeat the argument with the roles reversed.

■

Definition 1.39. The *dimension* $\dim_{\mathbb{F}} V$ of a finite-dimensional vector space V over a field $\mathbb{F}$ is the cardinality of any basis.[a]

[a] We usually write $\dim V$ if the field is understood, but be careful: see Exercise 1.4.6...

Examples 1.40. 1. The dimension is often part of the name of a vector space or is easily read off:

$$\dim \mathbb{R}^5 = 5, \qquad \dim \mathbb{F}^n = n, \qquad \dim M_{m\times n}(\mathbb{F}) = mn$$

2. Beware of *polynomials*! The standard basis of $P_n(\mathbb{R})$ is $\{1, x, \ldots, x^n\}$, whence $\dim P_n(\mathbb{R}) = n + 1$.

Corollary 1.41. *Suppose W is a subspace of a finite-dimensional space V. Then:*

1. $\dim W \leq \dim V$.
2. $\dim W = \dim V \implies W = V$.

Proof. By the Extension Theorem, we may extend any basis α of W to a basis β of V. Since $\alpha \subseteq \beta$, we plainly have $\dim W = |\alpha| \leq |\beta| = \dim V$. Part 2 is an exercise.

■

Summary: bases of finite-dimensional vector spaces We have not quite proved all of the following, but all should now seem at least reasonable.

1. Every such space has a basis and all have the same cardinality (the dimension).
2. We can extend a linearly independent set to a basis: a basis is a *maximal linearly independent set*.
3. Every spanning set contains a basis as a subset: a basis is a *minimal spanning set*.
4. A subset β is a basis of V if it satisfies *any two* of the following (it then satisfies the third):

$$|\beta| = \dim V, \qquad \beta \text{ is linearly independent}, \qquad \operatorname{Span}\beta = V$$

Example 1.42. We verify that $\beta = \left\{ \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ is a basis of $\mathbb{R}^3$.

Since $|\beta| = 3 = \dim \mathbb{R}^3$, we need only check linear independence:

$$a \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} + b \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix} + c \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \mathbf{0} \iff \begin{pmatrix} 1 & 2 & 1 \\ 3 & 1 & 1 \\ 4 & 5 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{0} \implies a = b = c = 0$$

since the matrix is invertible.

It is unnecessary, but the invertibility also shows directly that β is a spanning set: given $\mathbf{x} \in \mathbb{R}^3$,

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := \begin{pmatrix} 1 & 2 & 1 \\ 3 & 1 & 1 \\ 4 & 5 & 1 \end{pmatrix}^{-1} \mathbf{x} \implies \mathbf{x} = a \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} + b \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix} + c \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \in \operatorname{Span} \beta$$

Exercises 1.4 1. Prove carefully that $\beta = \{3\mathbf{i} + 2\mathbf{k}, 2\mathbf{i} + \mathbf{k}, \mathbf{j} + \mathbf{k}\}$ is a basis of $\mathbb{R}^3$.

2. Let $p(x) = 3 - 5x + 7x^2 \in P_2(\mathbb{R})$. With respect to the bases $\beta = \{1 - x, 1 + x^2, x - 2x^2\}$ and $\gamma = \{2 - x, x^2, 1 + x\}$, find the co-ordinate representations $[p]_\beta$ and $[p]_\gamma$.

3. As in Exercise 1.35, find a subset of S which is a basis of the vector space V.

   (a) $V = \mathbb{R}^3$, $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} \right\}$

   (b) $V = P_3(\mathbb{R})$, $S = \{1 + 2x,\ 1 + x + x^2,\ 2 + x - x^2,\ 3 + 2x,\ x - 2x^3\}$

4. Find a basis and thus the dimension of the following subspace of $\mathbb{F}^5$:

$$W = \{a_1\mathbf{e}_1 + \cdots + a_5\mathbf{e}_5 \in \mathbb{F}^5 : a_1 - a_3 - a_4 = 0\}$$

5. Let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be distinct vectors in a vector space V. Prove that if $\beta = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a basis of V, then $\gamma := \{\mathbf{u} + \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}, \mathbf{w}\}$ is also a basis of V.

6. $\mathbb{C}^3$ is a vector space over $\mathbb{C}$ *and* over $\mathbb{R}$. What are the values $\dim_{\mathbb{C}} \mathbb{C}^3$ and $\dim_{\mathbb{R}} \mathbb{C}^3$? State a basis in each case.

7. (a) Define $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. Prove that this is a vector space over $\mathbb{Q}$ and that $\beta = \{1, \sqrt{2}\}$ is a basis.

   (b) More generally, if $d \in \mathbb{Z}$ is not a perfect square, prove that $\dim_{\mathbb{Q}} \mathbb{Q}(\sqrt{d}) = 2$.

8. Explain the observation in the proof of the Existence Theorem 1.34: If $\operatorname{Span} X \neq \operatorname{Span} S$, then $\exists \mathbf{s}_{m+1} \in S$ such that $\mathbf{s}_{m+1} \notin \operatorname{Span} X$.

9. Given subsets X and S of the vector space V, compute the exchange T from the Exchange Theorem by mirroring Example 1.37.

   (a) $X = \left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix} \right\}$, $S = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, $V = \mathbb{R}^3$.

   (b) $X = \{1 - x, 2 + x^2, 1 + x^3\}$, $S = \{1, x, x^2, x^3\}$, $V = P_3(\mathbb{R})$.

   (c) $X = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix} \right\}$, $S = \{E^{11}, E^{12}, E^{21}, E^{22}\}$, $V = M_2(\mathbb{R})$.

10. Let $V = \{(x_n)_{n=1}^{\infty}\}$ be the set of all sequences of real numbers. This is a vector space over $\mathbb{R}$ under elementwise addition and scalar multiplication. For example: if $x_n = \frac{1}{n}$ and $y_n = 1 - \frac{1}{n^2}$, then $(x_n) + (y_n)$ is the sequence (z_n) with nth term

$$z_n = x_n + y_n = \frac{1}{n} + 1 - \frac{1}{n^2}$$

(a) For each $m \in \mathbb{N}$ define the sequence $E^m = (e_n^m)$ where

$$e_n^m = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{otherwise} \end{cases}$$

Thus $E^1 = (1, 0, 0, 0, \ldots)$ and $E^2 = (0, 1, 0, 0, \ldots)$, etc. Show that the set $X = \{E^m : m \in \mathbb{N}\}$ is a linearly independent subset of V.

(b) Is X a basis of V? Why/why not?

11. Let V be a vector space with dimension $n \geq 1$, and let S be a generating set.

(a) Show that S contains a linearly independent subset X.

(b) If X is a linearly independent subset of S, but X is *not* a basis, prove that $\exists \mathbf{s} \in S$ such that $X \cup \{\mathbf{s}\}$ is linearly independent.

(c) Prove that there exists a subset of S which is a basis of V.

(d) Prove that $|S| \geq n$.

(*This is asking you to modify the proof of the Existence Theorem. Note that you cannot assume that S is a finite set!*)

12. (Optional application) In this question we use linear algebra to find a polynomial of minimal degree through a set of points in the plane. Suppose that a_0, a_1 are distinct real numbers. Define the functions

$$f_0(x) = \frac{x - a_1}{a_0 - a_1}, \qquad f_1(x) = \frac{x - a_0}{a_1 - a_0}$$

It follows that

$$f_i(a_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad (*)$$

(a) Prove that f_0 and f_1 are linearly independent.

(b) Suppose that b_0, b_1 are real numbers. Show that the straight line passing through the points (a_0, b_0) and (a_1, b_1) lies in $\text{Span}\{f_0, f_1\}$ and that, consequently, $\{f_0, f_1\}$ forms a basis of the vector space of linear polynomials $P_1(\mathbb{R})$.

(c) Repeat parts (a) and (b) for any set of distinct values $a_0, a_1, \ldots, a_n$ to obtain polynomials $f_0, f_1, \ldots, f_n$ which satisfy $(*)$ and form a basis of $P_n(\mathbb{R})$. Hence or otherwise, prove that there is a unique degree $\leq n + 1$ polynomial passing through any points $(a_0, b_0), \ldots, (a_n, b_n)$ where the a_i are distinct.

(d) Hence or otherwise, find the unique degree 3 polynomial which passes through the points $(0, 1), (1, 4), (3, -1)$ and $(5, 10)$.

1.5 Maximal linearly independent subsets (non-examinable)

In the previous section, we showed that every finite-dimensional vector space has a basis. What about other vector spaces? Does every vector space have a basis?

Examples 1.43. To see the difficulty, consider two related spaces and the set $\beta = \{1, x, x^2, x^3, \cdots\}$.

1. The space of polynomials $P(\mathbb{R})$ has standard basis β (Exercise 1.3.12), and is therefore an infinite-dimensional space with a *countable* basis: it seems reasonable to write $\dim P(\mathbb{R}) = \aleph_0$.

2. The space V of formal *power series* with coefficients in $\mathbb{R}$ contains the vector

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$$

an *infinite* combination of the elements of β. Plainly β is not a basis of V. But does V have a basis and, if so, how can we find one?

There are two standard ways to tackle our problem.

1. Broaden the concept of linear combination/span to allow for *infinite sums.*[2] This introduces a new difficulty, *convergence,* which takes us into the realm of analysis and requires further definitions. If you later study Banach and Hilbert spaces, this is the approach you will follow. Indeed, in the context of power series, β is incredibly useful, *even more so than a basis* would be!

2. Appeal to *Zorn's Lemma*, a technical result equivalent to the (somewhat) controversial *axiom of choice.* This is the approach we'll follow for the remainder of the section.

Definition 1.44. Let $\mathcal{F}$ be a set of sets. A subset $\mathcal{C} \subseteq \mathcal{F}$ is a *chain*[a] in $\mathcal{F}$ if

$$\forall A, B \in \mathcal{C} \text{ either } A \subseteq B \text{ or } B \subseteq A$$

A chain $\mathcal{C}$ has an *upper bound* in $\mathcal{F}$ if there is some set $B \in \mathcal{F}$ such that

$$\forall A \in \mathcal{C} \text{ we have } A \subseteq B$$

A set $\beta \in \mathcal{F}$ is *maximal* it is a subset of no member of $\mathcal{F}$ but itself.

[a] Alternatively $\mathcal{C}$ is a *nest*, a *tower,* or is *totally ordered.*

The idea is to let $\mathcal{F}$ to be the set of all linearly independent subsets of a vector space V. Our goal is then to hunt for a maximal member of $\mathcal{F}$, since a basis β is precisely a *maximal linearly independent set* (see Exercise 1.5.1):

1. β is linearly independent.

2. The only linearly independent subset of V containing β is β itself.

[2] Definition 1.17 only allows us to conclude, by induction, that any *finite sum* of vectors $\sum_{i=1}^{n} \mathbf{v}_i$ is well-defined. In the abstract, i.e. without *limits,* an infinite sum $\sum_{n=1}^{\infty} \mathbf{v}_n$ has no meaning.

Examples 1.45. 1. Consider the standard basis $\beta = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ of $\mathbb{R}^3$. Clearly β is an upper bound for the following chain of linearly independent subsets

$$\mathcal{C} = \Big\{ \{\mathbf{i}\}, \{\mathbf{i}, \mathbf{j}\}, \{\mathbf{i}, \mathbf{j}, \mathbf{k}\} \Big\}$$

2. The basis $\beta = \{1, x, x^2, \ldots\}$ of $P(\mathbb{R})$ is an upper bound for the chain

$$\mathcal{C} = \Big\{ \{1\}, \{1, x\}, \{1, x, x^2\}, \ldots \Big\}$$

Read this example carefully: the ellipsis hides *infinitely many* subsets. In particular the upper bound β does not have to be an element of the chain! It is, however, the *union* $\beta = \bigcup_{U \in \mathcal{C}} U$ of all elements of the chain...

Axiom 1.46 (Zorn's Lemma). Let $\mathcal{F}$ be a non-empty family of sets. If every chain $\mathcal{C} \subseteq \mathcal{F}$ has an upper bound $M_{\mathcal{C}} \in \mathcal{F}$, then $\mathcal{F}$ has a maximal member.

Theorem 1.47. *Every vector space has a basis.*

Proof. If V is non-trivial, let $\mathcal{F} = \{\text{linearly independent subsets of } V\}$. Plainly this is non-empty. Suppose $\mathcal{C} \subseteq \mathcal{F}$ is a chain and define

$$M_{\mathcal{C}} := \bigcup_{U \in \mathcal{C}} U$$

We claim that $M_{\mathcal{C}}$ is an upper bound for $\mathcal{C}$ in $\mathcal{F}$. For this, we need to show two things:

1. $M_{\mathcal{C}} \in \mathcal{F}$: that is, $M_{\mathcal{C}}$ is a linearly independent set.
2. $\forall A \in \mathcal{C}$, we have $A \subseteq M_{\mathcal{C}}$.

The latter is obvious from the definition of union! For the former, suppose that $\mathbf{u}_1, \ldots, \mathbf{u}_n \in M_{\mathcal{C}}$ are distinct vectors such that

$$a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n = \mathbf{0}$$

By the total ordering of $\mathcal{C}$, we see[a] that $\exists U \in \mathcal{C}$ such that $\mathbf{u}_1, \ldots, \mathbf{u}_n \in U$. But each U is linearly independent, whence $a_1 = \cdots = a_n = 0$. It follows that $M_{\mathcal{C}} \in \mathcal{F}$.

Applying Zorn's lemma, we see that $\mathcal{F}$ has a maximal element β, which is necessarily a basis of V. ∎

[a]Since $\mathbf{u}_i \in M_{\mathcal{C}}$, $\exists U_i \in \mathcal{C}$ such that $\mathbf{u}_i \in U_i$. Now let $U = U_1 \cup \cdots \cup U_n$. By total ordering, one of these U_i contains all the others: this is U. Note that this only works because the subscript n is *finite*!

This argument (create an upper bound by taking the union over a chain before invoking Zorn's Lemma) is replicated in other areas of mathematics.[3] The results of the previous section may be generalized to cover infinite-dimensional vector spaces. A couple are outlined in the exercises.

[3]For instance, in abstract algebra to prove the existence of a maximal ideal in a ring.

Exercises 1.5 (Remember these are optional!)

1. As defined above, a set β is a *maximal linearly independent subset* of V if
   - β is linearly independent.
   - The only linearly independent subset of V containing β is β itself.

   The discussion on page 20 shows that every basis is a maximal linearly independent subset. Prove the converse:

   $$\beta \text{ a maximal linearly independent subset} \implies \beta \text{ is a basis}$$

   (*You cannot assume that β is finite: the entire point of this section is that it needn't be!*)

2. Show that categorization 4 on page 23 does not extend to infinite dimensions: specifically, state a linearly independent subset X of a vector space V such that $|X| = \dim V$, but such that X is *not* a basis of V.

3. Prove a more general version of Theorem 1.32: If β is a basis of V, then for all non-zero $\mathbf{v} \in V$ there is a unique *finite* subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq \beta$ and unique non-zero scalars $a_1, \ldots, a_n$ such that

   $$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$$

   Our only freedom is in the order of the vectors $\mathbf{v}_i$.

   (*Hint: obtain a contradiction by supposing $\mathbf{v} \in V$ is a non-zero vector which can be written as a linear combination of elements of β in two different ways*)

4. Prove the infinite-dimensional version of the Extension Theorem: if X is a linearly independent subset of a vector space V, then there exists a basis of V which contains X.

   (*Hint: let $\mathcal{F}$ be the set of all linearly independent subsets of V which contain X, and mimic the proof of Theorem 1.47*)

5. Consider the set $X = \{e^{\lambda x} : \lambda \in \mathbb{R}\}$. Investigate the idea that X is a linearly independent set in the vector space of continuous functions on $\mathbb{R}$, and the relationship of this to the *Vandermonde matrix.* It follows that Span X is a subspace of $C(\mathbb{R})$ with *uncountably infinite dimension.*

2 Linear Transformations and Matrices

A standard approach in algebra is to study collections of sets with a common structure and the maps between them which preserve that structure. In linear algebra this means vector spaces and the maps which behave nicely with respect to their defining structure, namely addition and scalar multiplication. Otherwise said, *linear maps* should preserve *linear combinations.*

2.1 Linear Maps, Compositions and Isomorphisms

Definition 2.1. Let V and W be vector spaces over the same field $\mathbb{F}$. A function $\mathrm{T} : V \to W$ is *linear* if $\forall \mathbf{v}_1, \mathbf{v}_2 \in V,\ \lambda \in \mathbb{F}$

1. $\mathrm{T}(\mathbf{v}_1 + \mathbf{v}_2) = \mathrm{T}(\mathbf{v}_1) + \mathrm{T}(\mathbf{v}_2)$
2. $\mathrm{T}(\lambda\mathbf{v}_1) = \lambda\mathrm{T}(\mathbf{v}_1)$

equivalently $\mathrm{T}(\lambda\mathbf{v}_1 + \mathbf{v}_2) = \lambda\mathrm{T}(\mathbf{v}_1) + \mathrm{T}(\mathbf{v}_2)$

The set of linear maps from V to W is denoted $\mathcal{L}(V, W)$. If $V = W$ we simply write $\mathcal{L}(V)$.

Warning! The definition looks very similar to that of a *subspace*. Make sure you know the difference!

You have already met many examples of linear maps in your mathematical career.

Examples 2.2.
1. For any V, W, the *zero function* $0 \in \mathcal{L}(V, W)$ maps everything to $\mathbf{0}_W \in W$, while the *identity function* $\mathrm{I} \in \mathcal{L}(V)$ leaves everything untouched:
$$\forall \mathbf{v} \in V, \quad 0(\mathbf{v}) = \mathbf{0}_W, \quad \mathrm{I}(\mathbf{v}) = \mathbf{v}$$
2. If $\mathbf{v} \in \mathbb{F}^n$ and $A \in M_{m\times n}(\mathbb{F})$, then *left-multiplication by A* is the linear map
$$\mathrm{L}_A : \mathbb{F}^n \to \mathbb{F}^m : \mathbf{v} \mapsto A\mathbf{v}$$
Verifying that this is linear is tedious: e.g., the i^{th} entry of the vector $A(\mathbf{x} + \mathbf{y})$ is
$$[A(\mathbf{x} + \mathbf{y})]_i = \sum_{j=1}^{n} a_{ij}(x_j + y_j) = \sum_{j=1}^{n} a_{ij}x_j + \sum_{j=1}^{n} a_{ij}y_j = [A\mathbf{x}]_i + [A\mathbf{y}]_i$$
precisely the i^{th} entry of the vector $A\mathbf{x} + A\mathbf{y}$. Scalar multiplication is similar.
3. Differentiation: Since $(\lambda f + g)' = \lambda f' + g'$, the function $\mathrm{T} : f \mapsto \frac{\mathrm{d}f}{\mathrm{d}x}$ is linear map defined on any vector space of differentiable functions.
4. Integration: $\mathrm{T} : C([a, b]) \to \mathbb{R} : f \mapsto \int_a^b f(x)\,\mathrm{d}x$ is linear, where $C([a, b])$ is the set of continuous functions $f : [a, b] \to \mathbb{R}$.

The following are easy to prove straight from Definition 2.1.

Lemma 2.3.
1. *Linear maps preserve zero:* $\mathrm{T}(\mathbf{0}_V) = \mathrm{T}(0\mathbf{v}) = 0\mathrm{T}(\mathbf{v}) = \mathbf{0}_W$.
2. *$\mathcal{L}(V, W)$ is a vector space whose identity is the zero function* $0 \in \mathcal{L}(V, W)$.

More important is the fact that linear maps are, as advertised, precisely those functions which preserve linear combinations.

Lemma 2.4. *A function* $\mathrm{T} : V \to W$ *is linear if and only if*

$$\forall \mathbf{v}_i \in V, a_i \in \mathbb{F}, \quad \mathrm{T}\left(\sum_{i=1}^{n} a_i \mathbf{v}_i\right) = \sum_{i=1}^{n} a_i \mathrm{T}(\mathbf{v}_i)$$

In particular, T *is defined by what it does to a basis: if* β *is a basis and we know all the values* $\mathrm{T}(\mathbf{v}_j)$ *for every* $\mathbf{v}_j \in \beta$*, then we know* T.

The proof is an exercise.

Example 2.5. The standard basis of $\mathbb{R}^3$ is $\beta = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. Suppose $\mathrm{T} \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ is such that

$$\mathrm{T}(\mathbf{i}) = \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \quad \mathrm{T}(\mathbf{j}) = \begin{pmatrix} 1 \\ 9 \end{pmatrix}, \quad \mathrm{T}(\mathbf{k}) = \begin{pmatrix} -4 \\ 7 \end{pmatrix}$$

Since T is defined on a basis, we can easily compute the entire map:

$$\mathrm{T}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\mathrm{T}(\mathbf{i}) + y\mathrm{T}(\mathbf{j}) + z\mathrm{T}(\mathbf{k}) = x\begin{pmatrix} 3 \\ 4 \end{pmatrix} + y\begin{pmatrix} 1 \\ 9 \end{pmatrix} + z\begin{pmatrix} -4 \\ 7 \end{pmatrix} = \begin{pmatrix} 3 & 1 & -4 \\ 4 & 9 & 7 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

It should be no surprise that the linear map T is in fact L_A where $A = \left(\begin{smallmatrix} 3 & 1 & -4 \\ 4 & 9 & 7 \end{smallmatrix}\right)$. We'll explore the relationship between linear maps, bases and matrices more fully in Section 2.3.

Compositions & Inverses

As functions, linear maps may be composed with each other, and might have inverses. Likely the only new information in the definition is notational.

Definition 2.6. 1. If $\mathrm{T} \in \mathcal{L}(V, W)$ and $\mathrm{U} \in \mathcal{L}(W, X)$, then the *composition* of T and U is defined by[a]

$$\mathrm{UT} : V \to X : \mathbf{v} \mapsto \mathrm{U}(\mathrm{T}(\mathbf{v}))$$

In the special case that $\mathrm{U} = \mathrm{T}$ (necessarily $V = W = X$), the composition is written T^2; similarly for higher powers $\mathrm{T}^3, \mathrm{T}^4$, etc.

2. $\mathrm{T} \in \mathcal{L}(V, W)$ is *invertible*, or an *isomorphism*, if has an *inverse*; a *function* $\mathrm{U} : W \to V$ for which

$$\mathrm{TU} = \mathrm{I}_W \quad \text{and} \quad \mathrm{UT} = \mathrm{I}_V$$

where $\mathrm{I}_W, \mathrm{I}_V$ are the identity maps on V, W respectively.

Vector spaces V, W are *isomorphic* if there exists an isomorphism $\mathrm{T} \in \mathcal{L}(V, W)$.

[a] For brevity, we write UT instead of $\mathrm{U} \circ \mathrm{T}$.

Examples 2.7. 1. If $\mathrm{T} \in \mathcal{L}(\mathbb{R}^2, P_2(\mathbb{R}))$ is defined by $\mathrm{T}\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) = a + bx$, then T is an isomorphism, with inverse $\mathrm{U} : a + bx \mapsto \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$.

2. (a) Let $A = \left(\begin{smallmatrix} 1 & 2 \\ -3 & 1 \end{smallmatrix}\right)$ and consider the linear map $\mathrm{L}_A \in \mathcal{L}(\mathbb{R}^2)$. Then

$$\mathrm{L}_A^2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -3 & 1 \end{pmatrix} \left(\begin{pmatrix} 1 & 2 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right) = \begin{pmatrix} 1 & 2 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} x + 2y \\ -3x + y \end{pmatrix} = \begin{pmatrix} -5x + 4y \\ -6x - 5y \end{pmatrix}$$

Unsurprisingly, this is the linear map L_{A^2}: indeed $A^2 = \left(\begin{smallmatrix} -5 & 4 \\ -6 & -5 \end{smallmatrix}\right)$.

(b) Now consider $\mathrm{U} = \mathrm{L}_B$ where $B = \frac{1}{7}\left(\begin{smallmatrix} 1 & -2 \\ 3 & 1 \end{smallmatrix}\right) = A^{-1}$ is the inverse matrix. Plainly

$$\mathrm{TU}\begin{pmatrix} x \\ y \end{pmatrix} = A\left(B\begin{pmatrix} x \\ y \end{pmatrix}\right) = AB\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \qquad \mathrm{UT}\begin{pmatrix} x \\ y \end{pmatrix} = (BA)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

whence U is an inverse of T, which is therefore an isomorphism.

3. Let $V = C^\infty(\mathbb{R})$ be the vector space of infinitely differentiable functions $f : \mathbb{R} \to \mathbb{R}$. Then $\mathrm{T}(f)(x) = f'(x) + f(x)$ and $\mathrm{U}(f)(x) = \int_0^1 f(x)\,\mathrm{d}x$ are linear maps $\mathrm{T}, \mathrm{U} \in \mathcal{L}(V)$. We compute:

$$\mathrm{UT}(f)(x) = \mathrm{U}\left(f'(x) + f(x)\right) = \int_0^1 f'(x) + f(x)\,\mathrm{d}x = f(1) - f(0) + \int_0^1 f(x)\,\mathrm{d}x$$

4. Define $\mathrm{T}, \mathrm{U} \in \mathcal{L}(P_1(\mathbb{R}))$ by $\mathrm{T}(f)(x) = f(x) + 3f'(x)$ and $\mathrm{U}(f)(x) = f(x) - 3f'(x)$. Then,

$$\mathrm{TU}(f)(x) = \mathrm{T}\big(f(x) - 3f'(x)\big) = f(x) - 3f'(x) + 3\big(f'(x) - 3f''(x)\big) = f(x)$$

since $f''(x) = 0$. We can similarly check that $\mathrm{UT} = \mathrm{I}_{P_1(\mathbb{R})}$ so that U is an inverse of T.

The examples suggest some simple results.

Lemma 2.8. *A composition of linear maps is linear.*

Proof. This follows from the linearity of both T and U:

$$\begin{aligned} \mathrm{UT}(\lambda \mathbf{v}_1 + \mathbf{v}_2) &= \mathrm{U}\Big(\mathrm{T}(\lambda \mathbf{v}_1 + \mathbf{v}_2)\Big) = \mathrm{U}\Big(\lambda \mathrm{T}(\mathbf{v}_1) + \mathrm{T}(\mathbf{v}_2)\Big) && \text{(linearity of T)} \\ &= \lambda \mathrm{U}\big(\mathrm{T}(\mathbf{v}_1)\big) + \mathrm{U}\big(\mathrm{T}(\mathbf{v}_2)\big) && \text{(linearity of U)} \\ &= \lambda \mathrm{UT}(\mathbf{v}_1) + \mathrm{UT}(\mathbf{v}_2) \end{aligned}$$

■

Lemma 2.9. *Let* $\mathrm{T} \in \mathcal{L}(V, W)$ *be an isomorphism. Then:*

1. *The inverse is unique: we call this function* T^{-1}.
2. *The inverse is an isomorphism:* $\mathrm{T}^{-1} \in \mathcal{L}(W, V)$ *is linear and invertible with* $(\mathrm{T}^{-1})^{-1} = \mathrm{T}$.
3. *If* $\mathrm{S} \in \mathcal{L}(W, X)$ *is invertible, then* $\mathrm{ST} \in \mathcal{L}(V, X)$ *is invertible with* $(\mathrm{ST})^{-1} = \mathrm{T}^{-1}\mathrm{S}^{-1}$.

