MAS220 ALGEBRA, SEMESTER 2

Prof. Neil Dummigan

University of Sheffield

Contents

1. VECTOR SPACES AND LINEAR MAPS

1.1. INTRODUCTION. .

In Semester 1 we saw groups and rings. A group is a set with a single binary operation, satisfying a certain list of axioms. We found that there are many natural examples of groups, making it seem to be a good definition, and explored the ways in which different groups can be related to each other (isomorphisms, homomorphisms, subgroups, quotient groups). We saw arithmetical examples such as $\mathbb{Z}$ and $\mathbb{R}^\times$, where the operation is addition or multiplication (respectively), and non-arithmetical examples such as S_n and O_2 where the operation is composition of bijections. We got a hint of how groups arising as symmetry groups can be applied to chemistry and physics.

A ring is a set with two binary operations, satisfying a certain list of axioms. The basic example is $\mathbb{Z}$ with the operations of addition and multiplication, and a ring can be thought of as some kind of generalised number system. We saw several exotic examples, some familiar, others new, including the complex numbers, Hamilton's quaternions, polynomial rings, matrix rings, the Gaussian integers and the finite field with 4 elements. Applications to computer graphics, flight simulation, QR codes, Fermat's two-square theorem and the board-game solitaire were mentioned, in more or less detail.

It is often useful to consider not just single numbers (or number-like elements), but ordered n-tuples of such things. For example, if you categorise your expenditure (in pounds) into accommodation, food, drink, transport, electricity, gas, phone, entertainment, clothes, other, then each month you are looking at an ordered 10-tuple of real numbers. In a digital image, the intensity of a pixel might be recorded as one of 256 levels, represented by an 8-bit number in binary. If a bit (0 or 1) is identified with the corresponding element $\overline{0}$ or $\overline{1}$ in $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$, then a 500×480 pixel image may be represented by an ordered n-tuple of elements of $\mathbb{F}_2$, where $n = 8 \times 500 \times 480 = 1,920,000$.

In both these examples we are selecting (for some n) from a set of ordered n-tuples of elements of some **field** F (a commutative ring in which every non-zero element has a multiplicative inverse). Such a set, with what will turn out to be the "right" operations, will be an example of a vector space.

1.2. VECTOR SPACES.

Definition 1.2.1. *Let F be a field (of "scalars"). A non-empty set V (of "vectors"), with two operations*

$+ : V \times V \to V$ ("vector addition") and

*$\cdot : F \times V \to V$ ("scalar multiplication") is an F-**vector space** if*

V1 $(V,+)$ *is an abelian group (let* $\underline{0}$ *denote the neutral element);*

V2 $\lambda(u+v)=\lambda u+\lambda v, \ \forall\,\lambda\in F,\, u,v\in V$*;*

V3 $(\lambda+\mu)v=\lambda v+\mu v, \ \forall\,\lambda,\mu\in F,\, v\in V$*;*

V4 $(\lambda\mu)v=\lambda(\mu v), \ \forall\,\lambda,\mu\in F,\, v\in V$*;*

V5 $1v=v, \ \forall\, v\in V$*.*

Example 1.2.2. *Let* F *be a field,* n *any positive integer. Then* $F^n:=\left\{\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix} : x_1,x_2,\ldots,x_n\in F\right\}$ *is an* F*-vector space, with addition and scalar multiplication defined by*

$$\begin{pmatrix}u_1\\u_2\\\vdots\\u_n\end{pmatrix}+\begin{pmatrix}v_1\\v_2\\\vdots\\v_n\end{pmatrix}:=\begin{pmatrix}u_1+v_1\\u_2+v_2\\\vdots\\u_n+v_n\end{pmatrix}, \qquad \lambda\begin{pmatrix}u_1\\u_2\\\vdots\\u_n\end{pmatrix}:=\begin{pmatrix}\lambda u_1\\\lambda u_2\\\vdots\\\lambda u_n\end{pmatrix}.$$

We have $\underline{0}=\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}$ *and* $-\begin{pmatrix}v_1\\v_2\\\vdots\\v_n\end{pmatrix}=\begin{pmatrix}-v_1\\-v_2\\\vdots\\-v_n\end{pmatrix}$.