We leave the proof as an exercise. We'll return to invertibility later.

Exercises 2.1

1. Show explicitly that the following are linear maps:
   (a) $\mathrm{T} : \mathbb{R}^2 \mapsto \mathbb{R}^3 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \left(\begin{smallmatrix} 3x \\ 2y-x \\ x \end{smallmatrix}\right)$
   (b) $\mathrm{T} : P_3(\mathbb{R}) \to P_3(\mathbb{R})$ defined by $\mathrm{T}(f)(x) = (x-1)f''(x) + 7f(x)$
   (c) $\mathrm{T} : M_n(\mathbb{F}) \to M_n(\mathbb{F}) : A \mapsto 3A - 2A^T$, where A^T is the transpose

2. Give a reason why the function $\mathrm{T} : \mathbb{R}^2 \to \mathbb{R}^2 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} xy \\ 3x - y \end{pmatrix}$ is non-linear.

3. Let $\mathrm{T}(f)(x) = f(x) + xf'(x)$ where $f(x) \in P_2(\mathbb{R})$.
   (a) Show that $\mathrm{T} \in \mathcal{L}(P_2(\mathbb{R}))$.
   (b) Compute the linear map $\mathrm{T}^2 \in \mathcal{L}(P_2(\mathbb{R}))$; that is, express $\mathrm{T}^2(f)(x)$ in terms of f and its derivatives.

4. Let $\mathrm{T}, \mathrm{U} \in \mathcal{L}(P_2(\mathbb{R}))$ be defined by
$$\mathrm{T}(f)(x) = 2f(x) + f''(x), \qquad \mathrm{U}(f)(x) = \frac{1}{4}(2f(x) - f''(x))$$
   Prove that $\mathrm{U} = \mathrm{T}^{-1}$.

5. Prove or disprove: $\mathrm{T} : \mathbb{R}^2 \to P_1(\mathbb{R}) : \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \mapsto ax - 2a + b$ is an isomorphism.
   (*Try to guess an inverse!*)

6. Prove, the 'equivalently' claim in Definition 2.1, that $\mathrm{T} : V \to W$ is linear if and only if
$$\forall \mathbf{v}_1, \mathbf{v}_2 \in V,\ \lambda \in \mathbb{F},\ \mathrm{T}(\lambda \mathbf{v}_1 + \mathbf{v}_2) = \lambda \mathrm{T}(\mathbf{v}_1) + \mathrm{T}(\mathbf{v}_2)$$

7. Prove explicitly that if $\mathrm{T}_1, \mathrm{T}_2 \in \mathcal{L}(V, W)$, then $\mathrm{T}_1 + \mathrm{T}_2$ is also a linear map.

8. Prove Lemma 2.4.

9. Prove all three parts of Lemma 2.9.

10. With reference to Lemma 2.9, explain why 'Isomorphic' is an equivalence relation on any set of vector spaces.

11. Prove that a linear map $\mathrm{T} \in \mathcal{L}(V, W)$ is an isomorphism if and only if it is *bijective*, that is,
    (a) *Injective*: $\forall \mathbf{v}_1, \mathbf{v}_2 \in V, \mathrm{T}(\mathbf{v}_1) = \mathrm{T}(\mathbf{v}_2) \implies \mathbf{v}_1 = \mathbf{v}_2$.
    (b) *Surjective*: $\forall \mathbf{w} \in W, \exists \mathbf{v} \in V$ such that $\mathbf{w} = \mathrm{T}(\mathbf{v})$.

2.2 The Rank–Nullity Theorem

We define two sets which are crucial for understanding linear maps.

Definition 2.10. Suppose $\mathrm{T} \in \mathcal{L}(V, W)$. Its *range/image* and *nullspace/kernel* are the sets

$$\mathcal{R}(\mathrm{T}) := \{\mathrm{T}(\mathbf{v}) \in W : \mathbf{v} \in V\}, \qquad \mathcal{N}(\mathrm{T}) := \{\mathbf{v} \in V : \mathrm{T}(\mathbf{v}) = \mathbf{0}_W\}$$

Example 2.11. Let $\mathrm{T} \in \mathcal{L}(P_3(\mathbb{R}), P_2(\mathbb{R}))$ be 'differentiate:'

$$\mathrm{T}(a + bx + cx^2 + dx^3) = b + 2cx + 3dx^2$$

Plainly $\mathcal{N}(\mathrm{T}) = \{a : a \in \mathbb{R}\} \subseteq P_3(\mathbb{R})$ is the space of constants, and

$$\mathcal{R}(\mathrm{T}) = \operatorname{Span}\{0, 1, 2x, 3x^2\} = \operatorname{Span}\{1, 2x, 3x^2\} = P_2(\mathbb{R})$$

The example immediately suggests that the range and nullspace are not merely sub*sets*...

Lemma 2.12. *The nullspace and range of $\mathrm{T} \in \mathcal{L}(V, W)$ are subspaces of V and W respectively.*

Proof. Everything follows from the formula $\mathrm{T}(\lambda\mathbf{v}_1 + \mathbf{v}_2) = \lambda\mathrm{T}(\mathbf{v}_1) + \mathrm{T}(\mathbf{v}_2)$.
$\mathcal{N}(\mathrm{T})$ is non-empty since $\mathrm{T}(\mathbf{0}) = \mathbf{0}$. Moreover, if $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{N}(\mathrm{T})$, so is $\lambda\mathbf{v}_1 + \mathbf{v}_2$, whence $\mathcal{N}(\mathrm{T})$ is a subspace of V. The range is similar. ∎

Definition 2.13. The *rank* and *nullity* of a linear map $\mathrm{T} \in \mathcal{L}(V, W)$ are

$$\operatorname{rank}\mathrm{T} := \dim\mathcal{R}(\mathrm{T}) \qquad \operatorname{null}\mathrm{T} := \dim\mathcal{N}(\mathrm{T})$$

Examples 2.14. 1. Revisiting Example 2.11, we see that $\operatorname{rank}\mathrm{T} = 3$ and $\operatorname{null}\mathrm{T} = 1$.

2. Let $\mathrm{L}_A(\mathbf{v}) = A\mathbf{v}$ where $A \in M_{m\times n}(\mathbb{F})$. Applied to the standard basis β of $\mathbb{F}^n$, we see that

$$\mathcal{R}(\mathrm{L}_A) = \operatorname{Span}\mathrm{L}_A(\beta) = \operatorname{Span}\{A\mathbf{e}_1, \ldots, A\mathbf{e}_n\}$$

Since $A\mathbf{e}_j$ is the j^{th} column of A, we see that $\mathcal{R}(\mathrm{L}_A)$ *is the column space of A.*
For example, if $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 3 & 3 \end{pmatrix}$, then

$$\mathcal{R}(L_A) = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 3 \\ 3 \end{pmatrix}\right\} = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \\ 3 \end{pmatrix}\right\}$$

since the third column is a linear combination of the others. We conclude that $\operatorname{rank}\mathrm{L}_A = 2$. To find the nullspace requires solving the system $A\mathbf{x} = \mathbf{0}$. After a few row operations,

$$\mathcal{N}(\mathrm{L}_A) = \left\{\begin{pmatrix} x \\ y \\ z \end{pmatrix} : x + z = 0 = y + z\right\} = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}\right\} \implies \operatorname{null}\mathrm{L}_A = 1$$

These values in fact satisfy one of the most crucial relationships in linear algebra.

Theorem 2.15 (Rank–Nullity). *If* $\mathrm{T} \in \mathcal{L}(V, W)$*, then* $\operatorname{rank}\mathrm{T} + \operatorname{null}\mathrm{T} = \dim V$.

Examples 2.16. First revisit Examples 2.14.

1. If $\mathrm{T} = \frac{\mathrm{d}}{\mathrm{d}x} \in \mathcal{L}(P_3(\mathbb{R}), P_2(\mathbb{R}))$, then, in accordance with the rank–nullity theorem,
$$\operatorname{rank}\mathrm{T} + \operatorname{null}\mathrm{T} = 3 + 1 = 4 = \dim P_3(\mathbb{R})$$

2. If $A = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 3 & 3 \end{smallmatrix}\right)$ so that $\mathrm{L}_A \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^4)$, then
$$\operatorname{rank}\mathrm{L}_A + \operatorname{null}\mathrm{L}_A = 2 + 1 = 3 = \dim \mathbb{R}^3$$

3. Let $\mathrm{T} \in \mathcal{L}(M_2(\mathbb{R}))$ be defined by $\mathrm{T}(A) = A + A^T$ where A^T is the transpose, that is,
$$\mathrm{T}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2a & b+c \\ b+c & 2d \end{pmatrix}$$
It should be clear that
$$\mathcal{R}(\mathrm{T}) = \left\{ \left(\begin{smallmatrix} p & q \\ q & r \end{smallmatrix}\right) : p, q, r \in \mathbb{R} \right\} \quad \mathcal{N}(\mathrm{T}) = \left\{ \left(\begin{smallmatrix} 0 & s \\ -s & 0 \end{smallmatrix}\right) : s \in \mathbb{R} \right\}$$
are, respectively, the subspaces of symmetric and skew-symmetric matrices. Plainly $\operatorname{rank}\mathrm{T} + \operatorname{null}\mathrm{T} = 3 + 1 = 4 = \dim M_2(\mathbb{R})$.

4. Let $\mathrm{T} \in \mathcal{L}(P(\mathbb{R}))$ be defined by $\mathrm{T}(f)(x) = f(x) + f(-x)$. It is easy to check that this is linear. Moreover,
$$f \in \mathcal{N}(\mathrm{T}) \iff f(-x) = -f(x) \iff f \text{ is odd} \iff f \in \operatorname{Span}\{x, x^3, x^5, x^7, \cdots\}$$
We may also check that the range consists of all even polynomials $\mathcal{R}(\mathrm{T}) = \operatorname{Span}\{1, x^2, x^4, \ldots\}$. The rank–nullity theorem holds even for this infinite-dimensional example, though it isn't very instructive.[4]

Before proving the rank–nullity theorem, we consider what a linear map does to a basis.

Lemma 2.17. *If* β *is a basis of* V*, then* $\mathrm{T}(\beta) := \{\mathrm{T}(\mathbf{v}) : \mathbf{v} \in \beta\}$ *is a spanning set for* $\mathcal{R}(\mathrm{T})$.

Proof. Let $\mathrm{T}(\mathbf{v}) \in \mathcal{R}(\mathrm{T})$. Then $\exists \mathbf{v}_i \in \beta$ such that $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$. By Lemma 2.4,
$$\mathrm{T}(\mathbf{v}) = \mathrm{T}\left(\sum_{i=1}^{n} a_i \mathbf{v}_i\right) = \sum_{i=1}^{n} a_i \mathrm{T}(\mathbf{v}_i) \in \operatorname{Span}(\mathrm{T}(\beta))$$
Thus $\mathcal{R}(\mathrm{T}) \leq \operatorname{Span}\mathrm{T}(\beta)$.
Conversely, the right hand side of Lemma 2.4 is a general element of $\operatorname{Span}\mathrm{T}(\beta)$ which is certainly in the range of T. Thus $\operatorname{Span}\mathrm{T}(\beta) \leq \mathcal{R}(\mathrm{T})$, and the subspaces are equal. ■

[4] If you're happy with addition of infinite cardinals, the rank–nullity theorem reads $\aleph_0 + \aleph_0 = \aleph_0$.

Proof of Rank–Nullity Theorem. Suppose ν is a basis of $\mathcal{N}(\mathrm{T})$. By the Extension Theorem, we may extend to a basis $\beta = \nu \cup \rho$ of V, where $\nu \cap \rho = \emptyset$. The proof now rests on two claims:

1. $\mathrm{T}(\rho)$ is a basis of $\mathcal{R}(\mathrm{T})$. First we verify the spanning property: by Lemma 2.17,

$$\mathcal{R}(\mathrm{T}) = \operatorname{Span} \mathrm{T}(\beta) = \operatorname{Span}\{\mathrm{T}(\mathbf{n}), \mathrm{T}(\mathbf{r}) : \mathbf{n} \in \nu, \mathbf{r} \in \rho\} = \operatorname{Span}\{\mathrm{T}(\mathbf{r}) : \mathbf{r} \in \rho\} = \operatorname{Span} \mathrm{T}(\rho)$$

For linear independence, suppose $\mathbf{r}_1, \ldots, \mathbf{r}_r \in \rho$ and compute:

$$\mathbf{0}_W = \sum_{i=1}^{r} a_i \mathrm{T}(\mathbf{r}_i) = \mathrm{T}\left(\sum_{i=1}^{r} a_i \mathbf{r}_i\right) \implies \sum_{i=1}^{r} a_i \mathbf{r}_i \in \mathcal{N}(\mathrm{T}) \cap \operatorname{Span} \rho = \{\mathbf{0}_V\} \qquad (*)$$

$$\implies a_1 = \cdots = a_n = 0 \qquad (\dagger)$$

We conclude that $\mathrm{T}(\rho)$ is a linearly independent spanning set of $\mathcal{R}(\mathrm{T})$.

2. $|\rho| = |\mathrm{T}(\rho)|$. We leave this to Exercise 2.1.3.

In conclusion:

$$\dim V = |\beta| = |\nu| + |\rho| \overset{(2)}{=} \operatorname{null} \mathrm{T} + |\mathrm{T}(\rho)| \overset{(1)}{=} \operatorname{null} \mathrm{T} + \operatorname{rank} \mathrm{T}$$

■

Note that the proof works when V is infinite-dimensional though the result is not so useful.

Injective & Surjective Linear Maps: Isomorphisms Revisited

It turns out that injectivity and surjectivity may be checked by considering the rank and nullity.

Theorem 2.18. *Suppose* $\mathrm{T} \in \mathcal{L}(V, W)$.

1. T *is injective* $\iff \mathcal{N}(\mathrm{T}) = \{\mathbf{0}\} \iff \operatorname{null} \mathrm{T} = 0$.
2. T *surjective* $\iff \mathcal{R}(\mathrm{T}) = W \implies \operatorname{rank} \mathrm{T} = \dim W$.
   Additionally, if W is finite-dimensional, then $\operatorname{rank} \mathrm{T} = \dim W \implies \mathrm{T}$ *surjective.*
3. *If* $\dim V = \dim W$ *is finite, then*

$$\mathrm{T} \text{ is injective} \iff \operatorname{null} \mathrm{T} = 0 \iff \operatorname{rank} \mathrm{T} = \dim V \iff \mathrm{T} \text{ is surjective}$$

In view of Exercise 2.1.11, if any one of these conditions holds then T *is an isomorphism*

Proof.

1. Injectivity means $\mathrm{T}(\mathbf{v}_1) = \mathrm{T}(\mathbf{v}_2) \implies \mathbf{v}_1 = \mathbf{v}_2$. The result follows follows quickly from,

$$\mathrm{T}(\mathbf{v}_1) = \mathrm{T}(\mathbf{v}_2) \overset{\text{linearity}}{\iff} \mathrm{T}(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0} \iff \mathbf{v}_1 - \mathbf{v}_2 \in \mathcal{N}(\mathrm{T})$$

2. $\mathcal{R}(\mathrm{T}) = W$ is the definition of surjectivity: certainly this implies $\operatorname{rank} \mathrm{T} = \dim W$. In finite dimensions we have the converse:

$$\mathcal{R}(\mathrm{T}) \leq W \text{ and } \dim \mathcal{R}(\mathrm{T}) = \operatorname{rank} \mathrm{T} = \dim W \implies \mathcal{R}(\mathrm{T}) = W$$

3. This follows immediately from the rank–nullity theorem and parts 1 & 2.

■

Examples 2.19. 1. Continuing a previous example, $\mathrm{T} = \frac{\mathrm{d}}{\mathrm{d}x} \in \mathcal{L}(P_3(\mathbb{R}), P_2(\mathbb{R}))$ is surjective ($\operatorname{rank} \mathrm{T} = 3 = \dim P_2(\mathbb{R})$) but not injective ($\operatorname{null} \mathrm{T} = 1 \neq 0$). The non-injectivity of T corresponds to the famous '$+C$' from calculus:

$$\frac{\mathrm{d}}{\mathrm{d}x}(p(x) + C) = \frac{\mathrm{d}}{\mathrm{d}x}p(x)$$

2. Consider $\mathrm{L}_A \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ where $A = \begin{pmatrix} 1 & 3 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ and compute the range and nullspace:

$\mathcal{R}(\mathrm{L}_A) = \operatorname{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} \right\} \implies \operatorname{rank} \mathrm{L}_A = 2 \neq \dim \mathbb{R}^3 \implies \mathrm{L}_A$ is not surjective.

$\mathcal{N}(\mathrm{L}_A) = \{\mathbf{0}\} \implies \mathrm{L}_A$ is injective.

3. Let $\mathrm{T} \in \mathcal{L}(P_2(\mathbb{R}))$ be defined by $\mathrm{T}(p)(x) = p'(x) + (x^2 - 1)\int_0^1 p(t)\,\mathrm{d}t$. Observe that

$$\begin{aligned}
\mathrm{T}(a + bx + cx^2) &= b + 2cx + (x^2 - 1)\left(a + \frac{1}{2}b + \frac{1}{3}c\right) \\
&= \frac{1}{2}b - a - \frac{1}{3}c + 2cx + \left(a + \frac{1}{2}b + \frac{1}{3}c\right)x^2 \\
&= 0 \iff c = 0 = \frac{1}{2}b - a = a + \frac{1}{2}b \iff a = b = c = 0
\end{aligned}$$

Since $\mathcal{N}(\mathrm{T}) = \{0\}$, we see that T is bijective and thus an isomorphism.

Corollary 2.20. *Suppose that V, W are vector spaces over the same field.*

1. *If $\mathrm{T} \in \mathcal{L}(V, W)$ is an isomorphism and β is a basis of V, then $\mathrm{T}(\beta)$ is a basis of W.*
2. *V and W are isomorphic if and only if $\dim V = \dim W$.*

The proof is little mostly a special case of the rank–nullity theorem: indeed you should re-read the proof to convince yourself that the restriction $\mathrm{T}_{\operatorname{Span}\rho} : \operatorname{Span}\rho \to \mathcal{R}(\mathrm{T})$ is indeed an isomorphism!

Proof. 1. This is a special case of part 1 of the proof of the rank–nullity theorem: we have $\mathcal{N}(\mathrm{T}) = \{\mathbf{0}\}$, whence $\nu = \emptyset$ and $\rho = \beta$.

2. ($\Rightarrow$) This is part 2 of the same proof: $\dim V = |\beta| = |\mathrm{T}(\beta)| = \dim W$.

($\Leftarrow$) Let $\dim V = \dim W$ and choose any bases β, γ of V, W. These have the same cardinality, whence $\exists f : \beta \to \gamma$ a bijection. By Lemma 2.4, f defines a unique linear map $\mathrm{T} \in \mathcal{L}(V, W)$: if $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \beta$, define

$$\mathrm{T}\left(\sum_{i=1}^{n} a_i \mathbf{v}_i\right) := \sum_{i=1}^{n} a_i f(\mathbf{v}_i)$$

It is straightforward, if tedious, to check that T is an isomorphism. ■

Examples 2.21. 1. $\mathbb{R}^6, P_5(\mathbb{R}), M_{2\times3}(\mathbb{R}), M_{3\times2}(\mathbb{R})$ are isomorphic since they all have dimension 6 over the same field $\mathbb{R}$. Explicit isomorphisms can be found by lining up the standard bases: e.g.

$$\mathrm{T}: M_{2\times3}(\mathbb{R}) \to P_5(\mathbb{R}) \text{ by } \mathrm{T}\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = a + bx + cx^2 + dx^3 + ex^4 + fx^5$$

2. Be careful with base fields!

   (a) $P_5(\mathbb{R})$ is not isomorphic to $\mathbb{C}^6$ since the (implied) base fields are different $\mathbb{R} \neq \mathbb{C}$.

   (b) Viewing $\mathbb{C}^6$ as a vector space over $\mathbb{R}$, we still fail to have isomorphicity since

   $$\dim_{\mathbb{R}} P_5(\mathbb{R}) = 6 \neq 12 = \dim_{\mathbb{R}} \mathbb{C}^6$$

   (c) As vector spaces over $\mathbb{R}$ the spaces $P_5(\mathbb{R})$ and $\mathbb{C}^3$ are isomorphic since both have *real* dimension 6. A suitable isomorphism is

   $$\mathrm{T}(a + bx + cx^2 + dx^3 + ex^5 + fx^5) = \begin{pmatrix} a+ib \\ c+id \\ e+if \end{pmatrix}$$

Exercises 2.2 1. For each function $\mathrm{T}: V \to W$: prove that T is linear, compute $\mathcal{N}(\mathrm{T})$ and $\mathcal{R}(\mathrm{T})$, and the rank and nullity, verify the Rank–Nullity theorem, and determine whether the function is injective or surjective.

   (a) $\mathrm{T}: \mathbb{R}^2 \to \mathbb{R}^3 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ 0 \\ 2x-y \end{pmatrix}$

   (b) $\mathrm{T}: M_n(\mathbb{F}) \to \mathbb{F} : A \mapsto \operatorname{tr} A$, where $\operatorname{tr} A = \sum_{i=1}^{n} A_{ii}$ is the *trace* of A.

   (c) $\mathrm{T}: P_4(\mathbb{R}) \to P_4(\mathbb{R}) : f \mapsto g$ where $g(x) = f(x) - (x^2+1)f''(x)$.

2. For each linear map, find the range and nullspace and compute the rank and nullity.

   (a) $\mathrm{T} = \mathrm{L}_A \in \mathcal{L}(\mathbb{C}^4)$ where $A = \begin{pmatrix} 1 & i & 1 & -i \\ 0 & 1 & 0 & -1 \\ i & 0 & i & 0 \\ i & -1 & i & 1 \end{pmatrix}$

   (b) $\mathrm{T} \in \mathcal{L}(V)$ where $\mathrm{T}(f)(x) = f''(x) - 4f(x)$ and $V = \operatorname{Span}\{e^{2x}, e^{-2x}, xe^{2x}, xe^{-2x}\}$.

3. Verify the following claims made during the proof of the rank–nullity theorem.

   (a) $(*)$ $\mathcal{N}(\mathrm{T}) \cap \operatorname{Span} \rho = \{\mathbf{0}_V\}$.

   (b) $(\dagger)$ $\sum_{i=1}^{r} a_i \mathbf{r}_i = \mathbf{0}_V \implies a_1 = \cdots = a_n = 0$.

   (c) $|\rho| = |\mathrm{T}(\rho)|$: do this by proving that $\mathrm{T}_\rho : \rho \to \mathrm{T}(\rho)$ is a bijection.

4. Suppose that $\dim V > \dim W$. Prove that there are no injective functions $\mathrm{T} \in \mathcal{L}(V, W)$.

5. Let $\mathrm{T} = \frac{\mathrm{d}}{\mathrm{d}x} \in \mathcal{L}(P(\mathbb{R}))$. Is T injective? Surjective? Why is this not a problem for Theorem 2.18?

6. Which of the following pairs are isomorphic? If yes, state an explicit isomorphism.

   (a) $\mathbb{F}^3$ and $P_3(\mathbb{F})$ (b) $\mathbb{F}^4$ and $P_3(\mathbb{F})$ (c) $M_2(\mathbb{R})$ and $P_3(\mathbb{R})$

   (d) $V = \{A \in M_2(\mathbb{R}) : \operatorname{tr} A = 0\}$ and $\mathbb{R}^4$ (e) $V = \{A \in M_2(\mathbb{R}) : \operatorname{tr} A = 0\}$ and $\mathbb{R}^3$

7. Let $\mathrm{T} \in \mathcal{L}(V, W)$ be an isomorphism and U a subspace of V. Prove that $\mathrm{T}(U) := \{\mathrm{T}(\mathbf{u}) : \mathbf{u} \in U\}$ is a subspace of W and that $\dim \mathrm{T}(U) = \dim U$.

8. (Hard) We prove the *first isomorphism theorem* for vector spaces and see its relation to the rank–nullity theorem. This should be familiar if you've studied group theory: you will need to recall the exercise on cosets from the first chapter. Throughout U is a subspace of V.

   (a) Let ${}^{V}\!/_{U} = \{\mathbf{v} + U : \mathbf{v} \in V\}$ be the set of cosets of U in V. Prove that the *canonical map*

   $$\gamma : V \to {}^{V}\!/_{U} : \mathbf{v} \mapsto \mathbf{v} + U$$

   is linear and has nullspace U.
   (*Thus every subspace of V is the nullspace of some linear map γ with* $\operatorname{dom} \gamma = V$)

   (b) Let $\mathrm{T} \in \mathcal{L}(V, W)$. Prove that

   $$\mathbf{v}_1 + \mathcal{N}(\mathrm{T}) = \mathbf{v}_2 + \mathcal{N}(\mathrm{T}) \iff \mathrm{T}(\mathbf{v}_1) = \mathrm{T}(\mathbf{v}_2)$$

   (c) Prove that the following is a well-defined isomorphism of vector spaces:

   $$\mu : {}^{V}\!/_{\mathcal{N}(\mathrm{T})} \to \mathcal{R}(\mathrm{T}) : \mathbf{v} + \mathcal{N}(\mathrm{T}) \mapsto \mathrm{T}(\mathbf{v})$$

   (d) By extending a basis of U to V, show that for any subspace $U \leq V$ we have

   $$\dim {}^{V}\!/_{U} + \dim U = \dim V$$

   Hence conclude the rank–nullity theorem.

2.3 The Matrix Representation of a Linear Map

Recall that if β is a basis of a n-dimensional vector space V over $\mathbb{F}$, then any vector $\mathbf{v} \in V$ has a unique co-ordinate representation $[\mathbf{v}]_\beta \in \mathbb{F}^n$. The same thing can be done for a linear map, resulting in a tight relationship between linear maps, bases and matrices.

Example 2.22. The linear map on $\mathbb{R}^2$ defined by $\mathrm{T}\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = \left(\begin{smallmatrix} y \\ x \end{smallmatrix}\right)$ is plainly left-multiplication by the matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Otherwise said, $\mathrm{T} = \mathrm{L}_A$. Observe that the *columns* of A are the result of applying T to the standard basis $\epsilon = \{\mathbf{i}, \mathbf{j}\}$:

$$\mathrm{T}(\mathbf{i}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \mathrm{T}(\mathbf{j}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

As the next result shows, given any linear map between finite dimensional spaces, choosing bases yields a representation of the map in terms of matrix multiplication.

Theorem 2.23 (Matrix representations). *Suppose that $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\gamma = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ are bases of V and W respectively.*

1. *If $\mathrm{T} \in \mathcal{L}(V, W)$ then the matrix*

$$A = \Big([\mathrm{T}(\mathbf{v}_1)]_\gamma \ \cdots \ [\mathrm{T}(\mathbf{v}_n)]_\gamma \Big) \in M_{m \times n}(\mathbb{F}) \tag{†}$$

   with j^{th} column $[\mathrm{T}(\mathbf{v}_j)]_\gamma$ is the unique matrix satisfying

$$\forall \mathbf{v} \in V, \quad [\mathrm{T}(\mathbf{v})]_\gamma = A[\mathbf{v}]_\beta \tag{$*$}$$

2. *Given $A \in M_{m \times n}(\mathbb{F})$, there is a unique linear map T satisfying $(*)$.*

Proof. 1. Suppose A is defined by (†) with ij^{th} entry $a_{ij} = \Big[[\mathrm{T}(\mathbf{v}_j)]_\gamma \Big]_i$, let $\mathbf{v} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n \in V$ be given, and compute: the column vector $A[\mathbf{v}]_\beta \in \mathbb{F}^m$ has i^{th} row

$$\begin{aligned} \Big(A[\mathbf{v}]_\beta \Big)_i &= \sum_{j=1}^n a_{ij} b_j = \sum_{j=1}^n \Big([\mathrm{T}(\mathbf{v}_j)]_\gamma \Big)_i b_j = \sum_{j=1}^n \Big(b_j [\mathrm{T}(\mathbf{v}_j)]_\gamma \Big)_i = \left(\sum_{j=1}^n [b_j \mathrm{T}(\mathbf{v}_j)]_\gamma \right)_i \\ &= \left(\left[\mathrm{T}\left(\sum_{j=1}^n b_j \mathbf{v}_j \right) \right]_\gamma \right)_i = \Big([\mathrm{T}(\mathbf{v})]_\gamma \Big)_i \qquad \text{(linearity/Lemma 2.4)} \end{aligned}$$

   A therefore satisfies $(*)$. Conversely, if $\epsilon = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis of $\mathbb{F}^n$ and A satisfies $(*)$, then

$$[\mathrm{T}(\mathbf{v}_j)]_\gamma = A[\mathbf{v}_j]_\beta = A\mathbf{e}_j$$

   is the j^{th} column of A: this proves uniqueness.

2. The co-ordinate representation $[\mathrm{T}(\mathbf{v})]_\gamma$ is unique and so $(*)$ uniquely defines T. ■

Definition 2.24. The matrix defined in (†) is the *matrix representation of* T *with respect to* β *and* γ: we write $A = [\mathrm{T}]^{\gamma}_{\beta}$. In the simplest case when $V = W$ and $\beta = \gamma$, we write $[\mathrm{T}]_{\beta}$.

The Theorem can be summarized by commutative diagrams: both options for travelling from V to $\mathbb{F}^m$ produce the same result.

$$\begin{array}{ccc} V & \xrightarrow{\ \mathrm{T}\ } & W \\ {\scriptstyle [\ \ ]_\beta}\downarrow & & \downarrow{\scriptstyle [\ \ ]_\gamma} \\ \mathbb{F}^n & \xrightarrow{A=[\mathrm{T}]^{\gamma}_{\beta}} & \mathbb{F}^m \end{array} \qquad\qquad \begin{array}{ccc} \mathbf{v} & \longmapsto & \mathrm{T}(\mathbf{v}) \\ \downarrow & & \downarrow \\ [\mathbf{v}]_\beta & \longmapsto & [\mathrm{T}(\mathbf{v})]_\gamma = [\mathrm{T}]^{\gamma}_{\beta}[\mathbf{v}]_\beta \end{array}$$

The big take-away is this:

$$\text{Linear Map} \underset{\text{Bases}}{\overset{\text{Choose}}{\leftrightsquigarrow}} \text{Matrix Multiplication}$$

More precisely, given bases of finite dimensional vector spaces, any linear map between them is equivalent to multiplication by a unique matrix.

Examples 2.25. 1. Recall the course introduction and the linear map T defined by 'rotate clockwise by 30° around the origin in $\mathbb{R}^2$.' With respect to the standard basis $\epsilon = \{\mathbf{i}, \mathbf{j}\}$, the matrix of T is

$$[\mathrm{T}]_\epsilon = \big([\mathrm{T}(\mathbf{i})]_\epsilon \quad \mathrm{T}[(\mathbf{j})]_\epsilon\big) = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$$

To rotate, say $\mathbf{v} = 2\mathbf{i} - 4\mathbf{j}$, we would compute

$$[\mathrm{T}(\mathbf{v})]_\epsilon = [\mathrm{T}]_\epsilon [\mathbf{v}]_\epsilon = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 2 \\ -4 \end{pmatrix} = \begin{pmatrix} \sqrt{3} - 2 \\ -1 - 2\sqrt{3} \end{pmatrix}$$

whence $\mathrm{T}(\mathbf{v}) = (\sqrt{3} - 2)\mathbf{i} + (-1 - 2\sqrt{3})\mathbf{j}$.

2. Recall Example 2.14.1. Let the standard bases of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$ be $\beta = \{1, x, x^2, x^3\}$ and $\gamma = \{1, x, x^2\}$. The matrix of $\mathrm{T} = \frac{\mathrm{d}}{\mathrm{d}x} \in \mathcal{L}(P_3(\mathbb{R}), P_2(\mathbb{R}))$ is then

$$[\mathrm{T}]^{\gamma}_{\beta} = \Big([\mathrm{T}(1)]_\gamma\ [\mathrm{T}(x)]_\gamma\ [\mathrm{T}(x^2)]_\gamma\ [\mathrm{T}(x^3)]_\gamma\Big) = \Big([0]_\gamma\ [1]_\gamma\ [2x]_\gamma\ [3x^2]_\gamma\Big) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

For instance, compare the calculations

$$[\mathrm{T}]^{\beta}_{\gamma} \begin{pmatrix} 2 \\ 5 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \\ 3 \end{pmatrix} \leftrightsquigarrow \frac{\mathrm{d}}{\mathrm{d}x}(2 + 5x + 3x^2 + x^3) = 5 + 6x + 3x^2$$

3. Linear maps often look nice with respect to a sensible basis. For example, let $T \in \mathcal{L}(\mathbb{R}^2)$ is defined by 'reflect in the line $y = 2x$.'

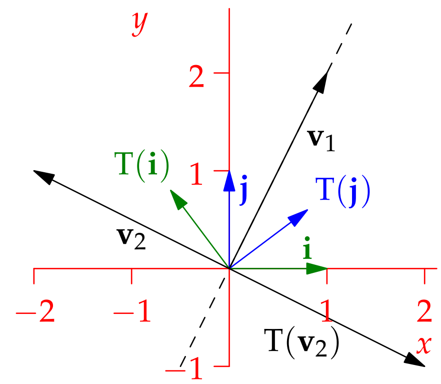
   The vectors $\mathbf{v}_1 = \binom{1}{2}$ and $\mathbf{v}_2 = \binom{-2}{1}$ satisfy

   $$T(\mathbf{v}_1) = \mathbf{v}_1 \qquad T(\mathbf{v}_2) = -\mathbf{v}_2$$

   Clearly $\beta = \{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis of $\mathbb{R}^2$, with respect to which

   $$[T]_\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

   This is much easier than finding the matrix with respect to the standard basis (exercise)



   $$[T]_\epsilon = \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix}$$

   We will revisit this idea later: the vectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ are *eigenvectors* for T; the matrix of a linear map with respect to an eigenbasis is always *diagonal*, with the *eigenvalues* down the diagonal.

4. Here is another example that simplifies nicely in terms of eigenvectors. Consider the linear map $T = L_A \in \mathcal{L}(\mathbb{R}^3)$ where

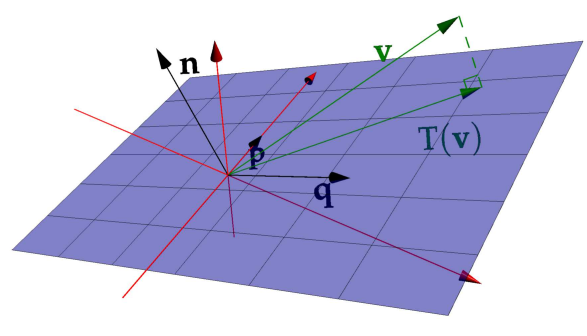
   $$A = \frac{1}{5}\begin{pmatrix} 4 & 0 & 2 \\ 0 & 5 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

   Plainly $A = [T]_\epsilon$ is the matrix of T with respect to the standard basis $\epsilon = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. Now consider

   $$\beta = \{\mathbf{n}, \mathbf{p}, \mathbf{q}\} = \left\{ \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \right\}$$

   It is easy to verify that β is linearly independent and thus a basis of $\mathbb{R}^3$. Moreover,

   $$T(\mathbf{n}) = \mathbf{0}, \qquad T(\mathbf{p}) = \mathbf{p}, \qquad T(\mathbf{q}) = \mathbf{q}$$



   from which the matrix is very simple

   $$[T]_\beta = \Big([T(\mathbf{n})]_\beta \; [T(\mathbf{p})]_\beta \; [T(\mathbf{q})]_\beta \Big)$$
   $$= \Big([\mathbf{0}]_\beta \; [\mathbf{p}]_\beta \; [\mathbf{q}]_\beta \Big) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   Since $\mathbf{p}$ and $\mathbf{q}$ are perpendicular ($\mathbf{p} \cdot \mathbf{n} = 0 = \mathbf{q} \cdot \mathbf{n}$) to $\mathbf{n}$, the matrix makes the physical interpretation of the linear map clear: T is the *orthogonal projection* onto the subspace $\text{Span}\{\mathbf{p}, \mathbf{q}\}$. It should also be from the map's interpretation as a projection that $\mathcal{N}(T) = \text{Span}\{\mathbf{n}\}$ and $\mathcal{R}(T) = \text{Span}\{\mathbf{p}, \mathbf{q}\}$.

Composition and Matrix Multiplication

It seems reasonable to expect that composition of linear maps corresponds to matrix multiplication. We merely have to be (very) careful with bases! Before seeing this, we engage in a little book-keeping.

Definition 2.26. The *identity matrix* $I_n \in M_n(\mathbb{F})$ has ij^{th} entry the *Kronecker delta symbol*

$$(I_n)_{ij} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Lemma 2.27. *Let V be an n-dimensional vector space with basis β and let $\mathrm{T} \in \mathcal{L}(V)$. Then*

$$[\mathrm{T}]_\beta = I_n \iff \mathrm{T} = \mathrm{I} \text{ is the identity map on } V$$

Proof. $(\Leftarrow)$ If $\mathrm{T} = \mathrm{I}$, then $\mathrm{T}(\mathbf{v}_i) = \mathbf{v}_i$ for each $\mathbf{v}_i \in \beta$. Plainly $[\mathrm{T}]_\beta = I_n$.
$(\Rightarrow)$ By the uniqueness of the matrix representation, $\mathrm{T} = \mathrm{I}$ is the only linear map with matrix I_n. ∎

Theorem 2.28. *Suppose $\mathrm{T} \in \mathcal{L}(V, W)$ and $\mathrm{U} \in \mathcal{L}(W, X)$ are linear maps and that V, W, X are finite-dimensional with bases β, γ, δ respectively. Then*

$$[\mathrm{UT}]_\beta^\delta = [\mathrm{U}]_\gamma^\delta [\mathrm{T}]_\beta^\gamma$$

In the common situation where $V = W = X$ and $\beta = \gamma = \delta$, this reduces to $[\mathrm{UT}]_\beta = [\mathrm{U}]_\beta [\mathrm{T}]_\beta$

Proof. Label the bases $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_l\}$, $\gamma = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ and $\delta = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, and the matrices $A = [\mathrm{T}]_\beta^\gamma$, $B = [\mathrm{U}]_\gamma^\delta$ and $C = [\mathrm{UT}]_\beta^\delta$. Observe first

$$[\mathrm{T}(\mathbf{v}_k)]_\gamma = \begin{pmatrix} A_{1k} \\ \vdots \\ A_{mk} \end{pmatrix} \implies \mathrm{T}(\mathbf{v}_k) = \sum_{j=1}^m \mathbf{w}_j A_{jk} \qquad (k^{\text{th}} \text{ column of } A)$$

$$[\mathrm{U}(\mathbf{w}_j)]_\delta = \begin{pmatrix} B_{1j} \\ \vdots \\ B_{nj} \end{pmatrix} \implies \mathrm{U}(\mathbf{w}_j) = \sum_{i=1}^n \mathbf{x}_i B_{ij} \qquad (j^{\text{th}} \text{ column of } B)$$

$$[\mathrm{UT}(\mathbf{v}_k)]_\delta = \begin{pmatrix} C_{1k} \\ \vdots \\ C_{nk} \end{pmatrix} \implies \mathrm{UT}(\mathbf{v}_k) = \sum_{i=1}^n \mathbf{x}_i C_{ik} \qquad (k^{\text{th}} \text{ column of } C)$$

Now put it together:

$$\sum_{i=1}^n \mathbf{x}_i C_{ik} = \mathrm{UT}(\mathbf{v}_k) = \mathrm{U}\left(\sum_{j=1}^m \mathbf{w}_j A_{jk}\right) = \sum_{j=1}^m \mathrm{U}(\mathbf{w}_j) A_{jk} = \sum_{j=1}^m \sum_{i=1}^n \mathbf{x}_i B_{ij} A_{jk} = \sum_{i=1}^n \mathbf{x}_i \left(\sum_{j=1}^m B_{ij} A_{jk}\right)$$

Since $\delta = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a basis we conclude that

$$C_{ik} = \sum_{j=1}^m B_{ij} A_{jk}$$

Otherwise said, $C = BA$. ∎

Taking the special case where $U = T^{-1}$ and $\delta = \beta$, we instantly conclude:

Corollary 2.29. *Suppose $T \in \mathcal{L}(V, W)$ is a map between n-dimensional spaces V, W with bases β, γ. Then T is an isomorphism if and only if its matrix $[T]_\beta^\gamma \in M_n(\mathbb{F})$ is invertible. Moreover*

$$[T^{-1}]_\gamma^\beta = \left([T]_\beta^\gamma\right)^{-1}$$

Examples 2.30. 1. Recall (Example 2.25.1) that the matrix of 'rotate clockwise by 30°' with respect to the standard basis $\epsilon = \{\mathbf{i}, \mathbf{j}\}$ of $\mathbb{R}^2$ is

$$[T]_\epsilon = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$$

It follows that T^2 (rotate clockwise by 60°) and T^3 (90°) have matrices

$$[T^2]_\epsilon = [T]_\epsilon[T]_\epsilon = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}^2 = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} \quad \text{and} \quad [T^3]_\epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

Moreover, the inverse of T (namely 'rotate 30° *counter-clockwise*') has matrix

$$[T^{-1}]_\epsilon = [T]_\epsilon^{-1} = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$$

2. Recall Example 2.25.4, and suppose $T \in \mathcal{L}(\mathbb{R}^3)$ is projection onto the plane perpendicular to $\mathbf{n} = -\mathbf{i} + 2\mathbf{k}$. Also let U be rotation by 60° clockwise around the $\mathbf{k}$-axis when viewed from above. With respect to the standard basis $\epsilon = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, and following the previous example,

$$[UT]_\beta = [U]_\beta[T]_\beta = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{4}{5} & 0 & \frac{2}{5} \\ 0 & 1 & 0 \\ \frac{2}{5} & 0 & \frac{1}{5} \end{pmatrix} = \frac{1}{10}\begin{pmatrix} 4 & 5\sqrt{3} & 2 \\ -4\sqrt{3} & 5 & -2\sqrt{3} \\ 4 & 0 & 2 \end{pmatrix}$$

3. Let $\beta = \{e^{-x}\cos 2x, e^{-x}\sin 2x\}$, $V = \operatorname{Span}\beta$ and consider $T = \frac{d}{dx} \in \mathcal{L}(V)$. By computing

$$T(e^{-x}\cos 2x) = -e^{-x}\cos 2x - 2e^{-x}\sin 2x, \qquad T(e^{-x}\sin 2x) = 2e^{-x}\cos 2x - e^{-x}\sin 2x$$

we see that the matrix of T with respect to β, and its inverse are

$$[T]_\beta = \begin{pmatrix} -1 & 2 \\ -2 & -1 \end{pmatrix} \implies [T^{-1}]_\beta = [T]_\beta^{-1} = \frac{1}{5}\begin{pmatrix} -1 & -2 \\ 2 & -1 \end{pmatrix}$$

Since T^{-1} computes *anti-derivatives*, the upshot is that we can do this using linear algebra!

$$\int ae^{-x}\cos 2x + be^{-x}\sin 2x \, dx = -\frac{a+2b}{5}e^{-x}\cos 2x + \frac{2a-b}{5}e^{-x}\sin 2x$$

Of course if you *really* prefer integration by parts...

A final bit of book-keeping: co-ordinate isomorphisms and matrices

Two colloquialisms are sometimes uttered in an attempt to summarize or simplify linear algebra.

1. Every vector space is really $\mathbb{F}^n$ in disguise.
2. Every linear map is really matrix multiplication in disguise.

While strictly incorrect, we can make these statements precise, at least under the imposition of two caveats: *when dimensions are finite* and *after choosing bases.*

Corollary 2.31. 1. *Suppose β is a basis of V and that $\dim_{\mathbb{F}} V = n$. Then the co-ordinate representation ϕ_β is an isomorphism*

$$\phi_\beta : V \to \mathbb{F}^n : \mathbf{v} \mapsto [\mathbf{v}]_\beta$$

2. *Additionally, suppose γ is a basis of W and that $\dim_{\mathbb{F}} W = m$. Then the vector space of linear maps $\mathcal{L}(V, W)$ is isomorphic to the space of matrices $M_{m\times n}(\mathbb{F})$ via the isomorphism*

$$\Phi_\beta^\gamma : \mathcal{L}(V, W) \to M_{m\times n}(\mathbb{F}) : \mathrm{T} \mapsto [\mathrm{T}]_\beta^\gamma$$

This is really just the finite-dimensional version of Corollary 2.20, but with explicit isomorphisms.

Proof. The co-ordinate representation ϕ_β is linear: if $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\} \subseteq \mathbb{F}^n$ is the standard basis,

$$\phi_\beta(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n = a_1\phi_\beta(\mathbf{v}_1) + \cdots + a_n\phi_\beta(\mathbf{v}_n)$$

Since $\mathcal{N}(\phi_\beta) = \{\mathbf{0}_V\}$, it is also plainly injective and thus an isomorphism.
Part 2 is similar. ∎

A peculiar difficulty with this discussion is that it can be hard to disentangle a *matrix* $A \in M_{m\times n}(\mathbb{F})$ from its associated *linear map* $\mathrm{L}_A : \mathbb{F}^n \to \mathbb{F}^m$. For reference, we summarize everything here.

Corollary 2.32. *Let $A, B \in M_{m\times n}(\mathbb{F})$.*

1. *If β, γ are the standard bases, then $[\mathrm{L}_A]_\beta^\gamma = A$*
2. $\mathrm{L}_A = \mathrm{L}_B \iff A = B$
3. $\mathrm{L}_{A+B} = \mathrm{L}_A + \mathrm{L}_B$ *and* $\mathrm{L}_{\lambda A} = \lambda \mathrm{L}_A$ *for all* $\lambda \in \mathbb{F}$
4. *If $\mathrm{T} \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$, then there is a unique $C \in M_{m\times n}(\mathbb{F})$ such that $\mathrm{T} = \mathrm{L}_C$*
5. *If $E \in M_{n\times p}(\mathbb{F})$, then $\mathrm{L}_{AE} = \mathrm{L}_A\mathrm{L}_E$*
6. *If $m = n$, then $\mathrm{L}_{I_n} = \mathrm{I}$*

Everything should be straightforward to prove given our previous results.

Exercises 2.3 1. Let $\mathrm{T} = \mathrm{L}_A \in \mathcal{L}(\mathbb{R}^2)$ be left-multiplication by $A = \begin{pmatrix} 1 & -2 \\ 3 & -1 \end{pmatrix}$.

   (a) Find the matrix $[\mathrm{T}]_\beta$ with respect to the basis $\beta = \{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}\}$.

   (b) Compute $\mathrm{T}(3\mathbf{i} + 4\mathbf{j})$ in two different ways, and make sure your answers agree!

2. Consider the linear map $\mathrm{T} = \frac{\mathrm{d}}{\mathrm{d}x} \in \mathcal{L}(P_3(\mathbb{R}), P_2(\mathbb{R}))$.

   (a) Compute $[\mathrm{T}]_\beta^\gamma$ with respect to the bases $\beta = \{1, x, x^2, x^3\}$ and $\gamma = \{1 - x, 1 + x, x^2 - 1\}$.

   (b) Verify that multiplication by $[\mathrm{T}]_\beta^\gamma$ correctly computes the derivative of the polynomial $p(x) = 2 + 5x + 3x^2 + x^3$.

   (c) Let $\mathrm{U} \in \mathcal{L}(P_2(\mathbb{R}), P_1(\mathbb{R}))$ also be 'differentiate,' so that $\mathrm{UT} = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ is the second-derivative. Compute the matrices of T, U and check that $[\mathrm{UT}]_\beta^\delta = [\mathrm{U}]_\gamma^\delta [\mathrm{T}]_\beta^\gamma$ when:

      i. β, γ and δ are the standard bases of $P_3(\mathbb{R})$, $P_2(\mathbb{R})$ and $P_1(\mathbb{R})$ respectively.

      ii. $\beta = \{1, x, x^2, x^3\}$, $\gamma = \{1 - x, 1 + x, x^2 - 1\}$ and $\delta = \{1, x\}$

3. Define $\mathrm{T} \in \mathcal{L}(P_3(\mathbb{R}), P_4(\mathbb{R}))$ by $\mathrm{T}(g)(x) = (x-1)g(x)$, let $f(x) = x + 2x^2 - 3x^3$ and suppose β, γ are the standard bases.

   (a) Compute $[f]_\beta$ and $[(x-1)f]_\gamma$.

   (b) Compute the matrix $[\mathrm{T}]_\beta^\gamma$ and check explicitly that $[(x-1)f]_\gamma = [\mathrm{T}]_\beta^\gamma [f]_\beta$.

4. Let $\mathrm{T} : P_3(\mathbb{R}) \to P_3(\mathbb{R})$ be the function defined by

$$\mathrm{T}(f)(x) = 2\int_0^1 f(t)\,\mathrm{d}t - 3\frac{\mathrm{d}}{\mathrm{d}x} f(x)$$

   (a) Give a short argument to justify the fact that T is linear.

   (b) Compute the matrix $[\mathrm{T}]_\beta$ of T with respect to the standard basis β of $P_3(\mathbb{R})$.

   (c) Find an explicit expression for the linear map $\mathrm{T}^2 \in \mathcal{L}(P_3(\mathbb{R}))$; that is, express $\mathrm{T}^2(f)(x)$ in terms of the integral and derivatives of $f(x)$.

   (d) Compute $[\mathrm{T}^2]_\beta$ using part (a), and check that it equals $[\mathrm{T}]_\beta^2$.

5. (*Recall Example 2.25.3*) Let $\mathrm{T} \in \mathcal{L}(\mathbb{R}^2)$ be the linear map 'reflect across the line $y = 2x$.' With respect to the standard basis, show that its matrix is $[\mathrm{T}]_\epsilon = \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix}$.

6. Find a basis of $\mathbb{R}^2$ with respect to which the linear map 'reflect across the line $x + 3y = 0$' has a diagonal matrix. Now find the matrix of this map with respect to the standard basis.

7. Let B be a fixed invertible $n \times n$ matrix. Prove that the following map is an isomorphism:

$$\Psi : M_n(\mathbb{F}) \to M_n(\mathbb{F}) : A \mapsto B^{-1}AB$$

8. Compute the integral $\int (2x - 3x^2)e^{3x}\,\mathrm{d}x$ *without* using integration by parts.

   (*Hint: Let $\beta = \{e^{3x}, xe^{3x}, x^2e^{3x}\}$ and invert the matrix of $\frac{\mathrm{d}}{\mathrm{d}x}$ with respect to β...*)

9. Give explicit proofs of Corollary 2.32 parts 1 & 5.

10. In the context of Corollary 2.31; suppose $\dim_{\mathbb{F}} V = n$ and that $\mathrm{T} : V \to \mathbb{F}^n$ is an isomorphism. Prove that $\mathrm{T} = \phi_\beta$ for some basis β.

2.4 The Change of Co-ordinate Matrix

Suppose V is finite-dimensional over $\mathbb{F}$ with distinct bases β, ϵ. We know from Corollary 2.31 that the co-ordinate maps are isomorphisms $V \to \mathbb{F}^n$:

$$\phi_\beta(\mathbf{v}) = [\mathbf{v}]_\beta, \qquad \phi_\epsilon(\mathbf{v}) = [\mathbf{v}]_\epsilon$$

Since inverses and compositions of isomorphisms are also isomorphisms, it follows that

$$\phi_\epsilon \circ \phi_\beta^{-1} : \mathbb{F}^n \to \mathbb{F}^n : [\mathbf{v}]_\beta \mapsto [\mathbf{v}]_\epsilon$$

is an isomorphism. Corollaries 2.32 and 2.31 force this isomorphism to be left-multiplication by an invertible matrix:

$$\exists Q_\beta^\epsilon \in M_n(\mathbb{F}) \text{ such that } \forall \mathbf{v} \in V,\ [\mathbf{v}]_\epsilon = \phi_\epsilon \circ \phi_\beta^{-1}(\mathbf{v}) = Q_\beta^\epsilon [\mathbf{v}]_\beta \qquad (*)$$

Definition 2.33. Q_β^ϵ is the *change of co-ordinate matrix* from β to ϵ.

Indeed $(*)$ makes it obvious how to compute: if $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ then $[\mathbf{v}_j]_\beta = \mathbf{e}_j$ is the j^{th} standard basis (column) vector of $\mathbb{F}^n$, and so…

Lemma 2.34. *The change of co-ordinate matrix from β to ϵ is the matrix of the identity linear map with respect to these bases: if $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then,*

$$Q_\beta^\epsilon = [\mathrm{I}_V]_\beta^\epsilon = \begin{pmatrix} [\mathbf{v}_1]_\epsilon & \cdots & [\mathbf{v}_n]_\epsilon \end{pmatrix}$$

It follows immediately that $Q_\epsilon^\beta = (Q_\beta^\epsilon)^{-1}$ and that $Q_\epsilon^\gamma Q_\beta^\epsilon = Q_\beta^\gamma$.

We are prioritizing ϵ in the above notation because, in many situations, one of the bases is a standard basis. In such a case, one can simply state Q_β^ϵ and invert to obtain Q_ϵ^β, as the next example illustrates.

Example 2.35. Consider the basis $\beta = \{1 - 3x, 2 + 5x\}$ of $P_1(\mathbb{R})$ and let $\epsilon = \{1, x\}$ be the standard basis. Then

$$Q_\beta^\epsilon = [\mathrm{I}_{P_1(\mathbb{R})}]_\beta^\epsilon = \begin{pmatrix} [1-3x]_\epsilon & [2+5x]_\epsilon \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -3 & 5 \end{pmatrix} \implies Q_\epsilon^\beta = \begin{pmatrix} 1 & 2 \\ -3 & 5 \end{pmatrix}^{-1} = \frac{1}{11}\begin{pmatrix} 5 & -2 \\ 3 & 1 \end{pmatrix}$$

To write, for instance $p(x) = 3 - x$ in terms of β, compute

$$[p]_\beta = Q_\epsilon^\beta [p]_\epsilon = \frac{1}{11}\begin{pmatrix} 5 & -2 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ -1 \end{pmatrix} = \frac{1}{11}\begin{pmatrix} 17 \\ 8 \end{pmatrix}$$
$$\implies 3 - x = \frac{17}{11}(1 - 3x) + \frac{8}{11}(2 + 5x)$$

While this approach doesn't save any time for a single calculation, it is much more efficient when one needs to convert many vectors to another basis. It is important to remember that the change of co-ordinate matrix merely tells you how the *co-ordinates* of a vector $\mathbf{v} \in V$ change when a basis changes: nothing happens to $\mathbf{v}$ itself!

Example 2.36. Here is a 3-dimensional example. Consider the basis $\beta = \{1+x, 2-x^2, 4-x^2\}$ of $P_2(\mathbb{R})$ and let $\epsilon = \{1, x, x^2\}$ be the standard basis. Then

$$Q_\beta^\epsilon = \Big([1+x]_\epsilon \ [2-x^2]_\epsilon \ [4-x^2]_\epsilon\Big) = \begin{pmatrix} 1 & 2 & 4 \\ 1 & 0 & 0 \\ 0 & -1 & -1 \end{pmatrix} \implies Q_\epsilon^\beta = \frac{1}{2}\begin{pmatrix} 0 & 2 & 0 \\ -1 & 1 & -4 \\ 1 & -1 & 2 \end{pmatrix}$$

To check that this makes sense, we check the co-ordinate representation of, say, $p(x) = 2 + 3x + 4x^2$ with respect to β:

$$[p]_\beta = Q_\epsilon^\beta [p]_\epsilon = \frac{1}{2}\begin{pmatrix} 0 & 2 & 0 \\ -1 & 1 & -4 \\ 1 & -1 & 2 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ -\frac{15}{2} \\ \frac{7}{2} \end{pmatrix}$$

$$p(x) = 3(1+x) - \frac{15}{2}(2-x^2) + \frac{7}{2}(4-x^2)$$

which is easily verified to by multiplying out. Of course, all this is predicated on being willing to invert a 3×3 matrix!

This process can be combined with matrix representations of linear maps.

Theorem 2.37. *Let* $\mathrm{T} \in \mathcal{L}(V)$ *where* V *has finite bases* ϵ *and* β*. Then the matrices of* T *satisfy*

$$[\mathrm{T}]_\beta = Q_\epsilon^\beta [\mathrm{T}]_\epsilon Q_\beta^\epsilon = Q_\epsilon^\beta [\mathrm{T}]_\epsilon (Q_\epsilon^\beta)^{-1}$$

Proof. Simply apply the right hand side to the representation of any vector $\mathbf{v} \in V$ with respect to β:

$$Q_\epsilon^\beta [\mathrm{T}]_\epsilon Q_\beta^\epsilon [\mathbf{v}]_\beta = Q_\epsilon^\beta [\mathrm{T}]_\epsilon [\mathbf{v}]_\epsilon = Q_\epsilon^\beta [\mathrm{T}(\mathbf{v})]_\epsilon = [\mathrm{T}(\mathbf{v})]_\beta = [\mathrm{T}]_\beta [\mathbf{v}]_\beta$$

■

The matrices of a linear map with respect to different bases are therefore *similar/conjugate.*

Example 2.38. We revisit Example 2.25.3 in this language. Let $\mathrm{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be reflection in the line $y = 2x$, let $\epsilon = \{\mathbf{i}, \mathbf{j}\}$ be the standard basis and $\beta = \{\mathbf{v}_1, \mathbf{v}_2\} = \{\mathbf{i} + 2\mathbf{j}, -2\mathbf{i} + \mathbf{j}\}$ be chosen to point parallel/perpendicular to the line of reflection. Since $\mathrm{T}(\mathbf{v}_1) = \mathbf{v}_1$ and $\mathrm{T}(\mathbf{v}_2) = -\mathbf{v}_2$ we saw that

$$[\mathrm{T}]_\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The change of co-ordinate matrices are then

$$Q_\beta^\epsilon = [\mathrm{I}_{\mathbb{R}^2}]_\beta^\epsilon = ([\mathbf{v}_1]_\epsilon \ [\mathbf{v}_2]_\epsilon) = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}, \qquad Q_\epsilon^\beta = (Q_\beta^\epsilon)^{-1} = \frac{1}{5}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$$

The matrix of T with respect to the standard basis ϵ is therefore

$$[\mathrm{T}]_\epsilon = Q_\beta^\epsilon [\mathrm{T}]_\beta Q_\epsilon^\beta = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \frac{1}{5}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix}$$

as we recovered earlier.

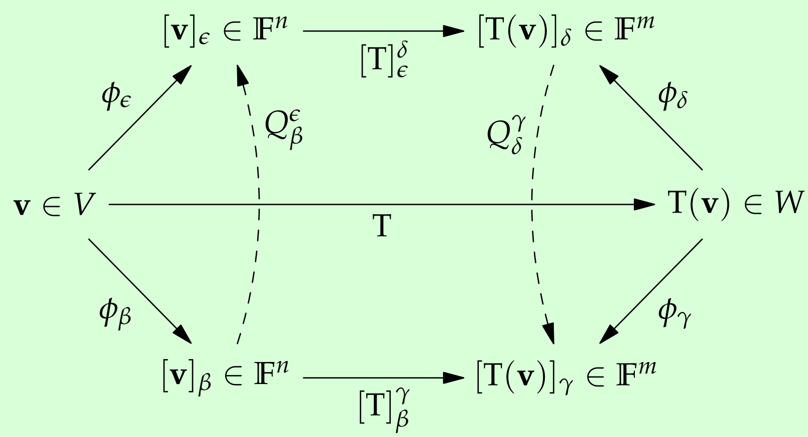
Change of basis in general (non-examinable)

The discussion generalizes to linear maps $\mathrm{T} \in \mathcal{L}(V, W)$ where we change bases of *both spaces*.

Theorem 2.39. *Suppose* $\mathrm{T} \in \mathcal{L}(V, W)$*, that* V *has bases* β, ϵ*, and* W *has bases* γ, δ*, where* $\dim V = n$ *and* $\dim W = m$*. Then*

$$[\mathrm{T}]_\beta^\gamma = Q_\delta^\gamma [\mathrm{T}]_\epsilon^\delta Q_\beta^\epsilon$$

where $Q_\beta^\epsilon \in M_n(\mathbb{F})$ *and* $Q_\delta^\gamma \in M_m(\mathbb{F})$ *are change of co-ordinate matrices. The relationships between these objects is summarized in the picture:*



Example 2.40. With respect to the standard bases $\epsilon = \{1, x, x^2, x^3\}$ and $\delta = \{1, x, x^2\}$, the derivative operator $\mathrm{T} = \frac{\mathrm{d}}{\mathrm{d}x} \in \mathcal{L}(P_3(\mathbb{R}), P_2(\mathbb{R}))$ has matrix

$$[\mathrm{T}]_\epsilon^\delta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Consider new bases $\beta = \{1 + x, 1 - x, 2x + x^2, x^3 - 1\}$ and $\gamma = \{1 - x, 2 + x^2, x\}$ of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$ respectively. The matrix of T with respect to β and γ is then

$$[\mathrm{T}]_\beta^\gamma = (Q_\gamma^\delta)^{-1} [\mathrm{T}]_\epsilon^\delta Q_\beta^\epsilon = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & -1 \\ 1 & -1 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 2 & -6 \\ 0 & 0 & 0 & 3 \\ 1 & -1 & 4 & -6 \end{pmatrix}$$

We can check this on an example: written with respect to β, let

$$p(x) = 3(1 + x) + 2(1 - x) - 4(2x + x^2) + 5(x^3 - 1)$$

$$\implies [p']_\gamma = [\mathrm{T}]_\beta^\gamma [p]_\beta = \begin{pmatrix} 1 & -1 & 2 & -6 \\ 0 & 0 & 0 & 3 \\ 1 & -1 & 4 & -6 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ -4 \\ 5 \end{pmatrix} = \begin{pmatrix} -37 \\ 15 \\ -45 \end{pmatrix}$$

which comports with

$$p'(x) = 3 - 2 - 4(2 + 2x) + 15x^2 = -37(1 - x) + 15(2 + x^2) - 45x$$

Exercises 2.4 1. Let $\epsilon = \{\mathbf{i}, \mathbf{j}\}$ be the standard basis of $\mathbb{R}^2$, and consider two further bases

$$\beta = \{-\mathbf{i} + 2\mathbf{j},\ 2\mathbf{i} - \mathbf{j}\}, \qquad \gamma = \{2\mathbf{i} + 5\mathbf{j},\ -\mathbf{i} - 3\mathbf{j}\}$$

Find the change of co-ordinate matrices Q_ϵ^β, Q_ϵ^γ and Q_β^γ.

2. Let $\epsilon = \{1, x, x^2\}$ and $\beta = \{1 - x, x + x^2, x^2 - 1\}$. Find the change of co-ordinate matrix Q_ϵ^β for $P_2(\mathbb{F})$. Check your answer by finding constants a, b, c such that

$$2 + 7x - 4x^2 = a(1 - x) + b(x + x^2) + c(x^2 - 1)$$

3. For each matrix A and basis β, find $[\mathrm{L}_A]_\beta$ and an invertible matrix Q such that $[\mathrm{L}_A]_\beta = Q^{-1}AQ$.

   (a) $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ and $\beta = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$

   (b) $A = \begin{pmatrix} 13 & 1 & 4 \\ 1 & 13 & 4 \\ 4 & 4 & 10 \end{pmatrix}$ and $\beta = \left\{ \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$

4. Recall the *trace* of a $n \times n$ matrix: $\operatorname{tr} C = \sum_{j=1}^{n} c_{jj}$.

   (a) Prove that $\operatorname{tr} AB = \operatorname{tr} BA$, provided both AB and BA are square.

   (b) Prove that if A and B are similar matrices ($B = Q^{-1}AQ$ for some Q), then $\operatorname{tr} A = \operatorname{tr} B$. (*The matrices of a linear map with respect to any two bases therefore always have the same trace*)

5. Suppose $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of $\mathbb{F}^n$ and that $\epsilon = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis. Prove that $Q_\epsilon^\beta \mathbf{v}_k = \mathbf{e}_k$ for each k.

   (*In this context, Q_ϵ^β is sometimes called a* change of basis *matrix, though this only makes sense in $\mathbb{F}^n$*)

6. Let R reflect in the line through the origin making angle θ with the positive x-axis in $\mathbb{R}^2$.

   (a) As in Example 2.38, use a change of co-ordinate matrix to find the matrix of R with respect to the standard basis $\epsilon = \{\mathbf{i}, \mathbf{j}\}$.

   (b) If the line of reflection has gradient m, state $[\mathrm{R}]_\epsilon$ in terms of m. When m is a rational number, what does this have to do with Pythagorean triples?

7. Let c and s be constants and consider the change of co-ordinates

$$\begin{cases} x = cu + sv \\ y = -su + cv \end{cases} \qquad (*)$$

   That is, if $\mathbf{x} = [\mathbf{x}]_\epsilon = \begin{pmatrix} x \\ y \end{pmatrix}$ is viewed with respect to $\epsilon = \{\mathbf{i}, \mathbf{j}\}$, then $[\mathbf{x}]_\beta = \begin{pmatrix} u \\ v \end{pmatrix}$ with respect to some new basis β.

   (a) Find β.

   (b) The curve

$$7x^2 - 6\sqrt{3}xy + 13y^2 = 16$$

   represents a conic in the plane. Assume that $c = \cos\theta$ and $s = \sin\theta$ for some unknown angle θ. Substitute, using $(*)$, for u and v in order to find a value of $\theta \in [0, 90°)$ for which the conic has no uv-term.

   Use your understanding of the basis β and the resulting change of co-ordinates to sketch the original conic.

3 Elementary Matrix Operations and Systems of Linear Equations

3.1 Elementary Matrix Operations and Elementary Matrices

In this chapter we develop a (hopefully!) familiar method for comparing matrices.

Definition 3.1. An *elementary row operation* is one of three transformations of the rows of a matrix:

Type I: Swap two rows;

Type II: Multiply a row by a non-zero constant;

Type III: Add to one row a scalar multiple of another.

Matrices are *row equivalent* if there exists a finite sequence of elementary row operations transforming one to the other.

The *elementary matrices* come in the same three families, each is the result of performing the corresponding row operation to the identity matrix:

Type I: E_{ij} is the identity matrix with rows i, j swapped;

Type II: $E_i^{(\lambda)}$ is the identity with the i^{th} diagonal entry replaced by $\lambda \neq 0$;

Type III: $E_{ij}^{(\lambda)}$ is the identity matrix with an additional λ in the ij^{th} entry.

Example 3.2. In $M_2(\mathbb{R})$ the elementary matrices are as follows:

$$E_{12} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad E_1^{(\lambda)} = \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}, \quad E_2^{(\lambda)} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}, \quad E_{12}^{(\lambda)} = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}, \quad E_{21}^{(\lambda)} = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$$

By subtracting three times the first row from the second, we see that the following are row equivalent:

$$\begin{pmatrix} 1 & 4 \\ -2 & 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 4 \\ -5 & -9 \end{pmatrix}$$

The crucial observation, stated in general below, is that this transformation is the result of multiplying by the corresponding elementary matrix:

$$E_{21}^{(-3)} \begin{pmatrix} 1 & 4 \\ -2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ -2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ -5 & -9 \end{pmatrix}$$

Theorem 3.3. *Let* T *be an elementary row operation acting on* $m \times n$ *matrices.*

1. T *is an isomorphism of* $M_{m\times n}(\mathbb{F})$ *with itself. Its inverse is an operation of the same type.*
2. $\mathrm{T}(A) = EA$ *where* E *is the elementary matrix* $\mathrm{T}(I_m)$ *obtained by applying* T *to the identity.*

In particular, the inverses of the three types of elementary matrix are

$$E_{ij}^{-1} = E_{ij}, \qquad \left(E_i^{(\lambda)}\right)^{-1} = E_i^{(\lambda^{-1})}, \qquad \left(E_{ij}^{(\lambda)}\right)^{-1} = E_{ij}^{(-\lambda)}$$

Proof. Note first that row operations never mix columns and neither does matrix multiplication: if $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the columns of A, then

$$\mathrm{T}(A) = \big(\mathrm{T}(\mathbf{a}_1) \cdots \mathrm{T}(\mathbf{a}_n)\big) \quad \text{and} \quad EA = \big(E\mathbf{a}_1 \cdots E\mathbf{a}_n\big)$$

It is therefore enough to prove the result when $n = 1$.

1. Suppose that T is of type III, adding to row i a multiple λ of row j. Let v_r, w_r denote the r^{th} entries of $\mathbf{v}, \mathbf{w} \in \mathbb{F}^m$ and let k be a scalar. The r^{th} entry of $\mathrm{T}(k\mathbf{v} + \mathbf{w})$ is plainly

$$\begin{cases} (kv_i + w_i) + \lambda(kv_j + w_j) = k(v_i + \lambda v_j) + (w_i + \lambda w_j) & \text{if } r = i \\ kv_r + w_r & \text{if } r \neq i \end{cases}$$

   which is certainly the r^{th} entry of $k\mathrm{T}(\mathbf{v}) + \mathrm{T}(\mathbf{w})$: thus $\mathrm{T} : \mathbb{F}^m \to \mathbb{F}^m$ is linear. Its inverse is plainly computed by *subtracting* from the i^{th} row λ times the j^{th}: an elementary operation of the same type. Operations of types I and II are similar.

2. By part 1, $\mathrm{T} = \mathrm{L}_E \in \mathcal{L}(\mathbb{F}^m)$ for some invertible matrix $E \in M_m(\mathbb{F})$. To compute E, simply apply T to the standard basis $\epsilon = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$:

$$E = \big([\mathrm{T}(\mathbf{e}_1)]_\epsilon \cdots [\mathrm{T}(\mathbf{e}_m)]_\epsilon\big) = \big(\mathrm{T}(\mathbf{e}_1) \cdots \mathrm{T}(\mathbf{e}_m)\big) = \mathrm{T}(\mathbf{e}_1 \cdots \mathbf{e}_n) = \mathrm{T}(I_m)$$

■

Column Operations Applying the above approach to columns yields the *elementary column operations.* Theorem 3.3 holds for column operations provided you multiply by matrices *on the right* $\mathrm{T}(A) = AE$ and replace 'row' with 'column.'

Example 3.4. Let $E = E_{21}^{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and compute:

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a + 3b & b & c \\ d + 3e & e & f \end{pmatrix}$$

E can be produced from the identity matrix by adding three times the second column to the first, precisely the effect it has as a column operation when multiplying on the right.

Exercises 3.1 1. Let $A = \begin{pmatrix} 2 & 4 \\ 1 & 7 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix}$.

   (a) Find a sequence of elementary matrices E^I, E^{II}, E^{III}, of the types indicated, so that

   $$B = E^{II} E^I E^{III} A$$

   (b) Hence find a matrix C such that $B = CA$. Is C the *only* matrix satisfying this equation?

   (c) Find another sequence of elementary matrices such that $B = E_k \cdots E_1 A$.

2. Let A be an $m \times n$ matrix. Prove that if B can be obtained from A by an elementary row operation, then B^T can be obtained from A^T by the corresponding elementary *column* operation. (*This essentially proves Theorem 3.3 for column operations.*)

3. For the matrices A, B in question 1, find a sequence of elementary matrices of any length/type such that $B = AE_1 \cdots E_k$.

3.2 The Rank of a Matrix and Matrix Inverses

Our goal is to use the row equivalence of matrices to provide systematic methods for computing ranks and inverses of linear maps. First we translate the notions of rank and nullity to matrices.

Definition 3.5. The *rank* and *nullity* of a matrix $A \in M_{m\times n}(\mathbb{F})$ are the rank/nullity of the linear map $\mathrm{L}_A : \mathbb{F}^n \to \mathbb{F}^m$ given by left-multiplication by A.

Our previous injectivity and surjectivity conditions immediately translate to this new language.

Lemma 3.6. *Let $A \in M_{m\times n}(\mathbb{F})$ and $\mathrm{L}_A : \mathbb{F}^n \to \mathbb{F}^m$ be the corresponding linear map.*

1. L_A *is injective* $\iff \operatorname{rank} A = n \iff \operatorname{null} A = 0$
2. L_A *is surjective* $\iff \operatorname{rank} A = m \iff \operatorname{null} A = n - m$
3. *(When $m = n$)* L_A *is an isomorphism* $\iff$ *A is invertible* $\iff \operatorname{rank} A = n \iff \operatorname{null} A = 0$

We now come to the crucial observations that permit easy calculations of ranks and inverses.

Theorem 3.7. *Let $A = (\mathbf{a}_1, \ldots, \mathbf{a}_n) \in M_{m\times n}(\mathbb{F})$ have columns $\mathbf{a}_j \in \mathbb{F}^n$.*

1. *The* column space $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ *of A has dimension* $\operatorname{rank} A$.
2. $\operatorname{rank} A$ *is invariant under multiplication by invertible matrices: if P and Q are invertible, then*

$$\operatorname{rank} PA = \operatorname{rank} AQ = \operatorname{rank} A$$

Proof. 1. For any vector $\mathbf{v} \in \mathbb{F}^n$, write $\mathbf{v} = \lambda_1 \mathbf{e}_1 + \cdots + \lambda_n \mathbf{e}_n$ with respect to the standard basis and observe that

$$\lambda_1 \mathbf{a}_1 + \cdots + \lambda_n \mathbf{a}_n = A\mathbf{v} = \mathrm{L}_A(\mathbf{v})$$

It is clear from this that the column space and $\mathcal{R}(L_A)$ are identical, whence $\operatorname{rank} A$ is the dimension of the column space.

2. We work with the range of the linear map $\mathrm{L}_{PA} = \mathrm{L}_P \mathrm{L}_A (= \mathrm{L}_P \circ \mathrm{L}_A)$:

$$\mathcal{R}(\mathrm{L}_{PA}) = \mathcal{R}(\mathrm{L}_P \mathrm{L}_A) = \{PA\mathbf{v} : \mathbf{v} \in \mathbb{F}^n\} = \mathrm{L}_P(\mathcal{R}(\mathrm{L}_A))$$

Since P is invertible, $\mathrm{L}_P : \mathcal{R}(\mathrm{L}_A) \to \mathrm{L}_P(\mathcal{R}(\mathrm{L}_A))$ is an isomorphism, whence

$$\dim \mathcal{R}(\mathrm{L}_A) = \dim \mathrm{L}_P(\mathcal{R}(\mathrm{L}_A)) = \dim \mathcal{R}(\mathrm{L}_{PA}) \implies \operatorname{rank} A = \operatorname{rank} PA$$

The result for AQ is even easier: we leave it to an exercise. ∎

By virtue of the theorem, we'll denote the column space and nullspace of A by $\mathcal{R}(A)$ and $\mathcal{N}(A)$ respectively rather than the lengthier $\mathcal{R}(\mathrm{L}_A)$ and $\mathcal{N}(\mathrm{L}_A)$.

Computing the Rank of a Matrix Recall that elementary row/column operations act via multiplication by invertible matrices: thus

> Elementary row/column operations are rank-preserving

Examples 3.8. 1. Recall Example 3.2, where we saw the row equivalence of $\left(\begin{smallmatrix}1&4\\-2&3\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&4\\-5&-9\end{smallmatrix}\right)$. Since the columns of these are linearly independent, the column spaces of both are $\mathbb{R}^2$ and both matrices plainly have rank 2. Indeed we can perform a sequence of row operations that make the rank even more obvious:

$$\begin{pmatrix}1&4\\-2&3\end{pmatrix}\xrightarrow{E_{21}^{(2)}}\begin{pmatrix}1&4\\0&11\end{pmatrix}\xrightarrow{E_{2}^{(\frac{1}{11})}}\begin{pmatrix}1&4\\0&1\end{pmatrix}\xrightarrow{E_{12}^{(-4)}}\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

Since all matrices have the same rank, the original clearly has rank 2.

2. Since 2×2 matrices are small, the row operation approach wasn't required. For a larger matrix however, it can be invaluable. For instance:

$$A=\begin{pmatrix}1&1&0&3\\2&0&1&1\\0&1&2&0\\0&2&5&-1\\2&1&1&3\end{pmatrix}\xrightarrow{E_{21}^{(-2)}E_{51}^{(-2)}}\begin{pmatrix}1&1&0&3\\0&-2&1&-5\\0&1&2&0\\0&2&5&-1\\0&-1&1&-3\end{pmatrix}\xrightarrow{E_{13}^{(-1)}E_{23}^{(2)}E_{43}^{(-2)}E_{53}^{(1)}}\begin{pmatrix}1&0&-2&3\\0&0&5&-5\\0&1&2&0\\0&0&1&-1\\0&0&3&-3\end{pmatrix}$$

$$\xrightarrow{E_{14}^{(2)}E_{24}^{(-5)}E_{34}^{(-2)}E_{54}^{(-3)}}\begin{pmatrix}1&0&0&1\\0&0&0&0\\0&1&0&2\\0&0&1&-1\\0&0&0&0\end{pmatrix}\xrightarrow{E_{34}E_{23}}\begin{pmatrix}1&0&0&1\\0&1&0&2\\0&0&1&-1\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

Since the fourth column is a linear combination of the first three (linearly independent) columns, we conclude that $\operatorname{rank} A = 3$. Alternatively, we could repeat using with column operations:

$$\begin{pmatrix}1&1&0&3\\2&0&1&1\\0&1&2&0\\0&2&5&-1\\2&1&1&3\end{pmatrix}\longrightarrow\cdots\longrightarrow\begin{pmatrix}1&0&0&0\\0&0&1&0\\-4&5&-3&0\\-10&12&-7&0\\0&1&0&0\end{pmatrix}$$

The first three columns are linearly independent, so again the rank is three. If we used a mixture of row and column operations, we could eventually transform A into the rank 3 matrix

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

Row or Column Operations: which to use? If your only purpose is to compute ranks, mixing row and column operations is fine. If you want something else, you may wish to stick to only one type: here's why.

Row Operations preserve the span of the *rows* of a matrix (the *row space*). This is important when matrices represent linear systems of equations. For example, below we transform a system of equations and the corresponding augmented matrix using row operations:

$$\begin{cases} x+3y=1 \\ 2x-y=16 \end{cases} \qquad \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 16 \end{pmatrix} \qquad \left(\begin{array}{cc|c} 1 & 3 & 1 \\ 2 & -1 & 16 \end{array}\right)$$

$$\begin{cases} x+3y=1 \\ -7y=14 \end{cases} \qquad \begin{pmatrix} 1 & 3 \\ 0 & -7 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 14 \end{pmatrix} \qquad \left(\begin{array}{cc|c} 1 & 3 & 1 \\ 0 & -7 & 14 \end{array}\right)$$

$$\begin{cases} x+3y=1 \\ y=2 \end{cases} \qquad \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad \left(\begin{array}{cc|c} 1 & 3 & -1 \\ 0 & 1 & 2 \end{array}\right)$$

$$\begin{cases} x=-1 \\ y=2 \end{cases} \qquad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \qquad \left(\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 2 \end{array}\right)$$

This familiar method relies on the fact that row-equivalent linear systems have identical solutions. When viewed as a matrix system (middle column) it should be clear that multiplication by elementary matrices must occur *on the left.*

Column Operations preserve the column space of a matrix. For instance, the above example shows that a simple basis of the column space of A is given by

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ -4 \\ -10 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 5 \\ 12 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -3 \\ -7 \\ 0 \end{pmatrix} \right\}$$

Row operations will change the column space and vice versa. If knowing these is important to you, stick to one type of operation!

The example generalizes:

Theorem 3.9. *A matrix $A \in M_{m\times n}(\mathbb{F})$ has* rank $A = r$ *if and only if there exists a finite sequence of row and column operations transforming A to the matrix*

$$D = \begin{pmatrix} I_r & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{pmatrix}$$

Here I_r is the $r \times r$ identity, with the remaining pieces being zero matrices of the given dimensions.

Otherwise said, there exist elementary $m \times m$ matrices $R_1, \ldots, R_k$ and elementary $n \times n$ matrices $C_1, \ldots, C_\ell$ such that

$$R_k \cdots R_1 A C_1 \cdots C_\ell = D$$

The proof is too long to give in full, but can be proved by tedious induction on the number of rows of A. We are far more interested in some corollaries, particularly involving the maximal rank case, that is when A is invertible.

Computing the inverse of a matrix Everything follows from a simple corollary.

Corollary 3.10. *Every invertible matrix A is a product of elementary matrices.*

In light of the Corollary, the last line of Theorem 3.9 can be rewritten so say that

$$\operatorname{rank} A = r \iff \exists \text{invertible } P, Q \text{ such that } PAQ = D$$

Proof. If $A \in M_n(\mathbb{F})$ is invertible, then rank $A = n$ whence $D = I_n$. It follows that there exist products P, Q of elementary matrices such that

$$PAQ = I_n \implies A = P^{-1}Q^{-1}$$

By Theorem 3.3, the right hand side and thus A is a product of elementary matrices. ∎

The proof yields a systematic method for calculating inverses:

- The identity matrix $I_n = QPA$ is the result of applying a sequence of row operations to A.
- $A^{-1} = QP = QPI_n$ is the result of applying the *same sequence* to I_n.
- Since row operations never mix up columns, we can find A^{-1} by applying row operations to the *augmented matrix* $(A \mid I)$ until the left side is the identity: the right side will then be A^{-1}, i.e.

$$(A \mid I) \text{ is row equivalent to } (I \mid A^{-1})$$

Example 3.11. We compute the inverse of $A = \left(\begin{smallmatrix} 1&2&3\\0&2&0\\3&0&1\end{smallmatrix}\right)$ by applying row operations to $(A \mid I_3)$:

$$\left(\begin{array}{ccc|ccc} 1&2&3&1&0&0\\0&2&0&0&1&0\\3&0&1&0&0&1\end{array}\right) \xrightarrow{E_{12}^{(-1)}} \left(\begin{array}{ccc|ccc} 1&0&3&1&-1&0\\0&2&0&0&1&0\\3&0&1&0&0&1\end{array}\right) \xrightarrow{E_{2}^{(\frac{1}{2})}} \left(\begin{array}{ccc|ccc} 1&0&3&1&-1&0\\0&1&0&0&\frac{1}{2}&0\\3&0&1&0&0&1\end{array}\right)$$

$$\xrightarrow{E_{31}^{(-3)}} \left(\begin{array}{ccc|ccc} 1&0&3&1&-1&0\\0&1&0&0&\frac{1}{2}&0\\0&0&-8&-3&3&1\end{array}\right) \xrightarrow{E_{3}^{(-\frac{1}{8})}} \left(\begin{array}{ccc|ccc} 1&0&3&1&-1&0\\0&1&0&0&\frac{1}{2}&0\\0&0&1&\frac{3}{8}&-\frac{3}{8}&-\frac{1}{8}\end{array}\right)$$

$$\xrightarrow{E_{13}^{(-3)}} \left(\begin{array}{ccc|ccc} 1&0&0&-\frac{1}{8}&\frac{1}{8}&\frac{3}{8}\\0&1&0&0&\frac{1}{2}&0\\0&0&1&\frac{3}{8}&-\frac{3}{8}&-\frac{1}{8}\end{array}\right) \implies A^{-1} = \frac{1}{8}\begin{pmatrix} -1&1&3\\0&4&0\\3&-3&-1\end{pmatrix}$$

It can easily be checked by multiplication that we have found the correct inverse matrix: the above indeed shows how to write A as a product of elementary matrices

$$I_3 = E_{13}^{(-3)}E_3^{(-\frac{1}{8})}E_{31}^{(-3)}E_2^{(\frac{1}{2})}E_{12}^{(-1)}A$$

$$\implies A = E_{12}^{(1)}E_2^{(2)}E_{31}^{(3)}E_3^{(-8)}E_{13}^{(3)} = \begin{pmatrix}1&1&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&2&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\3&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&-8\end{pmatrix}\begin{pmatrix}1&0&3\\0&1&0\\0&0&1\end{pmatrix}$$

It is also acceptable, though non-standard, to perform column operations on the augmented matrix $\left(\frac{A}{I_n}\right)$: just remember never to mix the two types of operation when computing the inverse!

Corollary 3.12. *For any matrix A we have* $\operatorname{rank} A = \operatorname{rank} A^T$.

Proof. $D = PAQ \implies D^T = Q^T A^T P^T$. Since Q^T and P^T are invertible, we immediately see that $\operatorname{rank} A^T = \operatorname{rank} D^T$. But $\operatorname{rank} D^T$ clearly equals $r = \operatorname{rank} D = \operatorname{rank} A$. ∎

In particular, the dimension of the *row space* (span of the rows of A) is also $\operatorname{rank} A$.

Maximum and Minimum Ranks of Compositions

As a final application we consider the rank of a composition in terms of its factors.

Example 3.13. If $\operatorname{rank} A = 3$ and $\operatorname{rank} B = 2$, we can appeal to block matrices such as in Theorem 3.9 to consider possible ranks of the product AB:

- $A = \begin{pmatrix} 1&0&0\\0&1&0\\0&0&1 \end{pmatrix}$, $B = \begin{pmatrix} 1&0\\0&1\\0&0 \end{pmatrix}$: $AB = B \implies \operatorname{rank} AB = 2$
- $A = \begin{pmatrix} 1&0&0&0\\0&1&0&0\\0&0&1&0 \end{pmatrix}$, $B = \begin{pmatrix} 0&0&0\\0&0&0\\1&0&0\\0&1&0 \end{pmatrix}$: $AB = \begin{pmatrix} 0&0&0\\0&0&0\\1&0&0 \end{pmatrix} \implies \operatorname{rank} AB = 1$
- $A = \begin{pmatrix} 1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0 \end{pmatrix}$, $B = \begin{pmatrix} 0&0&0\\0&0&0\\0&0&0\\1&0&0\\0&1&0 \end{pmatrix}$: $AB = \begin{pmatrix} 0&0&0\\0&0&0\\0&0&0 \end{pmatrix} \implies \operatorname{rank} AB = 0$

As the next result shows, these are essentially all the possibilities.

Theorem 3.14. *Let* $\mathrm{S}: U \to V$ *and* $\mathrm{T}: V \to W$ *be linear maps, then*[a]

$$\operatorname{rank}\mathrm{S} - \operatorname{null}\mathrm{T} \le \operatorname{rank}\mathrm{TS} \le \min\big(\operatorname{rank}\mathrm{T}, \operatorname{rank}\mathrm{S}\big)$$

$$\max\big(\operatorname{null}\mathrm{S}, \dim U - \operatorname{rank}\mathrm{T}\big) \le \operatorname{null}\mathrm{TS} \le \operatorname{null}\mathrm{T} + \operatorname{null}\mathrm{S}$$

The same relationships hold for matrices A, B, provided the product is AB is defined.

[a]The upper bounds are easier to remember due to their symmetry: use the rank–nullity theorem to recover the lower bounds instead of memorizing them!

Proof. If $\mathbf{w} \in \mathcal{R}(\mathrm{TS})$, then $\mathbf{w} = \mathrm{T}(\mathrm{S}(\mathbf{u}))$ for some $\mathbf{u} \in U$, from which $\mathbf{w} \in \mathcal{R}(\mathrm{T})$. We conclude that

$$\mathcal{R}(\mathrm{TS}) \le \mathcal{R}(\mathrm{T}) \implies \operatorname{rank}\mathrm{TS} \le \operatorname{rank}\mathrm{T}$$

If this were a claim about matrices, Corollary 3.12 could deal with the other part of the minimum: $\operatorname{rank} AB = \operatorname{rank}(AB)^T = \cdots$. Instead we consider null spaces and apply the rank-nullity theorem:

$$\mathbf{u} \in \mathcal{N}(\mathrm{S}) \implies \mathrm{S}(\mathbf{u}) = \mathbf{0} \implies \mathrm{TS}(\mathbf{u}) = \mathbf{0} \implies \mathbf{u} \in \mathcal{N}(\mathrm{TS})$$

from which

$$\begin{aligned} \mathcal{N}(\mathrm{S}) \le \mathcal{N}(\mathrm{TS}) &\implies \operatorname{null}\mathrm{S} \le \operatorname{null}\mathrm{TS} \implies \dim U - \operatorname{rank}\mathrm{S} \le \dim U - \operatorname{rank}\mathrm{TS} \\ &\implies \operatorname{rank}\mathrm{TS} \le \operatorname{rank}\mathrm{S} \end{aligned}$$

The remaining inequalities are an exercise. ∎

Example 3.15. Let $A \in M_{6\times 5}(\mathbb{R})$ and $B \in M_{5\times 4}(\mathbb{R})$, and suppose that $\operatorname{rank} A = 4$ and $\operatorname{rank} B = 3$. We find the maximum and minimum possible ranks of the product AB and give examples in each case.

First observe that since $\mathrm{L}_A : \mathbb{R}^5 \to \mathbb{R}^6$, the rank–nullity theorem says that $\operatorname{null} A = 5 - \operatorname{rank} A = 1$. Similarly $\operatorname{null} B = 4 - 3 = 1$.

By the Theorem,

$$2 = \operatorname{rank} B - \operatorname{null} A \le \operatorname{rank} AB \le \min(\operatorname{rank} A, \operatorname{rank} B) = 3$$

It is easy to cook up explicit matrices satisfying $\operatorname{rank} AB = 3$ as in the previous example: for instance

$$A = \begin{pmatrix} I_4 & O_{4\times 1} \\ O_{2\times 4} & O_{2\times 1} \end{pmatrix} = \begin{pmatrix} 1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&0\\0&0&0&0&0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&0\\0&0&0&0 \end{pmatrix} \implies AB = \begin{pmatrix} 1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&0\\0&0&0&0\\0&0&0&0 \end{pmatrix}$$

The idea for maximum ranks is to have the identity submatrices inside A and B overlap as much as possible.

By trying to make the identities overlap as little as possible—essentially squeezing as much of the range of A into the nullspace of A—we should can also create a minimal rank example: for instance, with the same A as above,

$$B = \begin{pmatrix} 0&0&0&0\\0&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1 \end{pmatrix} \implies AB = \begin{pmatrix} 0&0&0&0\\0&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&0\\0&0&0&0 \end{pmatrix}$$

Exercises 3.2 1. For each of the following matrices, compute the rank and the inverse if it exists:

(a) $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & -2 & 4 \\ 1 & 1 & -1 \\ 2 & 4 & 5 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & -1 & 2 \\ 2 & 0 & 1 & 0 \\ 0 & -1 & 1 & -3 \end{pmatrix}$

2. For each of the following linear transforms T, find the matrix of the linear map with respect to the standard bases, determine whether T is invertible, and compute T^{-1}, if it exists.

   (a) $\mathrm{T} : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ defined by $\mathrm{T}(f)(x) = (x+1)f'(x)$

   (b) $\mathrm{T} : \mathbb{R}^3 \to P_2(\mathbb{R})$ defined by $\mathrm{T}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = (a+b+c) + (a-b+c)x + ax^2$

   (c) $\mathrm{T} : M_2(\mathbb{R}) \to \mathbb{R}^4$ defined by $\mathrm{T}(A) = \begin{pmatrix} \operatorname{tr} A \\ \operatorname{tr} A^T \\ \operatorname{tr}(EA) \\ \operatorname{tr}(AE) \end{pmatrix}$ where $E = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

3. (a) Find $A \in M_{3\times 4}(\mathbb{R})$ and $B \in M_{4\times 3}(\mathbb{R})$ such that $\operatorname{rank} A = \operatorname{rank} B = 2$ and $\operatorname{rank} AB = 1$.

   (b) Suppose that $A \in M_{4\times 3}(\mathbb{R})$ and $B \in M_{3\times 5}(\mathbb{R})$ have $\operatorname{rank} A = 2$ and $\operatorname{rank} B = 3$. What is $\operatorname{rank} AB$?

4. (a) Let $\mathrm{S} \in \mathcal{L}(U,V)$ and $\mathrm{T} \in \mathcal{L}(V,W)$. Prove that

   $$\operatorname{null} \mathrm{TS} \le \operatorname{null} \mathrm{T} + \operatorname{null} \mathrm{S}$$

   Now apply the Rank–Nullity Theorem to finish the proof of Theorem 3.14.

(b) Let $A = \begin{pmatrix} I_a & O \\ O & O \end{pmatrix}$, $B = \begin{pmatrix} I_b & O \\ O & O \end{pmatrix}$ and $C = \begin{pmatrix} O & O \\ O & I_c \end{pmatrix}$ where a, b, c are the ranks of A, B, C respectively, and O indicates the zero matrix of the appropriate size. Suppose that

$$A \in M_{m\times n}(\mathbb{F}), \qquad B, C \in M_{n\times p}(\mathbb{F})$$

Compute AB and AC and check that

$$\operatorname{rank} AB = \min(\operatorname{rank} A, \operatorname{rank} B) \quad \text{and} \quad \operatorname{rank} AC = \max(0, \operatorname{rank} C - \operatorname{null} A)$$

This shows that the maximal and minimal ranks indicated in the Theorem can actually be achieved; the only caveat being that $\operatorname{rank} C - \operatorname{null} A$ *could be negative!*

5. Prove that for any $m \times n$ matrix A, we have $\operatorname{rank} A = 0 \iff A$ is the zero matrix.

6. (a) Prove that matrices $A, B \in M_{m\times n}(\mathbb{F})$ have the same rank if and only if $B = PAQ$ for some invertible P, Q.

   (b) Suppose that $\mathrm{T} \in \mathcal{L}(V, W)$ is a linear map between finite-dimensional vector spaces. Show that $\operatorname{rank}[\mathrm{T}]_\beta^\gamma$ is independent of the choice of bases β and γ.

   (*Of course, this value equals* $\operatorname{rank} \mathrm{T}$ *itself!*)

7. Write the invertible matrix $A = \begin{pmatrix} 3 & 2 \\ 10 & 6 \end{pmatrix}$ as a product of elementary matrices.

   For a challenge, see if you can do this for a general invertible matrix.

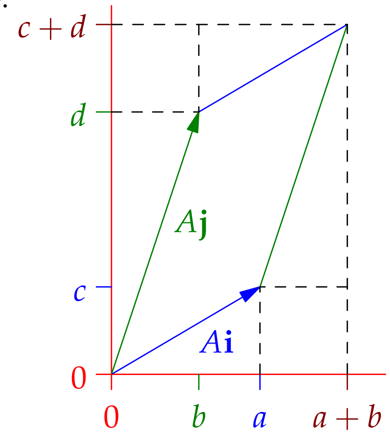
8. Let $\mathrm{T} \in \mathcal{L}(V, W)$ be given and suppose that $\mathrm{P} \in \mathcal{L}(W)$ and $\mathrm{Q} \in \mathcal{L}(V)$ are isomorphisms. Prove that

$$\operatorname{rank} \mathrm{PT} = \operatorname{rank} \mathrm{TQ} = \operatorname{rank} \mathrm{T}$$

   (*Your argument must work in infinite dimensions and thus without matrices*)

4 Determinants

To a square matrix A, it turns out that we can attach a single number, its *determinant*, which encapsulates the extent to which the linear map L_A enlarges or contracts space. For instance, consider a 2×2 matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Its columns are the result of multiplying the standard basis vectors $\mathbf{i}, \mathbf{j}$ by A:



$$A\mathbf{i} = \begin{pmatrix} a \\ c \end{pmatrix} \qquad A\mathbf{j} = \begin{pmatrix} b \\ d \end{pmatrix}$$

For simplicity, suppose $a, b, c, d > 0$ and that the columns are oriented as in the picture. The *unit square* spanned by $\mathbf{i}, \mathbf{j}$ is transformed by A to a *parallelogram*, whose area is

$$(a+b)(c+d) - 2bc - 2 \cdot \frac{1}{2}bd - 2 \cdot \frac{1}{2}ac = ad - bc$$

This one number neatly summarizes how left-multiplication by A changes the *area* of a shape.

4.1 Determinants of Order 2

Definition 4.1. The *determinant* $\det A = |A|$ of a 2×2 matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is the scalar

$$\det A = ad - bc$$

Example 4.2. If $A = \left(\begin{smallmatrix} 1 & 2 \\ 4 & 3 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 5 & 0 \\ 1 & -2 \end{smallmatrix}\right)$, then

$$\det A = 1 \cdot 3 - 2 \cdot 4 = -5, \qquad \det B = 5 \cdot (-2) - 0 \cdot 1 = -10$$

Note that $\det : M_2(\mathbb{F}) \to \mathbb{F}$ is a *non-linear* function; for instance

$$\det(A+B) = \begin{vmatrix} 6 & 2 \\ 5 & 1 \end{vmatrix} = 6 - 10 = -4 \neq \det A + \det B$$

However, determinant does play nicely with matrix multiplication:

$$\det AB = \begin{vmatrix} 7 & -4 \\ 23 & -6 \end{vmatrix} = -42 + 92 = 50 = \det A \det B$$

In the following results, we summarize the key properties of order 2 determinants: with the exception of the explicit inverse formula, these will eventually be seen to hold in higher dimensions.

Theorem 4.3 (Basic properties of order-two determinants). 1. $\det A^T = \det A$

2. $\det A = 0$ *if and only if the columns (rows) of A are parallel (linearly dependent).*
3. *Determinant is a bilinear function of the columns (rows) of A.*
4. $\det AB = \det A \det B$

These are easily verified directly: write $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, $B = \left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right)$, etc. The third property benefits from a little expansion: writing a matrix in terms of its columns, determinant can be thought of as a function

$$\det : \mathbb{F}^2 \times \mathbb{F}^2 \to \mathbb{F} : (\mathbf{a}_1\ \mathbf{a}_2) \mapsto \det(\mathbf{a}_1\ \mathbf{a}_2)$$

and property 3 claims that

$$\det(\lambda\mathbf{u} + \mathbf{v}, \mathbf{w}) = \lambda \det(\mathbf{u}, \mathbf{w}) + \det(\mathbf{v}, \mathbf{w}) \quad \text{and} \quad \det(\mathbf{u}, \lambda\mathbf{v} + \mathbf{w}) = \lambda \det(\mathbf{u}, \mathbf{v}) + \det(\mathbf{u}, \mathbf{w})$$

and similarly with regard to the rows of a matrix (a pair of row vectors).

Example 4.4. The fact that the first rows are identical means we can combine

$$-24 = -5 - 19 = \det\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} + \det\begin{pmatrix} 1 & 2 \\ 5 & -9 \end{pmatrix} = \det\begin{pmatrix} 1 & 2 \\ 9 & -6 \end{pmatrix} = -6 - 18 = -24$$

Properties 2 and 3 partly overlap with the effect of row/column operations on determinant.

Corollary 4.5 (Row/column operations).

Type I: *Swapping rows (or columns) changes the sign of* $\det A$

Type II: *Multiplying a row (or column) by* λ *multiplies* $\det A$ *by* λ

Type III: *Adding a multiple of one row (or column) to another leaves* $\det A$ *unchanged*

Proof. Simply combine the product formula with the list of all elementary matrices

$$\det EA = \det AE = \det E \det A$$

Type	Matrices	Determinant
I	$\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$	$\det E = -1$
II	$\left(\begin{smallmatrix} \lambda & 0 \\ 0 & 1 \end{smallmatrix}\right)$ $\left(\begin{smallmatrix} 1 & 0 \\ 0 & \lambda \end{smallmatrix}\right)$	$\det E = \lambda$
III	$\left(\begin{smallmatrix} 1 & \lambda \\ 0 & 1 \end{smallmatrix}\right)$ $\left(\begin{smallmatrix} 1 & 0 \\ \lambda & 1 \end{smallmatrix}\right)$	$\det E = 1$

Alternatively you can compute all possibilities directly. ∎

Corollary 4.6 (Inverses). *A is invertible (non-singular) if and only if $\det A \neq 0$. In such a case, $\det A^{-1} = \frac{1}{\det A}$ and*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Proof. (⇒) If A is invertible, the product formula tells us that

$$1 = \det I_2 = \det(AA^{-1}) = \det A \det A^{-1} \implies \det A \neq 0 \text{ and } \det A^{-1} = \frac{1}{\det A}$$

(⇐) Observe that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = (\det A) I_2$$

If $\det A \neq 0$, we divide by $\det A$ to see that A has an inverse given by the desired expression. ∎

Oriented Area and the Determinant

In the introduction, we considered a basic example of how determinant relates to area. We now proceed more formally.

Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = (\mathbf{u}\ \mathbf{v})$ be written in terms of its columns: we consider the parallelogram $\mathcal{P}$ obtained by applying A to the unit square, i.e. $\mathcal{P}$ is spanned by the vectors $\mathbf{u} = A\mathbf{i}$ and $\mathbf{v} = A\mathbf{j}$.

First observe (Theorem 4.3, part 2) that

$$\det A = \det(\mathbf{u}, \mathbf{v}) = 0 \iff \mathbf{u}, \mathbf{v} \text{ are parallel} \iff \mathcal{P} \text{ has zero area}$$

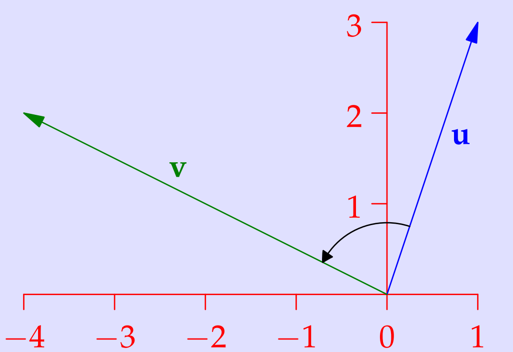
This gives a pictorial way to understand zero determinant and lack of invertibility: if $\mathbf{u}, \mathbf{v}$ are parallel then any purported inverse would map these to *parallel* vectors $A^{-1}\mathbf{u}, A^{-1}\mathbf{v}$ which couldn't span the original unit square! A linear map can *scale* area, but it cannot create area out of nothing.

Now consider the case when $\det A \neq 0$. In particular, this requires at least one of $a, c \neq 0$. It is straightforward to check that the matrix $R = \frac{1}{\sqrt{a^2+c^2}} \left(\begin{smallmatrix} a & c \\ -c & a \end{smallmatrix}\right)$ acts by *counter-clockwise rotation* and therefore *preserves area*. Now compute,

$$RA = \frac{1}{\sqrt{a^2+c^2}} \begin{pmatrix} a & c \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{\sqrt{a^2+c^2}} \begin{pmatrix} a^2+c^2 & ab+cd \\ 0 & \det A \end{pmatrix} \tag{$*$}$$

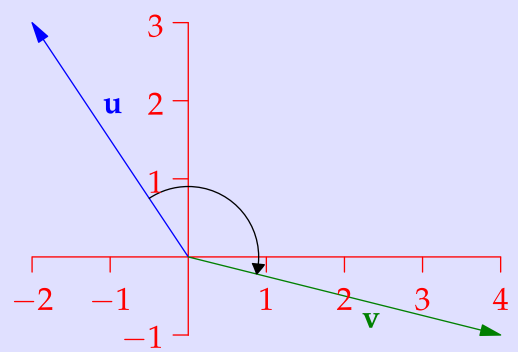
Since $a^2 + c^2 > 0$, we see that the sign of $\det A$ determines whether we rotate counter-clockwise ($\det A > 0$) or clockwise ($\det A < 0$) to get from the first to the second column of A.

Definition 4.7. An ordered pair of vectors $(\mathbf{u}, \mathbf{v})$ in $\mathbb{R}^2$ is *positively-oriented* if $\det(\mathbf{u}, \mathbf{v}) > 0$ and *negatively-oriented* if $\det(\mathbf{u}, \mathbf{v}) < 0$.



Positively-oriented: counter-clockwise rotation

$$\det(\mathbf{u}, \mathbf{v}) = \begin{vmatrix} 1 & -4 \\ 3 & 2 \end{vmatrix} = 14 > 0$$



Negatively-oriented: clockwise rotation

$$\det(\mathbf{u}, \mathbf{v}) = \begin{vmatrix} -2 & 4 \\ 3 & -1 \end{vmatrix} = -10 < 0$$

A short exercise following from $(*)$ results in a complete proof of the following:

Theorem 4.8. *The area of the parallelogram spanned by the columns of A is*

- $\det A$ *if the columns are positively-oriented;*
- $-\det A$ *if the columns are negatively-oriented.*

As such, $\det A$ is often known as the *oriented area* of a parallelogram.

Exercises 4.1 1. Compute the determinants of the following matrices and, if possible, their inverses:

(a) $\begin{pmatrix} 1 & 4 \\ 7 & -3 \end{pmatrix} \in M_2(\mathbb{R})$ (b) $\begin{pmatrix} i & 1+i \\ 1 & i \end{pmatrix} \in M_2(\mathbb{C})$ (c) $\begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix} \in M_2(\mathbb{Z}_5)$ (d) $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in M_2(\mathbb{Z}_5)$
(*Recall that* $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$ *is the field of remainders modulo* 5)

2. Explicitly prove all parts of Theorem 4.3.

3. Find the area of the triangles (*a triangle is half a parallelogram...*):

   (a) With vertices $(0, 0)$, $(2, 4)$ and $(1, -2)$.
   (b) With vertices $(2, 1)$, $(3, -2)$ and $(-5, 4)$.

4. (a) Show that the matrix $R = \frac{1}{\sqrt{a^2+c^2}} \begin{pmatrix} a & c \\ -c & a \end{pmatrix}$ acts by counter-clockwise rotation.
(*Hint: the columns of R are related by* $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ -c \end{pmatrix} = \begin{pmatrix} c \\ a \end{pmatrix}$)
   (b) Complete the proof of Theorem 4.8.

5. Let $\mathbb{F}$ be a field and $\Delta : \mathbb{F}^2 \times \mathbb{F}^2 \to \mathbb{F}$ be a *bilinear function*: $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{F}^2, \lambda \in \mathbb{F}$,

$$\Delta(\lambda \mathbf{u} + \mathbf{v}, \mathbf{w}) = \lambda \Delta(\mathbf{u}, \mathbf{w}) + \Delta(\mathbf{v}, \mathbf{w}) \quad \text{and} \quad \Delta(\mathbf{u}, \lambda \mathbf{v} + \mathbf{w}) = \lambda \Delta(\mathbf{u}, \mathbf{v}) + \Delta(\mathbf{u}, \mathbf{w})$$

   (a) We say that Δ is *alternating* if $\Delta(\mathbf{u}, \mathbf{u}) = 0$ for all $\mathbf{u} \in \mathbb{F}^2$.
      i. Prove that Δ alternating $\implies \Delta(\mathbf{v}, \mathbf{u}) = -\Delta(\mathbf{u}, \mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$.
      ii. Prove the converse to part (a) (*provided* $2 \neq 0$ in $\mathbb{F}$!).
   (b) Prove that if Δ is an alternating bilinear function satisfying $\Delta(\mathbf{i}, \mathbf{j}) = 1$, then $\Delta = \det$.

6. (*A link to multivariable calculus*) Let $D, E \subseteq \mathbb{R}^2$ and $T : D \to E$ be a change of co-ordinates

$$T(u, v) = \big(x(u, v), y(u, v)\big)$$

Assume T is bijective and that the partial derivatives of x, y with respect to u, v exist and are continuous. Note that T does not have to be linear!
Let $P = (u_0, v_0) \in D$, and consider small positive quantities $\Delta u, \Delta v$ to define points

$$Q = (u_0 + \Delta u, v_0), \qquad R = (u_0, v_0 + \Delta v)$$

The area of the parallelogram spanned by $\overrightarrow{PQ} = \mathbf{i}\Delta u$ and $\overrightarrow{PR} = \mathbf{j}\Delta v$ is therefore $\Delta u \Delta v$.

Prove that the parallelogram spanned by $\overrightarrow{T(P)T(Q)}$ and $\overrightarrow{T(P)T(R)}$ has

$$\text{Area} \approx \left| \det \begin{pmatrix} x_u(P) & x_v(P) \\ y_u(P) & y_v(P) \end{pmatrix} \right| \Delta u \Delta v$$

where $x_u = \frac{\partial x}{\partial u}$, etc., denote partial derivatives.

This determinant is the Jacobian $J(T) = \frac{\partial(x,y)}{\partial(u,v)}$. *The above is essentially the justification for the change of variables formula for double integrals:*

$$\iint_E f(x, y)\, \mathrm{d}x\mathrm{d}y = \iint_D f\big(x(u, v), y(u, v)\big) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| \mathrm{d}u\mathrm{d}v$$

4.2 Higher-Order Determinants

We now extend the definition of determinant to any size of square matrix. The goal is to establish all the basic properties seen for order 2 determinants. This will take a little time…

Definition 4.9. Let $A \in M_n(\mathbb{F})$. For each i, j we defined the ij^{th} *minor* of A to be the matrix $\tilde{A}_{ij}$ obtained by deleting the i^{th} row and j^{th} column of A. The *determinant* of A to be the sum

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det \tilde{A}_{1j} = a_{11} \det \tilde{A}_{11} - a_{12} \det \tilde{A}_{12} + \cdots + (-1)^{1+n} a_{1n} \det \tilde{A}_{1n}$$

This is known as the *cofactor expansion* of $\det A$ along the first row of A.

The difficulty should be immediately obvious: the definition is *inductive*! For instance the determinant of $A \in M_4(\mathbb{F})$ is defined in terms of the *four* 3×3 determinants, each of which is computed using *three* 2×2 determinants: in total we need *twelve* order 2 determinants!

Example 4.10. Let $A = \begin{pmatrix} 1 & 0 & 2 & -3 \\ 0 & 1 & 2 & -1 \\ -2 & 3 & 2 & 6 \\ 1 & 2 & 1 & 1 \end{pmatrix}$, then

$$\begin{aligned}
\det A &= 1 \begin{vmatrix} 1 & 2 & -1 \\ 3 & 2 & 6 \\ 2 & 1 & 1 \end{vmatrix} - 0 \begin{vmatrix} 0 & 2 & -1 \\ -2 & 2 & 6 \\ 1 & 1 & 1 \end{vmatrix} + 2 \begin{vmatrix} 0 & 1 & -1 \\ -2 & 3 & 6 \\ 1 & 2 & 1 \end{vmatrix} - (-3) \begin{vmatrix} 0 & 1 & 2 \\ -2 & 3 & 2 \\ 1 & 2 & 1 \end{vmatrix} \\
&= \begin{vmatrix} 1 & 2 & -1 \\ 3 & 2 & 6 \\ 2 & 1 & 1 \end{vmatrix} + 2 \begin{vmatrix} 0 & 1 & -1 \\ -2 & 3 & 6 \\ 1 & 2 & 1 \end{vmatrix} + 3 \begin{vmatrix} 0 & 1 & 2 \\ -2 & 3 & 2 \\ 1 & 2 & 1 \end{vmatrix}
\end{aligned}$$

Next we compute the remaining 3×3 determinants using the cofactor expansion:

$$\begin{aligned}
\det A &= \left(\begin{vmatrix} 2 & 6 \\ 1 & 1 \end{vmatrix} - 2 \begin{vmatrix} 3 & 6 \\ 2 & 1 \end{vmatrix} + (-1) \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} \right) \\
&\qquad + 2 \left(0 \begin{vmatrix} 3 & 6 \\ 2 & 1 \end{vmatrix} - \begin{vmatrix} -2 & 6 \\ 1 & 1 \end{vmatrix} + (-1) \begin{vmatrix} -2 & 3 \\ 1 & 2 \end{vmatrix} \right) \\
&\qquad\quad + 3 \left(0 \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} - \begin{vmatrix} -2 & 2 \\ 1 & 1 \end{vmatrix} + 2 \begin{vmatrix} -2 & 3 \\ 1 & 2 \end{vmatrix} \right) \\
&= \begin{vmatrix} 2 & 6 \\ 1 & 1 \end{vmatrix} - 2 \begin{vmatrix} 3 & 6 \\ 2 & 1 \end{vmatrix} - \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} - 2 \begin{vmatrix} -2 & 6 \\ 1 & 1 \end{vmatrix} - 2 \begin{vmatrix} -2 & 3 \\ 1 & 2 \end{vmatrix} - 3 \begin{vmatrix} -2 & 2 \\ 1 & 1 \end{vmatrix} + 6 \begin{vmatrix} -2 & 3 \\ 1 & 2 \end{vmatrix}
\end{aligned}$$

Each of the remaining 2×2 determinants can now be evaluated:

$$\det A = -4 - 2(-9) - (-1) - 2(-8) - 2(-7) - 3(-4) + 6(-7) = 15$$

While the above calculation was assisted by the fact that we only needed to compute three 3×3 determinants, it is still very slow-going. As the order n gets larger, things becomes ugly very quickly. In order to facilitate more rapid calculations it is useful to develop some of the properties we saw in the previous section. We begin by computing the determinant of the $n \times n$ identity matrix I_n: The cofactor expansion along the first row yields

$$\det I_n = 1 \cdot \det I_{n-1} - 0 \det (\tilde{I_n})_{12} + 0 \det (\tilde{I_n})_{13} - \cdots + (-1)^{1+n} \cdot 0 \det (\tilde{I_n})_{1n} = \det I_{n-1}$$

By induction, since $\det I_2 = 1$, we conclude:

Lemma 4.11. $\det I_n = 1$.

The next result combines several basic properties, all of which we have checked for order 2 determinants. Several of the arguments rely on lengthy inductions.

Theorem 4.12.

1. *$\det : M_n(\mathbb{F}) \to \mathbb{F}$ is a linear function of each row when the other rows are fixed.*
2. *If A has a row of zeros, then $\det A = 0$.*
3. *The determinant can be evaluated using the cofactor expansion along* any *row:*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det \tilde{A}_{ij}$$

4. *If A has two identical rows then $\det A = 0$.*

Sketch Proof.

1. This is by induction. We know the statement is true for all 2×2 matrices. Fix n and suppose the claim is true for all $n \times n$ matrices. Let A be order $n+1$, with i^{th} row a linear combination

$$\mathbf{a}_i^T = \lambda \mathbf{b}_i^T + \mathbf{c}_i^T \qquad (*)$$

Also let B, C be the matrices obtained by replacing row i of A with $\mathbf{b}^T$ and $\mathbf{c}^T$ respectively. There are two cases.

(a) If $i = 1$, then the cofactor expansion along the first row is plainly linear since B, C have the same minors as A:

$$a_{1j} = \lambda b_j + c_j \implies \det A = \sum_{j=1}^{n+1} (-1)^{1+j} (\lambda b_j + c_j) \left| \tilde{A}_{1j} \right| = \lambda \det B + \det C$$

(b) If $i > 1$, then each of the $n \times n$ minors $\tilde{A}_{1j}$ has its $(i-1)^{\text{th}}$ row observing the same linear combination $(*)$. By the induction hypothesis,

$$\det A = \sum_{j=1}^{n+1} (-1)^{1+j} a_{1j} \left| \tilde{A}_{1j} \right| = \sum_{j=1}^{n+1} (-1)^{1+j} a_{1j} (\lambda \left| \tilde{B}_{1j} \right| + \left| \tilde{C}_{1j} \right|) = \lambda \det B + \det C$$

2. Re-run the proof of part 1 without C and taking $\lambda = 0$: i.e. suppose $(*)$ is simply $\mathbf{a}^T = 0\mathbf{b}^T$ where $\mathbf{b}^T$ is any row vector.
3. We omit the argument since it requires a *long* induction based on parts 1 and 2.
4. This is Exercise 4.2.5.

■

Examples 4.13. 1. Look for rows with many zeros when computing the determinant! Compare, for instance, the cofactor expansions of the following along the first and second rows:

First row: $\det\begin{pmatrix}5&3&2\\0&0&2\\7&2&1\end{pmatrix} = 5\begin{vmatrix}0&2\\2&1\end{vmatrix} - 3\begin{vmatrix}0&2\\7&1\end{vmatrix} + 2\begin{vmatrix}0&0\\7&2\end{vmatrix} = 5\cdot(-4) - 3\cdot(-14) + 2\cdot 0 = 22$

Second row: $\det\begin{pmatrix}5&3&2\\0&0&2\\7&2&1\end{pmatrix} = -0 + 0 - 2\begin{vmatrix}5&3\\7&2\end{vmatrix} = 22$

2. Here we take advantage of linearity and the fact that the first two rows are nearly identical before expanding along the second row of the resulting 3×3 determinant:

$$\det\begin{pmatrix}1&2&3&4\\1&2&3&0\\1&2&0&-1\\1&0&2&5\end{pmatrix} = \begin{vmatrix}1&2&3&4\\1&2&3&4\\1&2&0&-1\\1&0&2&5\end{vmatrix} - \begin{vmatrix}1&2&3&4\\0&0&0&4\\1&2&0&-1\\1&0&2&5\end{vmatrix} = 0 - 4\begin{vmatrix}1&2&3\\1&2&0\\1&0&2\end{vmatrix} = -4\left(-\begin{vmatrix}2&3\\0&2\end{vmatrix} + 2\begin{vmatrix}1&3\\1&2\end{vmatrix}\right) = 24$$

Elementary matrices and the determinant

We now consider the effect of elementary row operations on the determinant.

Corollary 4.14. *Let A be an $n\times n$ matrix and let E an elementary matrix. For each type:*

Type I: $\det EA = -\det A$, *and* $\det E = -1$;

Type II: If E multiplies a row by λ, then $\det EA = \lambda \det A$, *and* $\det E = \lambda$;

Type III: $\det EA = \det A$ *and* $\det E = 1$.

Warning! Multiplying every row by λ yields $\det(\lambda A) = \lambda^n \det A$

It is worth taking stock for a moment:

- For order two determinants (Corollary 4.5) these results followed from the multiplicative property $\det AB = \det A \det B$.
- For higher order, we need to prove directly (using Theorem 4.12). Indeed the Corollary establishes the limited multiplicative property $\det EA = \det E \det A$ whenever E is elementary. We will use this fact in the next section to prove the full multiplication formula.

Proof. 1. See Exercise 4.2.5

2. Suppose $E = E_i^{(\lambda)}$ multiplies row i by λ. Since det is linear in each row, we immediately see that $\det EA = \lambda \det A$.

3. Suppose $E_{ik}^{(\lambda)}$ adds λ times row k to row i. Since determinant is linear in row i, we see that

$$\det EA = \det A + \lambda \det B$$

where B is the matrix obtained from A by replacing row i with row k. Since B has two identical rows, we conclude that $\det EA = \det A$.

Letting $A = I_n$ in each case gives the advertised values for $\det E$. ∎

Examples 4.15. 1. In case the argument for part 3 of the theorem is unclear, here is an example: if $E = E_{21}^{(5)} = \begin{pmatrix} 1 & 0 \\ 5 & 1 \end{pmatrix}$ and $A = \begin{pmatrix} 4 & 2 \\ 3 & -1 \end{pmatrix}$, then

$$B = \begin{pmatrix} 4 & 2 \\ 4 & 2 \end{pmatrix} \text{ and } \det EA = \det \begin{pmatrix} 4 & 2 \\ 3+20 & -1+10 \end{pmatrix} = \det \begin{pmatrix} 4 & 2 \\ 3 & -1 \end{pmatrix} + 5 \det \begin{pmatrix} 4 & 2 \\ 4 & 2 \end{pmatrix}$$

2. We can use all the above results to assist us in computing determinants. For instance,

$$\det \begin{pmatrix} 1 & 1 & 2 & -3 \\ 0 & 3 & 1 & 0 \\ 0 & 1 & 2 & -1 \\ 2 & 7 & 4 & -6 \end{pmatrix} = 0 + 3 \begin{vmatrix} 1 & 2 & -3 \\ 0 & 2 & -1 \\ 2 & 4 & -6 \end{vmatrix} - \begin{vmatrix} 1 & 1 & -3 \\ 0 & 1 & -1 \\ 2 & 7 & -6 \end{vmatrix} + 0 \quad \text{(cofactor expansion 2}^{\text{nd}}\text{ row)}$$

$$= 3 \begin{vmatrix} 1 & 2 & -3 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{vmatrix} - \begin{vmatrix} 1 & 1 & -3 \\ 0 & 1 & -1 \\ 0 & 5 & 0 \end{vmatrix} \quad \text{(type III row operations simplify third rows)}$$

$$= 0 - \left(0 - 5 \begin{vmatrix} 1 & -3 \\ 0 & -1 \end{vmatrix} + 0 \right) \quad \text{(equal rows/cofactor expansion along third row)}$$

$$= 5(-1-0) = -5$$

We are now in a position to establish the relationships between determinant, products and inverses.

Theorem 4.16. *Let $A, B \in M_n(\mathbb{F})$. Then:*

1. $\det AB = \det A \det B$
2. *A is invertible* $\iff$ $\det A \neq 0$, *in which case* $\det A^{-1} = \frac{1}{\det A}$

Proof. We first establish both results when A is non-invertible (singular). Since the row space has dimension $\operatorname{rank} A < n$, at least one row (row i say) is a linear combination of the others:

$$\mathbf{a}_i^T = \sum_{k \neq i} c_k \mathbf{a}_k^T$$

Applying row operations of type III, namely multiplication by $E_{ik}^{(-c_k)}$, we see that

$$\det A = \det \left(\prod_{k \neq i} E_{ik}^{(-c_k)} A \right) = 0$$

since the right hand matrix has row i identically zero. Moreover, $\operatorname{rank} AB \leq \operatorname{rank} A < n$ so that AB is also singular. We conclude that $\det AB = 0 = \det A \det B$.

Now suppose A is invertible. Then $A = E_k \cdots E_1$ is a product of elementary matrices. The the multiplicative property for elementary matrices (Corollary 4.14) now proves the general result!

$$\det AB = \det E_k \cdots \det E_1 \det B = \det(E_k \cdots E_1) \det B = \det A \det B$$

Let $B = I$ to see that $\det A = \det E_k \cdots \det E_1 \neq 0$. Finally, let $B = A^{-1}$ to see that

$$1 = \det I_n = \det AA^{-1} = \det A \det A^{-1} \implies \det A^{-1} = \frac{1}{\det A}$$

∎

Only one basic property of determinant now remains.

Theorem 4.17. $\det A^T = \det A$.

Proof. First observe that $\det A = 0 \iff \det A^T = 0$, since $\operatorname{rank} A = \operatorname{rank} A^T$.

Suppose now that A is invertible. Then $A = E_k \cdots E_1$ is the product of elementary matrices. But then

$$A^T = E_1^T \cdots E_k^T$$

is also the product of elementary matrices. Moreover, E_i^T is an elementary matrix of the same type *and determinant* as E_i. It follows that

$$\det A^T = \det E_1^T \cdots \det E_k^T = \det E_1 \cdots \det E_k = \det A$$

■

Any statements regarding rows or row operations now apply equally to columns and vice versa. In particular, we may compute $\det A$ using the cofactor expansion along any row or *down any column.*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det \tilde{A}_{ij} \qquad \text{(expansion along } i^{\text{th}} \text{ row)}$$

$$= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det \tilde{A}_{ij} \qquad \text{(expansion along } j^{\text{th}} \text{ column)}$$

Example 4.18. It is sensible to evaluate the determinant of the following matrix using the cofactor expansion down the 3rd column:

$$\det \begin{pmatrix} 1 & 2 & 0 & 4 \\ -2 & 3 & 1 & 5 \\ 2 & 4 & 2 & 8 \\ -1 & -3 & 0 & 2 \end{pmatrix} = -\begin{vmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ -1 & -3 & 2 \end{vmatrix} + 2\begin{vmatrix} 1 & 2 & 4 \\ -2 & 3 & 5 \\ -1 & -3 & 2 \end{vmatrix} = 0 + 2\begin{vmatrix} 1 & 2 & 4 \\ 0 & 7 & 13 \\ 0 & -1 & 6 \end{vmatrix} \qquad \text{(rank} = 0 \text{ and row operations)}$$

$$= 2\begin{vmatrix} 7 & 13 \\ -1 & 6 \end{vmatrix} = 110 \qquad \text{(cofactor first column)}$$

Order 3 Determinants and Volume

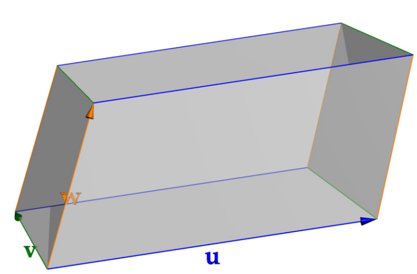
As in two dimensions, we have a geometric interpretation.

Theorem 4.19. *Let* $A = (\mathbf{u}, \mathbf{v}, \mathbf{w}) \in M_3(\mathbb{R})$ *have columns* $\mathbf{u}, \mathbf{v}, \mathbf{w}$.

1. $\det A = [\mathbf{u}, \mathbf{v}, \mathbf{w}] = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ *is the* scalar triple product *of its columns*
2. *The volume of the parallelepiped spanned by* $\mathbf{u}, \mathbf{v}, \mathbf{w}$ *is given by* $|\det A|$.

The ordered triple $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ is said to be *positively-oriented* if $\det A > 0$. By the theorem, this is if and only if $\mathbf{u}$ lies on the same side of the the $\mathbf{v}, \mathbf{w}$-plane as the normal vector $\mathbf{v} \times \mathbf{w}$: this is the familiar *right-hand rule.* The picture represents a positively oriented triple.

One can also extend this interpretation of oriented volume to higher dimensions.



Example 4.20. The parallelepiped spanned by the vectors $\left\{ \left(\begin{smallmatrix}1\\1\\2\end{smallmatrix}\right), \left(\begin{smallmatrix}0\\2\\6\end{smallmatrix}\right), \left(\begin{smallmatrix}-1\\1\\-2\end{smallmatrix}\right) \right\}$ has volume

$$\left|\det \left(\begin{smallmatrix}1&0&-1\\1&2&1\\2&6&-2\end{smallmatrix}\right)\right| = |-12| = 12$$

Since the determinant is negative, the three vectors are *negatively-oriented.*

Cramer's Rule

Finally, we present an application to the solution of $n \times n$ systems of linear equations $A\mathbf{x} = \mathbf{b}$:

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \qquad\vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n = b_n \end{cases} \qquad \text{where } A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \ \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Theorem 4.21. *Suppose $A\mathbf{x} = \mathbf{b}$ where A is invertible. Then the unique solution $\mathbf{x}$ has k^{th} entry*

$$x_k = \frac{1}{\det A} \det M_k$$

where M_k is the matrix obtained by replacing column k of A by $\mathbf{b}$.

Proof. For each k, define $X_k = A^{-1}M_k$. Since the columns of A and M_k, except the k^{th}, are identical

$$X_k\mathbf{e}_i = A^{-1}M_k\mathbf{e}_i = \begin{cases} A^{-1}\mathbf{b} = \mathbf{x} & \text{if } i = k \\ A^{-1}A\mathbf{e}_i = \mathbf{e}_i & \text{if } i \neq k \end{cases}$$

Therefore X_k is the identity matrix except that the k^{th} column is the solution $\mathbf{x}$. Use the k^{th} row cofactor expansion of X_k to see that

$$x_k = (-1)^{k+k} x_k \det I_{n-1} = \det X_k = \det A^{-1} \det M_k$$

∎

Cramer's rule is particularly useful *if* you only want to find a small number of the solution values $x_1, \ldots, x_n$. If you want to find *all* the values, you are usually better off solving the system using an augmented matrix approach, or even computing the inverse A^{-1}.

Example 4.22. We find y by applying Cramer's rule to the system $\begin{cases} 2x + 3y + 7z &= 1 \\ x - y + 3z &= 2 \\ 3x + 5y + z &= 8 \end{cases}$

$$y = \frac{\left|\begin{smallmatrix}2&1&7\\1&2&3\\3&8&1\end{smallmatrix}\right|}{\left|\begin{smallmatrix}2&3&7\\1&-1&3\\3&5&1\end{smallmatrix}\right|} = \frac{2\begin{vmatrix}2&3\\8&1\end{vmatrix} - \begin{vmatrix}1&3\\3&1\end{vmatrix} + 7\begin{vmatrix}1&2\\3&8\end{vmatrix}}{2\begin{vmatrix}-1&3\\5&1\end{vmatrix} - 3\begin{vmatrix}1&3\\3&1\end{vmatrix} + 7\begin{vmatrix}1&-1\\3&5\end{vmatrix}} = \frac{-44+8+14}{-32+24+56} = \frac{-22}{48} = -\frac{11}{24}$$

If you want to find x and z using this method, you need to do a lot of calculating!

Exercises 4.2 1. Compute the determinants of the following matrices using any method you like:

$$A = \begin{pmatrix} -1 & 2+i & 3 \\ 1-i & i & 1 \\ 3i & 2 & -1+i \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -2 & 3 & -12 \\ -5 & 12 & -14 & 19 \\ -9 & 22 & -20 & 31 \\ -4 & 9 & -14 & 15 \end{pmatrix}$$

2. Prove that the determinant of an upper triangular matrix is the product of the terms on its diagonal.

3. We saw that a general 4×4 determinant requires the computation of $4 \cdot 3 = 12$ determinants of order 2. How many order 2 determinants does an order n determinant require?

4. Establish the 'diagonal' method for computing a 3×3 determinant: as the sum of the products of the falling diagonals minus the products of the rising diagonals:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$\det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{31}a_{22}$$

Now use this to quickly compute the determinant of $A = \left(\begin{smallmatrix} 1 & 3 & 0 \\ 2 & -1 & 1 \\ 1 & 0 & 2 \end{smallmatrix}\right)$.

(Warning! *This method is special to order 3 determinants: if $n \geq 4$ you have to calculate the slow way!*)

5. (a) Prove part 4 of Theorem 4.12: a matrix with two identical rows has determinant zero.
   (b) Prove the type I case of Corollary 4.14: switching two rows changes the sign of the determinant.

   (*Hint: You can prove these in either order. You'll need an induction for whichever you do first, then the other should follow quite easily...*)

6. Find the determinant of $A = \begin{pmatrix} 1 & 4 & 1 & 1 \\ 2 & 3 & 0 & 1 \\ 2 & 2 & 0 & 1 \\ 2 & 1 & -1 & 1 \end{pmatrix}$

7. Use Cramer's rule to solve the system of equations $\begin{cases} 2x + y - 3z = 1 \\ x - 2y + z = 0 \\ 3x + 4y - 2z = -5 \end{cases}$

8. Suppose that n is odd and that A is skew-symmetric ($A^T = -A$). Prove that A is singular. Can we say anything if n is even?

9. Suppose $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq \mathbb{F}^n$. Prove that β is a basis if and only if $\det(\mathbf{v}_1 \cdots \mathbf{v}_n) \neq 0$.

10. Let $\dim V = n$, let $\mathrm{T} \in \mathcal{L}(V)$ and suppose β is a basis of V. Define $\det \mathrm{T} := \det[\mathrm{T}]_\beta$. Explain why this definition is independent of the choice of basis β.

11. (a) Suppose $X, Y \in M_n(\mathbb{F})$ can be written in block form

$$X = \left(\begin{array}{c|c} A & B \\ \hline O & I \end{array}\right) \qquad Y = \left(\begin{array}{c|c} I & C \\ \hline O & D \end{array}\right)$$

where O is a zero matrix and I an identity. Prove that $\det X = \det A$ and $\det Y = \det D$.

(b) Use part (a) to prove that $\det \left(\begin{array}{c|c} A & B \\ \hline O & C \end{array}\right) = \det A \det C$

12. Consider the system of equations

$$\begin{cases} 2x + 3y = 1 \\ x + 4y = 2 \end{cases} \iff \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

(a) Compute the inverse of the square matrix and, supposing $x, y \in \mathbb{R}$, find the solution.

(b) Suppose the above are now equations in the field $\mathbb{Z}_7$ of remainders modulo 7:

$$\begin{cases} 2x + 3y \equiv 1 \pmod 7 \\ x + 4y \equiv 2 \pmod 7 \end{cases}$$

Compute the inverse of the matrix in $\mathbb{Z}_7$ and use it to find x, y.

(c) What happens if you try to solve the system in $\mathbb{Z}_5$? Instead find the solutions to the system

$$\begin{cases} 2x + 3y \equiv 4 \pmod 5 \\ x + 4y \equiv 2 \pmod 5 \end{cases}$$

13. (a) Use Cramer's rule to prove the *classical adjoint* formula for the inverse:

$$(A^{-1})_{ij} = \frac{(-1)^{i+j}}{\det A} \det \tilde{A}_{ji}$$

Moreover, conclude that $\sum_{j=1}^{n} (-1)^{i+j} a_{ij} \left|\tilde{A}_{kj}\right| = \begin{cases} \det A & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}$

(b) Use part (a) to compute the inverse of $A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix}$

(c) Suppose that $A \in M_n(\mathbb{Z})$ is an invertible square matrix, all of whose entries are integers. Prove that the entries of A^{-1} are all integers if and only if $\det A = \pm 1$.

4.3 A Characterization of Determinant: non-examinable

In Exercise 4.1.5 we can give an alternative interpretation of the determinant. This construction can be done in general, though it involves a more advanced type of vector space.

Definition 4.23. Let V be a vector space over $\mathbb{F}$ and let $f : V \times \cdots \times V \to \mathbb{F}$ be a function from k copies of V to the field. We say that f is *k-multilinear* if it is linear in each entry: for each j,

$$f(\mathbf{v}_1, \dots, \mathbf{v}_{j-1}, \mathbf{x} + \lambda\mathbf{y}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_k) = f(\mathbf{v}_1, \dots, \mathbf{v}_{j-1}, \mathbf{x}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_k) + \lambda f(\mathbf{v}_1, \dots, \mathbf{v}_{j-1}, \mathbf{y}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_k)$$

An *alternating k-form* on V is a k-multilinear function f which evaluates to zero whenever two entries in the domain are equal:

$$f(\dots, \mathbf{v}, \dots, \mathbf{v}, \dots) = 0$$

Since the codomain of an alternating k-form is the (one-dimensional) vector space $\mathbb{F}$, the set of all alternating k-forms is a vector space over $\mathbb{F}$ in its own right, denoted $\bigwedge^k V^*$.

This may seem very abstract, but you've already seen an example: the determinant! Viewed as a function

$$\det : \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$$

Theorem 4.12 (part 1) and Corollary 4.14 say that $\det \in \bigwedge^n (\mathbb{F}^n)^*$ is an alternating n-form. We now observe something very special about this vector space.

Theorem 4.24. *If* $\dim V = n$*, then* $\dim \bigwedge^n V^* = 1$.

Proof. Let $f \in \bigwedge^n V^*$. Since f multilinear, it is determined completely by its values when applied to a basis $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V. In particular, it is determined by the n^n scalars

$$f(\mathbf{w}_1, \dots, \mathbf{w}_n) \quad \text{where each} \quad \mathbf{w}_i \in \beta \tag{$*$}$$

Since f is alternating, this is non-zero only if $\{\mathbf{w}_1, \dots, \mathbf{w}_n\} = \beta$ is the whole basis. Moreover, for any $\mathbf{x}, \mathbf{y} \in V$

$$\begin{aligned} 0 &= f(\dots, \mathbf{x} + \mathbf{y}, \dots, \mathbf{x} + \mathbf{y}, \dots) \\ &= f(\dots, \mathbf{x}, \dots, \mathbf{x}, \dots) + f(\dots, \mathbf{x}, \dots, \mathbf{y}, \dots) + f(\dots, \mathbf{y}, \dots, \mathbf{x}, \dots) + f(\dots, \mathbf{y}, \dots, \mathbf{y}, \dots) \\ &= f(\dots, \mathbf{x}, \dots, \mathbf{y}, \dots) + f(\dots, \mathbf{y}, \dots, \mathbf{x}, \dots) \qquad \text{(since } f \text{ is alternating)} \\ \implies & f(\dots, \mathbf{y}, \dots, \mathbf{x}, \dots) = -f(\dots, \mathbf{x}, \dots, \mathbf{y}, \dots) \end{aligned}$$

Applying the 'entry swapping' to the values $(*)$, we see that we may rearrange the order of the basis vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ so that they are in exactly the same order as they appear in β, at the cost of a $\pm$-sign. Otherwise said, f is *completely determined by the single value* $f(\mathbf{v}_1, \dots, \mathbf{v}_n)$.

A specific alternating n-form f may be defined by choosing $f(\mathbf{v}_1, \dots, \mathbf{v}_n) = 1$. If $g \in \bigwedge^n V^*$ is another alternating n-form, then

$$g = af \quad \text{where} \quad a = g(\mathbf{v}_1, \dots, \mathbf{v}_n)$$

All alternating n-forms are a scalar multiple of f, and so $\dim \bigwedge^n V^* = 1$. ■

This immediately leads to an alternative *non-inductive* definition of determinant.

Definition 4.25. The order n determinant is the unique alternating n-form $\det \in \bigwedge^n (\mathbb{F}^n)^*$ for which $\det(\mathbf{e}_1 \cdots \mathbf{e}_n) = 1$.

This is very sneaky. Rather than defining determinant inductively in terms of smaller determinants, we demand only that it satisfy certain properties. The problem with this definition is that it is very difficult to compute with explicitly, so it is good that we have the more elementary discussion to rely on. The alternative formulation is much more useful in the abstract.

The *exterior algebra* of alternating forms is widely applied in modern Geometry and Physics with a purpose analogous to how determinants measure (hyper-)volume.

Exercises 4.3

1. Let V be a vector space over $\mathbb{F}$ with dual space $V^* = \mathcal{L}(V, \mathbb{F})$. For any $f, g \in V^*$, define the *wedge product* $f \wedge g$ by
$$\forall \mathbf{v}, \mathbf{w} \in V, \quad f \wedge g(\mathbf{v}, \mathbf{w}) := \det \begin{pmatrix} f(\mathbf{v}) & f(\mathbf{w}) \\ g(\mathbf{v}) & g(\mathbf{w}) \end{pmatrix}$$
Prove that $f \wedge g \in \bigwedge^2 V^*$.

2. Let $\{f_1, f_2, f_3\}$ be the dual basis to $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ in $\mathbb{R}^3$ (that is $f_i(\mathbf{e}_j) = \delta_{ij}$). Prove that the cross-product of vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ is given by
$$\mathbf{v} \times \mathbf{w} = \begin{pmatrix} f_2 \wedge f_3(\mathbf{v}, \mathbf{w}) \\ f_3 \wedge f_1(\mathbf{v}, \mathbf{w}) \\ f_1 \wedge f_2(\mathbf{v}, \mathbf{w}) \end{pmatrix}$$
A similar approach can be used to construct an analogue of the cross product of three *vectors in* $\mathbb{R}^4$*: if you want a challenge, try to figure out how to create a vector orthogonal to* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^4$ *using three-forms.*

3. If $k \le n = \dim V$, prove that $\dim \bigwedge^k V^* = \binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient, namely the number of size-k subsets of a size n set.
(*Hint: given a basis* β *of* V*, what values are needed to defined* f*?*)

5 Diagonalization

5.1 Eigenvalues and Eigenvectors

Suppose V has a finite basis $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. We've seen that a linear map $\mathrm{T} \in \mathcal{L}(V)$ corresponds to multiplication by a matrix $[\mathrm{T}]_\beta \in M_n(\mathbb{F})$:

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

The most desirable situation is when this matrix is *diagonal*: otherwise said, $\exists \lambda_i \in \mathbb{F}$ such that

$$[\mathrm{T}]_\beta = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \quad \text{corresponding to} \quad \forall i,\ \mathrm{T}(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$$

Each vector $\mathbf{v}_i$ is transformed by T in a simple way: without meaningfully changing its direction.

Definition 5.1. Suppose V is a vector space over $\mathbb{F}$ and that $\mathrm{T} \in \mathcal{L}(V)$.

1. A *non-zero* vector $\mathbf{v} \in V$ is an *eigenvector*[a] of T with *eigenvalue* $\lambda \in \mathbb{F}$ if $\mathrm{T}(\mathbf{v}) = \lambda \mathbf{v}$.
2. The eigenvalues/vectors of $A \in M_n(\mathbb{F})$ are those of $\mathrm{L}_A \in \mathcal{L}(\mathbb{F}^n)$: the equation is $A\mathbf{v} = \lambda \mathbf{v}$.
3. If V is finite-dimensional, we say that T is *diagonalizable* if there exists a basis β of eigenvectors: otherwise said, $[\mathrm{T}]_\beta$ is diagonal. We call β an *eigenbasis.*

[a] In German, *eigen* indicates ownership: the term was coined by David Hilbert to indicate how eigenvalues and eigenvectors belong to a linear map. Earlier mathematicians used the word *characteristic* in a similar context.

Example 5.2. If $\beta = \{\mathbf{v}_1, \mathbf{v}_2\}$ is an eigenbasis for $A = \left(\begin{smallmatrix} 2 & 1 \\ 3 & 4 \end{smallmatrix}\right)$, then for $\mathbf{v}_j = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ and $\lambda_j = \lambda$, we have

$$\begin{cases} 2x + y = \lambda x \\ 3x + 4y = \lambda y \end{cases} \iff \begin{cases} y = (\lambda - 2)x \\ 3x = (\lambda - 4)y \end{cases} \iff \begin{cases} y = (\lambda - 2)x \\ 3x = (\lambda - 4)(\lambda - 2)x \end{cases}$$
$$\iff (\lambda - 4)(\lambda - 2) = 3 \qquad (*)$$

since $x = y = 0$ does not produce a basis vector. The polynomial has solutions $\lambda_1 = 5$, $\lambda_2 = 1$ which, upon substitution into the original equations, result in the eigenvectors[a]

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Plainly L_A is diagonalizable since $[\mathrm{L}_A]_\beta = \left(\begin{smallmatrix} 5 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is diagonal, and we conclude that $\{\mathbf{v}_1, \mathbf{v}_2\}$ really is an eigenbasis. Moreover, if $\epsilon = \{\mathbf{i}, \mathbf{j}\}$ is the standard basis, then

$$A = [\mathrm{L}_A]_\epsilon = Q_\beta^\epsilon [\mathrm{L}_A]_\beta Q_\epsilon^\beta = \begin{pmatrix} 1 & -1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 3 & 1 \end{pmatrix}^{-1}$$

where Q_β^ϵ is the change of co-ordinate matrix: thus $A = QDQ^{-1}$ where D is diagonal.

[a] There is some freedom here: any non-zero scalar multiples of $\mathbf{v}_1, \mathbf{v}_2$ are also eigenvectors; $[\mathrm{L}_A]_\beta$ is unchanged.

Warnings! The definition and example should remind you of the following critical facts:

- $\mathbf{0}$ is *never* an eigenvector! It is completely uninteresting to observe that $\mathbf{v} = \mathbf{0}$ solves *every* equation of the form $\mathrm{T}(\mathbf{v}) = \lambda\mathbf{v}$.

- If $\mathbf{v}$ is an eigenvector of T with eigenvalue λ, then so is any *non-zero* scalar multiple:

$$\mathrm{T}(k\mathbf{v}) = k\mathrm{T}(\mathbf{v}) = k\lambda\mathbf{v} = \lambda(k\mathbf{v})$$

Indeed in Example 5.2, the (infinitely many) eigenvectors of A have the form

$$a\mathbf{v}_1 = \begin{pmatrix} a \\ 3a \end{pmatrix}, \quad b\mathbf{v}_2 = \begin{pmatrix} -b \\ b \end{pmatrix} \quad \text{where } a, b \neq 0$$

What we really care about are *linearly independent* eigenvectors, of which A has only *two*: $\mathbf{v}_1, \mathbf{v}_2$. While strictly nonsense, it is common and acceptable to state that "A has two eigenvectors," rather than the more precise "A has two *linearly independent* eigenvectors."

Is every linear map diagonalizable? Does every linear map have eigenvectors?

These are the most obvious questions arising from the definition: the answers to both are a resounding *no*! To illustrate, here are several examples where we obtain many eigenvectors or very few.

Examples 5.3. 1. Let $A = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}$. If $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$ is an eigenvector with eigenvector λ, then

$$\begin{cases} x + 4y = \lambda x \\ y = \lambda y \end{cases} \implies xy + 4y^2 = \lambda xy = xy \implies y = 0 \implies \lambda = 1$$

A is non-diagonalizable: it has one independent eigenvector $\mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ with eigenvalue $\lambda = 1$.

2. The matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$ acts by rotation counter-clockwise by 90° in $\mathbb{R}^2$. Since $A\mathbf{v}$ is perpendicular to $\mathbf{v}$, we see that A has no eigenvectors! In particular, A is not diagonalizable.

   However, see Example 5.7.3 for what happens when A is viewed as a *complex* matrix.

3. Let $\mathrm{T} = \frac{\mathrm{d}}{\mathrm{d}x}$ be defined by differentiation on some vector space of functions V.

   A non-zero function $f \in V$ is an eigenvector (*eigenfunction*) of T with eigenvalue λ if and only if it satisfies the *natural growth equation* $f' = \lambda f$. As seen in calculus/ODEs, all solutions have the form $f(x) = ce^{\lambda x}$ where c is constant. Here are three specific cases:

   (a) If V is the space of all differentiable functions, then T has *infinitely many* linearly independent[a] eigenvectors $f(x) = e^{\lambda x}$. In this context diagonalizability is meaningless since V is infinite-dimensional.

   (b) If $V = P(\mathbb{R})$ is the space of polynomials, then T has exactly one independent eigenvector $f(x) = 1$ with eigenvalue $\lambda = 0$.

   (c) Let $\beta = \{\sin x, \cos x\}$ and $V = \mathrm{Span}_{\mathbb{R}}\{\sin x, \cos x\}$, then $[\mathrm{T}]_\beta = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is the matrix above, and so T has no eigenvectors.

   (d) If $\beta = \{e^x, e^{2x}, e^{5x}\}$ and $V = \mathrm{Span}_{\mathbb{R}}\,\beta$, then T is diagonalizable; indeed β is an eigenbasis.

[a] That these functions are linearly independent is a little tricky and was discussed in the first chapter.

Finding Eigenvalues and Eigenvectors in Finite-Dimensions

As Example 5.3.3 shows, linear operators on infinite-dimensional vector spaces can have eigenvectors, though the computation of such is usually case-specific. In the finite-dimensional situation, we can approach matters systematically. First we observe that we need only consider *matrices.*

Lemma 5.4. *Let* $\mathrm{T} \in \mathcal{L}(V)$ *where* $\dim V = n$ *and suppose* β *is a basis of* V*. Then*

$$\mathrm{T}(\mathbf{v}) = \lambda\mathbf{v} \iff [\mathrm{T}]_\beta[\mathbf{v}]_\beta = \lambda[\mathbf{v}]_\beta$$

Otherwise said:

- T *has the same eigenvalues as any matrix of* T *with respect to any basis.*
- *The co-ordinate isomorphism* $\phi_\beta : V \to \mathbb{F}^n : \mathbf{v} \mapsto [\mathbf{v}]_\beta$ *maps eigenvectors of* T *to those of* $[\mathrm{T}]_\beta$*.*

The lemma says that to compute the eigenvalues and eigenvectors of T, we simply compute those of its matrix $[\mathrm{T}]_\beta$ with respect to any basis β and then translate.

With this identification out of the way, let $A \in M_n(\mathbb{F})$ have eigenvector $\mathbf{v}$ with eigenvalue λ. Observe:

$$A\mathbf{v} = \lambda\mathbf{v} \iff (A - \lambda I)\mathbf{v} = \mathbf{0} \tag{†}$$

where I is the identity matrix. Since $\mathbf{v} \neq \mathbf{0}$, the nullspace $\mathcal{N}(A - \lambda I)$ is non-trivial. Indeed

$$\begin{aligned} \lambda \text{ is an eigenvalue of } A &\iff \operatorname{null}(A - \lambda I) > 0 \iff \operatorname{rank}(A - \lambda I) < n \\ &\iff \det(A - \lambda I) = 0 \end{aligned}$$

where we used the Rank–Nullity Theorem and a standard property of the determinant.

Definition 5.5. The *characteristic polynomial* of a matrix A is $p(t) := \det(A - tI)$.
When $\dim V = n$, the *characteristic polynomial* of $\mathrm{T} \in \mathcal{L}(V)$ may be computed with respect to any basis β of V

$$p(t) = \det(\mathrm{T} - t\mathrm{I}) := \det[\mathrm{T} - t\mathrm{I}]_\beta = \det([\mathrm{T}]_\beta - tI_n)$$

In either case, the *characteristic equation* is $p(t) = 0$.

Plainly λ is an eigenvalue if and only if $p(\lambda) = 0$. Once we have an eigenvalue, (†) says that the corresponding eigenvectors lie in the nullspace $\mathcal{N}(A - \lambda I)$. To summarize:

Theorem 5.6. *Let* $A \in M_n(\mathbb{F})$*.*

1. *The characteristic polynomial* $p(t)$ *is a degree* n *polynomial in* t *with leading term* $(-1)^n t^n$*.*
2. A *has at most* n *eigenvalues, precisely the solutions to the characteristic equation* $p(t) = 0$*.*
3. *An eigenvector with eigenvalue* λ *is any non-zero element of the* eigenspace $E_\lambda := \mathcal{N}(A - \lambda I)$*.*

Once part 1 is proved, the rest follows immediately from our above discussion and the fact that a degree n polynomial has at most n solutions. Before seeing this, we revisit our past examples in this language and see another.

Examples 5.7. 1. (Example 5.2) $A = \left(\begin{smallmatrix} 2 & 1 \\ 3 & 4 \end{smallmatrix}\right)$ has characteristic polynomial

$$p(t) = \det\begin{pmatrix} 2-t & 1 \\ 3 & 4-t \end{pmatrix} = (2-t)(4-t) - 3 = t^2 - 6t + 5 = (t-5)(t-1)$$

recovering the eigenvalues $\lambda_1 = 5$ and $\lambda_2 = 1$. We can now find the nullspaces:

- $\lambda_1 = 5$: $\mathcal{N}(A - \lambda_1 I) = \mathcal{N}\begin{pmatrix} -3 & 1 \\ 3 & -1 \end{pmatrix} = \text{Span}\begin{pmatrix} 1 \\ 3 \end{pmatrix}$
- $\lambda_2 = 1$: $\mathcal{N}(A - \lambda_2 I) = \mathcal{N}\begin{pmatrix} 1 & 1 \\ 3 & 3 \end{pmatrix} = \text{Span}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

We may therefore choose two independent eigenvectors $\mathbf{v}_1 = \left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right)$, $\mathbf{v}_2 = \left(\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right)$: these form an eigenbasis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

2. (Example 5.3.2) $A = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ has characteristic equation $p(t) = \det\left(\begin{smallmatrix} -t & -1 \\ 1 & -t \end{smallmatrix}\right) = t^2 + 1 = 0$. Since this has no solutions (in $\mathbb{R}$), we see that A has no eigenvalues. However, if we consider $A \in M_2(\mathbb{C})$ as a *complex* matrix, then there are two eigenvalues $\lambda_1 = i$ and $\lambda_2 = -i$: indeed

$$\mathcal{N}(A - \lambda_1 I) = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} = \text{Span}\begin{pmatrix} i \\ 1 \end{pmatrix} \quad \text{and} \quad \mathcal{N}(A - \lambda_2 I) = \text{Span}\begin{pmatrix} -i \\ 1 \end{pmatrix}$$

so we may choose two independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{C}^2$. These form a basis and so A is diagonalizable as a complex matrix.

3. Let $\mathrm{T} \in \mathcal{L}(P_2(\mathbb{R}))$ be defined by

$$\mathrm{T}(f)(x) = \int_0^2 f(x)\,\mathrm{d}x + (x-3)f'(x)$$

With respect to the standard basis, we have the matrix $A = [\mathrm{T}]_\epsilon = \left(\begin{smallmatrix} 2 & -1 & \frac{8}{3} \\ 0 & 1 & -6 \\ 0 & 0 & 2 \end{smallmatrix}\right)$ whose eigenvalues are the solutions of the characteristic equation

$$0 = p(t) = \det(A - tI) = (2-t)^2(1-t) \iff t = 1, 2$$

Now compute the nullspaces:

- $\lambda_1 = 1$: $\mathcal{N}(A - \lambda_1 I) = \mathcal{N}\begin{pmatrix} 1 & -1 & \frac{8}{3} \\ 0 & 0 & -6 \\ 0 & 0 & 1 \end{pmatrix} = \text{Span}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$
- $\lambda_2 = 2$: $\mathcal{N}(A - \lambda_2 I) = \mathcal{N}\begin{pmatrix} 0 & -1 & \frac{8}{3} \\ 0 & -1 & -6 \\ 0 & 0 & 0 \end{pmatrix} = \text{Span}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

We may therefore choose two independent eigenvectors of A; $\mathbf{v}_1 = \left(\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right)$ and $\mathbf{v}_2 = \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right)$. These correspond to polynomials $f_1, f_2 \in P_2(\mathbb{R})$ or *eigenfunctions* of T:

$$\mathbf{v}_1 = [f_1]_\epsilon \implies f_1(x) = 1 + x$$
$$\mathbf{v}_2 = [f_2]_\epsilon \implies f_2(x) = 1$$

It is easily checked directly that $\mathrm{T}(f_1) = f_1$ and $\mathrm{T}(f_2) = 2f_2$. Since T has insufficient independent eigenvectors, we see that it is *not* diagonalizable.

We now give an induction argument to complete the proof of Theorem 5.6: that $p(t)$ is a degree n polynomial. First observe the following obvious fact: in any cofactor expansion of the determinant, one never multiplies an entry of a matrix by itself…

Lemma 5.8. *If $B(t)$ is a square matrix, k of whose entries are linear functions of t with the rest constant, then* $\det B(t)$ *is a polynomial in t with degree $\le k$.*

The main argument is a little difficult to follow, so first consider an example where we expand the characteristic polynomial of a 3×3 matrix along the first row.

$$B = \begin{pmatrix} 1 & 3 & 2 \\ -1 & 0 & 4 \\ 1 & -2 & 5 \end{pmatrix} \implies p(t) = \begin{vmatrix} 1-t & 3 & 2 \\ -1 & -t & 4 \\ 1 & -2 & 5-t \end{vmatrix}$$
$$= (1-t)\begin{vmatrix} -t & 4 \\ -2 & 5-t \end{vmatrix} - 3\begin{vmatrix} -1 & 4 \\ 1 & 5-t \end{vmatrix} + 2\begin{vmatrix} -1 & -t \\ 1 & -2 \end{vmatrix}$$
$$= (b_{11} - t)\det \tilde{B}_{11}(t) - b_{12}\det \tilde{B}_{12}(t) + b_{13}\det \tilde{B}_{13}(t)$$

In each case $\tilde{B}_{1j}(t)$ is the $1j^{\text{th}}$ minor of the matrix $B - tI$. Observe that

$$\deg(\det \tilde{B}_{1j}(t)) = \begin{cases} 2 & \text{if } j = 1 \\ 1 & \text{otherwise} \end{cases} \implies \deg(p(t)) = 3$$

This is essentially the induction step in the following proof with $n = 2$.

Proof of Theorem 5.6, part 1. Since only n entries of the matrix $A - tI$ contain t, the Lemma tells us that the maximum degree of $p(t) = \det(A - tI)$ is n.

It remains to prove that the leading term of $p(t)$ is $(-1)^n t^n$: we prove by induction on n.

(*Base Case*) If $n = 1$, then $A = (a)$ and so $p(t) = -t + a$ as required.

(*Induction Step*) Fix $n \in \mathbb{N}$ and assume for every matrix $A \in M_n(\mathbb{F})$ that

$$p(t) = (-1)^n t^n + \cdots$$

Let $B \in M_{n+1}(\mathbb{F})$ and compute using the cofactor expansion along the first row:

$$\det(B - tI) = (b_{11} - t)\det \tilde{B}_{11}(t) - b_{12}\det \tilde{B}_{12}(t) + \cdots$$

where $\tilde{B}_{1j}(t)$ is the $n \times n$ minor obtained by deleting the 1^{st} row and j^{th} column of $B - tI$. There are two cases:

If $j = 1$: $\tilde{B}_{11}(t) = B_{11} - tI \in M_n(\mathbb{F})$. By the induction hypothesis its determinant is a degree n polynomial with leading term $(-1)^n t^n$. It follows that

$$(b_{11} - t)\det \tilde{B}_{11}(t) = (-1)^{n+1} t^{n+1} + \text{lower order terms}$$

If $j \ge 2$: $\tilde{B}_{1j}(t) \in M_n(\mathbb{F})$ where $n - 1$ of the entries contain a t: we have deleted the first row and j^{th} column and thus removed *two* of the diagonal t-terms from $B - tI$. By the Lemma, $\det \tilde{B}_{1j}(t)$ is a polynomial of degree *at most* $n - 1$.

Summing these polynomials completes the proof. ■

Exercises 5.1 1. For each matrix $A \in M_n(\mathbb{F})$, find the eigenvalues and a set of linearly independent eigenvectors. If an eigenbasis exists, state an invertible matrix Q and a diagonal matrix D such that $A = QDQ^{-1}$.

(a) $A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \in M_2(\mathbb{R})$ (b) $A = \begin{pmatrix} 0 & -2 & -3 \\ -1 & 1 & -1 \\ 2 & 2 & 5 \end{pmatrix} \in M_3(\mathbb{R})$

(c) $A = \begin{pmatrix} i & 1 \\ 2 & -i \end{pmatrix} \in M_2(\mathbb{C})$ (d) $A = \begin{pmatrix} 2 & 0 & -1 \\ 4 & 1 & -4 \\ 2 & 0 & -1 \end{pmatrix} \in M_3(\mathbb{R})$

2. For each linear operator T on a vector space V, find an ordered basis β such that $[\mathrm{T}]_\beta$ is diagonal.

(a) $V = P_2(\mathbb{R})$ and $\mathrm{T}\big(f(x)\big) = xf'(x) + f(2)x + f(3)$

(b) $V = P_3(\mathbb{R})$ and $\mathrm{T}\big(f(x)\big) = xf'(x) + f''(x) - f(2)$

3. If A and B are similar matrices ($B = QAQ^{-1}$ for some Q), prove that $\mathbf{v}$ is an eigenvector of A if and only if $Q\mathbf{v}$ is an eigenvector of B with the same eigenvalue.

4. Prove that the characteristic polynomial $p(t) = \det(\mathrm{T} - t\mathrm{I}) = \det([\mathrm{T}]_\beta - tI)$ of a linear map $\mathrm{T} \in \mathcal{L}(V)$ is independent of the choice of basis β used in its computation.

5. Suppose $A \in M_n(\mathbb{C})$ is a *real* matrix with eigenvalue $\lambda \in \mathbb{C}$ and eigenvector $\mathbf{v} \in \mathbb{C}^n$.

(a) Prove that the complex conjugate $\overline{\mathbf{v}}$ is also an eigenvector. What is its eigenvalue?

(b) Prove that if $\mathbf{v} = a\mathbf{w}$ for some (complex) scalar a and *real vector* $\mathbf{w} \in \mathbb{R}^n$, then $\lambda \in \mathbb{R}$.

(c) Is the converse of part (b) true? Explain. In particular, if $\lambda \in \mathbb{R}$, consider the real and imaginary parts of $\mathbf{v}$

$$\mathbf{x} := \frac{1}{2}(\mathbf{v} + \overline{\mathbf{v}}) \qquad \mathbf{y} := \frac{1}{2i}(\mathbf{v} - \overline{\mathbf{v}})$$

and prove that $\dim_{\mathbb{R}} \operatorname{Span}\{\mathbf{x}, \mathbf{y}\} = \dim_{\mathbb{C}} \operatorname{Span}\{\mathbf{v}, \overline{\mathbf{v}}\}$. What does this mean for the eigenvectors of A?

6. Let $p(t) = (-1)^n t^n + c_{n-1}t^{n-1} \cdots + c_0$ be the characteristic polynomial of a matrix A.

(a) Prove that $c_0 = \det A$ and hence conclude that A is invertible if and only if $c_0 \neq 0$.

(b) Prove that $p(t) = (a_{11} - t)(a_{22} - t) \cdots (a_{nn} - t) + q(t)$ where $\deg q(t) \leq n - 2$. Hence argue that $c_{n-1} = (-1)^{n-1} \operatorname{tr} A$.
(*Hint: try an induction proof*)

5.2 Diagonalizability

We have now seen how to compute eigenvectors in finite dimensions, and observed that diagonalizability is equivalent to the existence of an eigenbasis. In this section we consider the question of when an eigenbasis might exist.

Theorem 5.9. *Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are eigenvectors of $\mathrm{T} \in \mathcal{L}(V)$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent.*

Proof. We prove by induction on k. The base case $k = 1$ is trivial.

Fix $k \in \mathbb{N}$: for the induction hypothesis, suppose *every* set of k eigenvectors corresponding to k distinct eigenvalues is linearly independent. To obtain a contradiction, suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}\}$ is a *linearly dependent* set of $k+1$ eigenvectors corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_{k+1}$. WLOG, we may assume

$$\exists a_1, \ldots, a_k \in \mathbb{F} \quad \text{such that} \quad a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k + \mathbf{v}_{k+1} = \mathbf{0} \tag{$*$}$$

Apply T to this linear dependence and substitute for $\mathbf{v}_{k+1}$ using $(*)$:

$$\sum_{j=1}^{k} a_j\lambda_j\mathbf{v}_j + \lambda_{k+1}\mathbf{v}_{k+1} = \mathbf{0} \implies \sum_{j=1}^{k} a_j(\lambda_j - \lambda_{k+1})\mathbf{v}_j = \mathbf{0}$$
$$\implies a_j(\lambda_j - \lambda_{k+1}) = 0 \implies a_j = 0$$

where we used the linear independence of $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and the distinctness of the $\lambda_1, \ldots, \lambda_{k+1}$. But this shows that $\mathbf{v}_{k+1} = \mathbf{0}$ is not an eignevector: contradiction. We conclude that $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ is linearly independent.

By induction, the result is proved. ∎

Suppose $\dim V = n$ and that the degree n characteristic polynomial of $\mathrm{T} \in \mathcal{L}(V)$ has distinct roots;[5]

$$p(t) = (-1)^n(t - \lambda_1)\cdots(t - \lambda_n) = (\lambda_1 - t)\cdots(\lambda_n - t)$$

where $\lambda_1, \ldots, \lambda_n$ are the distinct eigenvalues of T. Since each λ_j implies the existence of at least one eigenvector $\mathbf{v}_j$, the Theorem says that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent and thus a basis of V (an *eigenbasis* for T). We therefore have a simple sufficient condition for the diagonalizability of T.

Corollary 5.10. *Suppose $\dim_{\mathbb{F}} V = n$ and $\mathrm{T} \in \mathcal{L}(V)$. If T has n distinct eigenvalues (equivalently $p(t)$ has n distinct roots in the field $\mathbb{F}$), then T is diagonalizable.*

To orient ourselves, recall Examples 5.7.

1. $A = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} \in M_2(\mathbb{R})$ has distinct eigenvalues $\lambda = 1, 5 \in \mathbb{R}$ and is diagonalizable.
2. $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{C})$ has distinct eigenvalues $\lambda = \pm i \in \mathbb{C}$ and is diagonalizable.
3. $\mathrm{T} \in \mathcal{L}(P_2(\mathbb{R}))$ defined by $\mathrm{T}(f)(x) = \int_0^2 f(x)\,\mathrm{d}x + (x-3)f'(x)$ has only two distinct eigenvalues $\lambda = 1, 2$ and is non-diagonalizable.

[5]From algebra, every degree n polynomial has *at most* n distinct roots.

After reviewing the examples, it might feel as if the Corollary should be biconditional. However, a trivial example says not: the identity map $\mathrm{I}_V \in \mathrm{T}(V)$ has only one eigenvalue $\lambda = 1$ but is plainly diagonalizable (by *any* basis!). We now develop a *necessary* condition for diagonalizability.

Definition 5.11. A degree n polynomial $p(t)$ *splits over a field* $\mathbb{F}$ if it factorizes completely over $\mathbb{F}$. Otherwise said, $\exists c, \alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that

$$p(t) = c(t - \alpha_1) \cdots (t - \alpha_n)$$

The values $\alpha_1, \ldots, \alpha_n$ are the *roots* or *zeros* of the polynomial.

Example 5.12. $p(t) = t^2 + 4 = (t - 2i)(t + 2i)$ does not split over $\mathbb{R}$, but does split over $\mathbb{C}$.

Theorem 5.13. *If* T *is diagonalizable, then its characteristic polynomial splits.*

Proof. Let $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an eigenbasis, then $[\mathrm{T}]_\beta$ is diagonal with the eigenvalues down the diagonal. But then the characteristic polynomial of T splits:

$$p(t) = \det([\mathrm{T}]_\beta - tI) = (\lambda_1 - t) \cdots (\lambda_n - t)$$

■

Putting Corollary 5.10 and Theorem 5.13 together, we have

$$p(t) \text{ has distinct roots} \implies \mathrm{T} \text{ diagonalizable} \implies p(t) \text{ splits}$$

Our 'identity' observation above shows that these conditions are not equivalent. Here is another example of repeated eigenvalues.

Examples 5.14. The polynomial $p(t) = (6 - t)(4 - t)^2$ splits but does not have three distinct roots. This is not an idle example, for p is the characteristic polynomial of *many* linear maps, some diagonalizable, some not. For instance:

1. $A = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$ is diagonalizable (it's already diagonal!) with eigenbasis $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$.
2. $B = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{pmatrix}$ is non-diagonalizable. To verify this, observe that

$$\mathcal{N}(B - 6I) = \mathcal{N}\begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & -2 \end{pmatrix} = \operatorname{Span}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$
$$\mathcal{N}(B - 4I) = \mathcal{N}\begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \operatorname{Span}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

We can therefore find only two independent eigenvectors $\mathbf{v}_1 = \mathbf{i}$ and $\mathbf{v}_2 = \mathbf{j}$.

To obtain a fuller description of diagonalizability, we need to come to terms with the discrepancy above: $p(t)$ has root ($\lambda = 4$) with multiplicity *two*, but we can only find *one* independent eigenvector ($\mathbf{v}_2 = \mathbf{j}$) with this eigenvalue $B\mathbf{v}_2 = 4\mathbf{v}_2$.

Definition 5.15. Suppose V is finite-dimensional and that $\mathrm{T} \in \mathcal{L}(V)$ has an eigenvalue λ.

1. The *geometric multiplicity* of λ is the dimension $\dim E_\lambda$ of its *eigenspace*[a]

$$E_\lambda := \mathcal{N}(\mathrm{T} - \lambda \mathrm{I})$$

2. The *algebraic multiplicity* $\operatorname{mult}(\lambda)$ of λ is the highest power m for which $(t-\lambda)^m$ is a factor of the characteristic polynomial $p(t)$. Otherwise said, there exists a polynomial $q(t)$ such that

$$p(t) = (t-\lambda)^m q(t) \quad \text{and} \quad q(\lambda) \neq 0$$

[a] $\mathbf{v}$ is an eigenvector with eigenvalue λ if and only if $\mathbf{v} \in E_\lambda$ is *non-zero*.

Example (5.14, mark II). Here are the eigenspaces and multiplicities for B: note how the algebraic and geometric multiplicities differ.

eigenvalue λ	6	4
algebraic multiplicity $\operatorname{mult}(\lambda)$	1	2
eigenspace E_λ	Span $\mathbf{i}$	Span $\mathbf{j}$
geometric multiplicity $\dim E_\lambda$	1	1

We can now state the main result.

Theorem 5.16. *Suppose* $\dim V = n$ *and that* $\mathrm{T} \in \mathcal{L}(V)$ *has* distinct *eigenvalues* $\lambda_1, \ldots, \lambda_k$.

1. *For each eigenvalue* λ_i*, we have* $\dim E_{\lambda_i} \leq \operatorname{mult}(\lambda_i)$.
2. *The following are equivalent:*
   (a) T *is diagonalizable.*
   (b) *The characteristic polynomial of* T *splits and* $\dim E_{\lambda_i} = \operatorname{mult}(\lambda_i)$ *for each i;*

$$p(t) = (\lambda_1 - t)^{\dim E_{\lambda_1}} \cdots (\lambda_k - t)^{\dim E_{\lambda_k}}$$

   (c) $\sum_{i=1}^k \dim E_{\lambda_i} = n$.
   (d) $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$.

We'll prove this shortly, but first, here are two examples where the calculations have been omitted.

Examples 5.17. 1. $A = \begin{pmatrix} 2&0&0&0 \\ 0&3&1&0 \\ 0&0&3&1 \\ 0&0&0&3 \end{pmatrix}$ is non-diagonalizable: $p(t) = (2-t)(3-t)^3$ splits, but,

λ	2	3
$\operatorname{mult}(\lambda)$	1	3
E_λ	Span $\mathbf{e}_1$	Span $\mathbf{e}_2$
$\dim E_\lambda$	1	1

$$\dim E_3 \neq \operatorname{mult}(3), \quad \text{and} \quad \sum_{i=1}^{2} \dim E_{\lambda_i} = 2 < 4$$

2. Let $B = \begin{pmatrix} -1 & 6 & 0 & 0 \\ -2 & 6 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$ is diagonalizable. Indeed $p(t) = (2-t)(3-t)^3$ splits, and we have

λ	2	3
$\text{mult}(\lambda)$	1	3
E_λ	$\text{Span}(2\mathbf{e}_1 + \mathbf{e}_2)$	$\text{Span}\{3\mathbf{e}_1 + 2\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$
$\dim E_\lambda$	1	3

and $\mathbb{R}^4 = E_2 \oplus E_3$

From the table, we can read off an eigenbasis with respect to which the map is diagonal

$$\beta = \left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\} \implies [\mathrm{L}_B]_\beta = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Proof. 1. Let $r = \dim E_\lambda$ and extend a basis β_λ of E_λ to a basis $\beta = \beta_\lambda \cup \gamma$ of V.
Since $\mathrm{T}(\mathbf{v}) = \lambda \mathbf{v}$ for all $\mathbf{v} \in E_\lambda$, we see that the matrix of T has block form $[\mathrm{T}]_\beta = \left(\begin{array}{c|c} \lambda I_r & A \\ \hline O & B \end{array}\right)$ for some matrices A, B, from which the characteristic polynomial of T is

$$p(t) = \det \left(\begin{array}{c|c} (\lambda - t) I_r & A \\ \hline O & B - tI_{n-r} \end{array}\right) = (\lambda - t)^r \det(B - tI_{n-r})$$

It follows that $(\lambda - t)^{\dim E_\lambda}$ divides $p(t)$, and so $\dim E_\lambda \leq \text{mult}(\lambda)$.

2. We give a brief summary:

(a) $\implies$ (b) If T is diagonalizable, then $p(t)$ splits by Theorem 5.13, whence $\sum \text{mult}(\lambda_i) = n$. The cardinality n of an eigenbasis is *at most* $\sum \dim E_{\lambda_i}$. Combined with part 1, we have equality of multiplicities:

$$n \leq \sum \dim E_{\lambda_i} \leq \sum \text{mult}(\lambda_i) = n \implies \dim E_{\lambda_i} = \text{mult}(\lambda_i)$$

(b) $\implies$ (c) $p(t)$ splits $\implies n = \sum \text{mult}(\lambda_i) = \sum \dim E_{\lambda_i}$

(c) $\implies$ (d) This requires an induction on the number of distinct eigenvalues.
For the induction step, fix $j < k$ and let $\mathbf{v}_{j+1}$ be an eigenvector with eigenvalue λ_{j+1}. If $\mathbf{v}_{j+1} \in E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_j}$ then there exist eigenvectors $\mathbf{v}_i \in E_{\lambda_i}$ and $a_i \in \mathbb{F}$ for which

$$\mathbf{v}_{j+1} = a_1 \mathbf{v}_1 + \cdots + a_j \mathbf{v}_j$$

But this contradicts the linear independence of the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_{j+1}\}$ (Theorem 5.9).

By induction, $E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$ exists; by assumption it has dimension $n = \dim V$ and thus equals V.

(d) $\implies$ (a) is trivial since (d) says there exists a basis of eigenvectors. ∎

Exercises 5.2 1. For each matrix, find its characteristic polynomial, its eigenvalues/spaces, its algebraic and geometric multiplicities and decide if it is diagonalizable.

(a) $A = \begin{pmatrix} 4 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$ (b) $B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -4 & 0 & 4 \\ 0 & 0 & 1 & 0 \\ 0 & -6 & 0 & 6 \end{pmatrix}$

2. Let $\mathrm{T} = \frac{\mathrm{d}}{\mathrm{d}x}$ be the derivative operator.
   (a) If we consider $\mathrm{T} = \mathcal{L}(P_2(\mathbb{R}))$, show that T is *not* diagonalizable.
   (b) More generally, what is the characteristic polynomial of $\mathrm{T} \in \mathcal{L}(P_n(\mathbb{R}))$? Why is it clear that T is non-diagonalizable?

3. Diagonalize $A = \left(\begin{smallmatrix} 1 & 4 \\ 2 & 3 \end{smallmatrix}\right) \in M_2(\mathbb{R})$, and thus find an expression for A^n for any $n \in \mathbb{N}$.

4. Show that the characteristic polynomial of $A = \left(\begin{smallmatrix} 3 & -4 \\ 4 & 3 \end{smallmatrix}\right)$ does not split over $\mathbb{R}$. Diagonalize A over $\mathbb{C}$.

5. Suppose T is a linear operator on a finite dimensional vector space V and that β is a basis of V with respect to which $[\mathrm{T}]_\beta$ is diagonal. Prove that the characteristic polynomial of T splits.

6. Suppose $\mathrm{T} \in \mathcal{L}(V)$ is invertible with eigenvalue λ. Prove that λ^{-1} is an eigenvalue of T^{-1} with the same eigenspace. If T is diagonalizable, prove that T^{-1} is diagonalizable.

7. If $p(t)$ splits, prove that $\det \mathrm{T} = \lambda_1^{\mathrm{mult}(\lambda_1)} \cdots \lambda_k^{\mathrm{mult}(\lambda_k)}$ is the product of its distinct eigenvalues up to (algebraic) multiplicity.

5.3 Invariant Subspaces and the Cayley–Hamilton Theorem

Eigenspaces of a linear map provide a simple example of a special type of subspace.

Definition 5.18. Suppose $\mathrm{T} \in \mathcal{L}(V)$. A subspace W of V is T-*invariant* if $\mathrm{T}(W) = \{\mathrm{T}(\mathbf{w}) : \mathbf{w} \in W\}$ is a subspace of W. In such a case we define the *restriction* $\mathrm{T}_W \in \mathcal{L}(W)$ by

$$\mathrm{T}_W : W \to W : \mathbf{w} \mapsto \mathrm{T}(\mathbf{w})$$

Much can often be understood about a linear map by considering its invariant subspaces.
We start by extending the proof of Theorem 5.16 (part 1) to *any* invariant subspace.

Theorem 5.19. *Suppose $\mathrm{T} \in \mathcal{L}(V)$, that $\dim V$ is finite and that $W \leq V$ is T-invariant. Then the characteristic polynomial $p_W(t)$ of T_W divides that of T.*

Proof. Extend a basis β_W of W to a basis $\beta = \beta_W \cup \gamma$ of V. Then $\exists A, B$ such that

$$[\mathrm{T}]_\beta = \left(\begin{array}{c|c} [\mathrm{T}_W]_{\beta_W} & A \\ \hline O & B \end{array}\right) \implies p(t) = \det([\mathrm{T}_W]_{\beta_W} - tI)\det(B - tI) = p_W(t)\det(B - tI)$$

∎

Examples 5.20. 1. Every eigenspace E_λ is T-invariant: $\forall \mathbf{w} \in E_\lambda$, we have, $\mathrm{T}(\mathbf{w}) = \lambda\mathbf{w} \in E_\lambda$.
Restricted to the eigenspace, the linear map is simply $\mathrm{T}_{E_\lambda} = \lambda \mathrm{I}$, with characteristic polynomial $p_\lambda(t) = (\lambda - t)^{\dim E_\lambda}$. This divides $p(t)$, as seen in Theorem 5.16.

2. If $A = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{pmatrix}$, then $\mathrm{L}_A \in \mathcal{L}(\mathbb{R}^3)$ has an invariant subspace $W = \operatorname{Span}\{\mathbf{i}, \mathbf{j}\}$. It is easy to check that, with respect to the standard basis, the restriction of L_A to W has matrix $\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ Since both this and A are upper triangular, we quickly verify that

$$p(t) = (1-t)(2-t)(3-t) = (2-t)p_W(t)$$

We now consider a more general type of invariant subspace.

Definition 5.21. Let $\mathrm{T} \in \mathcal{L}(V)$ and $\mathbf{v} \in V$. The T-*cyclic subspace generated by* $\mathbf{v}$ is the span

$$\langle \mathbf{v} \rangle = \operatorname{Span}\{\mathbf{v}, \mathrm{T}(\mathbf{v}), \mathrm{T}^2(\mathbf{v}), \ldots\}$$

The T-cyclic subspace $\langle \mathbf{v} \rangle$ is the *smallest* T-invariant subspace containing $\mathbf{v}$ (see Exercise 5.3.4).

Examples 5.22. 1. If $A = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -4 & 1 \\ 0 & 0 & -4 \end{pmatrix}$, then the eigenspaces are L_A-cyclic subspaces:

$$E_5 = \operatorname{Span}\mathbf{i} = \langle \mathbf{i} \rangle, \qquad E_{-4} = \operatorname{Span}\mathbf{j} = \langle \mathbf{j} \rangle$$

There are other examples, for instance $\langle \mathbf{k} \rangle = \operatorname{Span}\{\mathbf{j}, \mathbf{k}\}$ is L_A-cyclic, but is not an eigenspace.

2. $\dim \langle \mathbf{v} \rangle = 1 \iff \mathbf{v}$ is an eigenvector of T.

3. Not every T-invariant subspace is T-cyclic: for instance, if $\mathrm{T} = \mathrm{I}$ is the identity, then every subspace is T-invariant, however only the *one-dimensional* subspaces are T-cyclic!

For T-cyclic subspaces, we can extend Theorem 5.19 further.

Theorem 5.23. *Let V be finite dimensional, $\mathrm{T} \in \mathcal{L}(V)$, and suppose $W = \langle \mathbf{w} \rangle$ is* T*-invariant with* $\dim W = k$. *Then:*

1. $\beta_W = \{\mathbf{w}, \mathrm{T}(\mathbf{w}), \ldots, \mathrm{T}^{k-1}(\mathbf{w})\}$ *is a basis of* W.
2. *If* $\mathrm{T}^k(\mathbf{w}) + a_{k-1}\mathrm{T}^{k-1}(\mathbf{w}) + \cdots + a_0\mathbf{w} = \mathbf{0}$, *then the characteristic polynomial of* T_W *is*
$$p_W(t) = (-1)^k \left(t^k + a_{k-1}t^{k-1} + \cdots + a_1 t + a_0\right)$$
3. $p_W(\mathrm{T}_W) = 0$.

Proof. 1. Let i be maximal such that $\{\mathbf{w}, \mathrm{T}(\mathbf{w}), \ldots, \mathrm{T}^{i-1}(\mathbf{w})\}$ is linearly independent. Observe:

- Plainly i exists since a maximal linearly independent set is finite ($\dim W < \infty$).
- By the maximality of i, $\mathrm{T}^i(\mathbf{w}) \in \operatorname{Span}\{\mathbf{w}, \mathrm{T}(\mathbf{w}), \ldots, \mathrm{T}^{i-1}(\mathbf{w})\}$; by induction this extends to
$$j \geq i \implies \mathrm{T}^j(\mathbf{w}) \in \operatorname{Span}\{\mathbf{w}, \mathrm{T}(\mathbf{w}), \ldots, \mathrm{T}^{i-1}(\mathbf{w})\}$$
It follows that $W = \operatorname{Span}\{\mathbf{w}, \mathrm{T}(\mathbf{w}), \ldots, \mathrm{T}^{i-1}(\mathbf{w})\}$.

We conclude that $\{\mathbf{w}, \mathrm{T}(\mathbf{w}), \ldots, \mathrm{T}^{i-1}(\mathbf{w})\}$ is a basis of W, whence $i = k$.

2. Expand the characteristic polynomial along the first row:
$$p_W(t) = \det([\mathrm{T}_W]_{\beta_W} - tI_k) = \begin{vmatrix} -t & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & -t & 0 & & 0 & -a_1 \\ 0 & 1 & -t & & 0 & -a_2 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & & -t & -a_{k-2} \\ 0 & 0 & 0 & \cdots & 1 & -a_{k-1}-t \end{vmatrix}$$
$$= -t \begin{vmatrix} -t & 0 & & 0 & -a_1 \\ 1 & -t & & 0 & -a_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & -t & -a_{k-2} \\ 0 & 0 & \cdots & 1 & -a_{k-1}-t \end{vmatrix} + (-1)^k a_0 \begin{vmatrix} 1 & -t & 0 & \cdots & 0 \\ 0 & 1 & -t & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & & \ddots & -t \\ 0 & 0 & \cdots & & 1 \end{vmatrix}$$
The second matrix has determinant 1, yielding the $(-1)^k a_0$ term. The first is $-t$ multiplied by a determinant of the same type but one dimension lower. An induction finishes things off.

3. Write $\mathrm{S} \in \mathcal{L}(V)$ for the linear map
$$\mathrm{S} := p_W(\mathrm{T}) = (-1)^k(\mathrm{T}^k + a_{k-1}\mathrm{T}^{k-1} + \cdots + a_0\mathrm{I})$$
Part 2 says $\mathrm{S}(\mathbf{w}) = \mathbf{0}$. Since S is a polynomial in T, it commutes with all powers of T:
$$\forall i,\ \mathrm{S}(\mathrm{T}^i(\mathbf{w})) = \mathrm{T}^i(\mathrm{S}(\mathbf{w})) = \mathbf{0}$$
Since S is zero on the basis β_W of W, we see that S_W is the zero function. ∎

While the previous result is a little intense, the punchline of the discussion is thankfully much cleaner.

Corollary 5.24 (Cayley–Hamilton). *A linear map satisfies its characteristic polynomial.*

Proof. Let $\mathbf{v} \in V$ and consider the T-cyclic subspace $W = \langle \mathbf{v} \rangle$ generated by $\mathbf{v}$. By Theorem 5.19, the characteristic polynomial $p_W(t)$ of the restriction T_W satisfies

$$p(t) = q_W(t)p_W(t)$$

for some polynomial $q_W(t)$. However, Theorem 5.23 part 3 says that $p_W(\mathrm{T})(\mathbf{v}) = \mathbf{0}$, whence

$$p(\mathrm{T})(\mathbf{v}) = \mathbf{0}$$

Since we may apply this reasoning to *any* $\mathbf{v} \in V$, we conclude that $p(\mathrm{T}) \equiv 0$ is the *zero function.* ■

The Cayley–Hamilton Theorem is used extensively to develop the idea of diagonalizability in inner product spaces and in the discussion of Jordan canonical forms. We will simply apply it to the calculation of inverses and large powers of a linear map.

Examples 5.25. 1. (Example 5.2) $A = \left(\begin{smallmatrix} 2 & 1 \\ 3 & 4 \end{smallmatrix}\right)$ has $p(t) = t^2 - 6t + 5$ and we confirm:

$$A^2 - 6A = \begin{pmatrix} 7 & 6 \\ 18 & 19 \end{pmatrix} - 6\begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} = -5I$$

It may seem like a strange thing to do for this matrix, but the characteristic equation can be used to calculate the inverse of A:

$$A^2 - 6A + 5I = 0 \implies A(A - 6I) = -5I \implies A^{-1} = \frac{1}{5}(6I - A) = \frac{1}{5}\begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix}$$

2. (Example 5.7.3) We use the Cayley–Hamilton Theorem to compute T^4 when

$$A = [\mathrm{T}]_\epsilon = \begin{pmatrix} 2 & -1 & \frac{8}{3} \\ 0 & 1 & -6 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{with} \quad p(t) = (2-t)^2(1-t) = 4 - 8t + 5t^2 - t^3$$

By Cayley–Hamilton,

$$\begin{aligned} \mathrm{T}^4 &= \mathrm{T} \circ \mathrm{T}^3 = \mathrm{T} \circ (5\mathrm{T}^2 - 8\mathrm{T} + 4I) = 5\mathrm{T}^3 - 8\mathrm{T}^2 + 4\mathrm{T} = 5(5\mathrm{T}^2 - 8\mathrm{T} + 4\mathrm{I}) - 8\mathrm{T}^2 + 4\mathrm{T} \\ &= 17\mathrm{T}^2 - 36\mathrm{T} + 20\mathrm{I} \end{aligned}$$

We can easily(!) compute the matrix:

$$[\mathrm{T}^4]_\epsilon = 17A^2 - 36A + 20I = 17\begin{pmatrix} 4 & -3 & \frac{50}{3} \\ 0 & 1 & -18 \\ 0 & 0 & 4 \end{pmatrix} - 36\begin{pmatrix} 2 & -1 & \frac{8}{3} \\ 0 & 1 & -6 \\ 0 & 0 & 2 \end{pmatrix} + 20\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 16 & -15 & \frac{562}{3} \\ 0 & 1 & -90 \\ 0 & 0 & 16 \end{pmatrix}$$

It follows, for example, that

$$\mathrm{T}^4(35 - 3x^2) = 35 \cdot 16 - 3\left(\frac{562}{3} - 90x + 16x^2\right) = -2 + 270x - 48x^2$$

3. The linear map $T \in \mathcal{L}(P_2(\mathbb{R}))$ defined by $T(f(x)) = f(x) + (x-1)f'(x)$ has characteristic polynomial

$$p(t) = (1-t)(2-t)(3-t) = -t^3 + 6t^2 - 11t + 6$$

as is easily seen by computing the matrix of T with respect to the standard basis $\{1, x, x^2\}$. By Cayley–Hamilton, we conclude that $T^3 = 6T^2 - 11T + 6I$. You can also check this explicitly, after first computing

$$T^2(f)(x) = f(x) + 3(x-1)f'(x) + (x-1)^2 f''(x)$$
$$T^3(f)(x) = f(x) + 7(x-1)f'(x) + 6(x-1)^2 f''(x)$$

We can also apply Cayley–Hamilton to find the inverse of T:

$$I = \frac{1}{6}(T^3 - 6T^2 + 11T) \implies T^{-1} = \frac{1}{6}(T^2 - 6T + 11I)$$
$$\implies T^{-1}(f)(x) = f(x) - \frac{1}{2}(x-1)f'(x) + \frac{1}{6}(x-1)^2 f''(x)$$

Warning! This is only the inverse of T viewed as a linear transformation of $P_2(\mathbb{R})$! If we change the vector space, the formula for the inverse will also change…

Exercises 5.3 1. Find a basis for the T-cyclic subspace $\langle \mathbf{v} \rangle$ of the given linear map:

(a) $T\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} a+b \\ b-c \\ a+c \\ a+d \end{pmatrix}$ on $\mathbb{R}^4$, where $\mathbf{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

(b) $T(f)(x) = f''(x)$ on $P_3(\mathbb{R})$, where $\mathbf{v} = x^3$

(c) $T(f)(x) = f''(x) + f(x)$ on $\text{Span}\{1, \sin x, \cos x, x\sin x, x\cos x\}$ where $\mathbf{v} = 1 + x\sin x$.

2. If $A = \begin{pmatrix} 4 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$, find *three distinct* invariant subspaces $W \le \mathbb{R}^4$ such that $\dim W = 3$.
(*Hint: What is* $A\mathbf{e}_2$*?*)

3. Let $T \in \mathcal{L}(V)$ and $\mathbf{v} \in V$. Prove that $\dim \langle \mathbf{v} \rangle = 1 \iff \mathbf{v}$ is an eigenvector of T.

4. We earlier remarked that the T-cyclic subspace $\langle \mathbf{v} \rangle$ is the smallest T-invariant subspace of V containing $\mathbf{v}$. To flesh this out, prove the following explicitly:

   (a) $\langle \mathbf{v} \rangle$ is T-invariant.

   (b) If $W \le V$ is T-invariant and $\mathbf{v} \in W$, then $\langle \mathbf{v} \rangle \le W$.

5. Consider the linear map $L_A : \mathbb{R}^3 \to \mathbb{R}^3$ where $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & 0 & 2 \end{pmatrix}$

   (a) Find the L_A-cyclic subspace generated by each $\mathbf{v} \in \mathbb{R}^3$. In particular, prove that $\langle \mathbf{v} \rangle = \mathbb{R}^3 \iff ac \neq 0$.
   (*Hint: first compute* $\det(\mathbf{v}\ A\mathbf{v}\ A^2\mathbf{v})$ *for any* $\mathbf{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$)

   (b) Check that the Cayley–Hamilton Theorem is satisfied for L_A.

6. Let $\mathrm{T}(f)(x) = f'(x) + \frac{1}{x}\int_0^x f(t)\,\mathrm{d}t$ be a linear map $\mathrm{T} \in \mathcal{L}(P_2(\mathbb{R}))$.

   (a) Find the characteristic polynomial of T and identify its eigenspaces. Is it diagonalizable?

   (b) Find $a, b, c \in \mathbb{R}$ such that $\mathrm{T}^3 = a\mathrm{T}^2 + b\mathrm{T} + c\mathrm{I}$.

   (c) What are $\dim \mathcal{L}(P_2(\mathbb{R}))$ and $\dim \operatorname{Span}\{\mathrm{T}^k : k \in \mathbb{N}_0\}$? Explain.

7. Recall Exercise 5.25.3. Find an explicit expression for $\mathrm{T}^{-1}(f)(x)$ (i.e. using derivatives!) when T is viewed as a linear transformation of $P_1(\mathbb{R})$.

8. For any matrix $A \in M_n(\mathbb{F})$, prove that

$$\dim \operatorname{Span}\{I, A, A^2, \ldots\} \leq n$$

9. Suppose $A \in M_n(\mathbb{F})$ has characteristic polynomial

$$p(t) = (-1)^n t^n + c_{n-1}t^{n-1} + \cdots + c_0$$

   (a) Prove that if A is invertible, then

$$A^{-1} = -\frac{1}{c_0}\Big((-1)^n A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1 I\Big)$$

   (b) Use this to find the inverse of T in Exercise 6.

   (c) If A is upper-triangular and invertible, prove that A^{-1} is also upper-triangular.