Example 1.2.3. *Let* F *be a field,* n *any positive integer. Then* $F_n:=\{(x_1,x_2,\ldots,x_n) : x_1,x_2,\ldots,x_n\in F\}$ *is an* F*-vector space, with addition and scalar multiplication defined by*

$$(a_1,\ldots,a_n)+(b_1,\ldots,b_n)=(a_1+b_1,\ldots,a_n+b_n), \quad \lambda(a_1,\ldots,a_n)=(\lambda a_1,\ldots,\lambda a_n).$$

We have $\underline{0}=(0,\ldots,0)$ *and* $-(a_1,\ldots,a_n)=(-a_1,\ldots,-a_n)$.

These two examples are really the same thing written two different ways, but it is convenient to distinguish between them. In either case, it is easy to see how each axiom follows from the corresponding axiom for the ring F, applied component-by-component.

Example 1.2.4. *Special case* $n=1$*;* F *is an* F*-vector space.*

Example 1.2.5. *The set* $\mathbb{R}[x]$*, of polynomials in one variable with real coefficients, is an* $\mathbb{R}$*-vector space. Given any* $f,g\in\mathbb{R}[x]$*, we add them the usual way. Though we can also multiply together any two such polynomials to view* $\mathbb{R}[x]$ *as a ring, if we look only at the case that one of them is a constant* $\lambda\in\mathbb{R}$*, that is scalar multiplication. In other words, we are restricting the multiplication operation* $\cdot:\mathbb{R}[x]\times\mathbb{R}[x]\to\mathbb{R}[x]$ *to* $\cdot:\mathbb{R}\times\mathbb{R}[x]\to\mathbb{R}[x]$*.*

Example 1.2.6. *Taking $F = \mathbb{C}$ in Example 1.2.4, $\mathbb{C}$ is a $\mathbb{C}$-vector space. But restricting $\cdot : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ to $\cdot : \mathbb{R} \times \mathbb{C} \to \mathbb{C}$, we may view it as an $\mathbb{R}$-vector space.*

1.3. SOME SIMPLE CONSEQUENCES OF THE AXIOMS.

Proposition 1.3.1. *Let V be an F-vector space.*

(1) $\lambda\,\underline{0} = \underline{0} \ \ \forall \lambda \in F$.

(2) $0\,v = \underline{0} \ \ \forall v \in V$.

(3) $-(\lambda\,v) = (-\lambda)v \ \ \forall \lambda \in F, v \in V$.

(4) $-v = (-1)v \ \ \forall v \in V$.

Proof. (1)

$$\lambda\,\underline{0} \overset{V1}{=} \lambda(\underline{0} + \underline{0}) \overset{V2}{=} \lambda\,\underline{0} + \lambda\,\underline{0}.$$

Adding $-(\lambda\,\underline{0})$ to both sides, and using V1, $\underline{0} = \lambda\,\underline{0}$, as required.

(2)

$$0\,v \overset{R1}{=} (0+0)v \overset{V3}{=} 0\,v + 0\,v.$$

Adding $-(0\,v)$ to both sides, and using V1, $\underline{0} = 0\,v$, as required.

(3)

$$(-\lambda)v + \lambda v \overset{V3}{=} ((-\lambda) + \lambda)v = 0\,v \overset{(2)}{=} \underline{0}.$$

Adding $-(\lambda\,v)$ to both sides, and using V1, $(-\lambda)v = -(\lambda\,v)$, as required.

(4)

$$-v \overset{V5}{=} -(1\,v) \overset{(3),\lambda=1}{=} (-1)v.$$

□

1.4. SUBSPACES, SPANS. .

We looked at subgroups of groups and subrings of rings. Likewise we should look at subspaces of vector spaces.

Definition 1.4.1. *Let V be an F-vector space. An F-vector* **subspace** *of V is a subset U of V that is a vector space under the same operations $+$ and $\cdot$.*

Where F is understood, we might just call U a subspace. Elsewhere you may see it called a linear subspace. Sometimes it is written $U \leq V$.

Subspace criterion (Proof omitted.) Given an F-vector space V, and a subset $U \subseteq V$, to show that U is an F-vector subspace of V it suffices to show

SS1 $U \neq \emptyset$;

SS2 $u_1, u_2 \in U \implies u_1 + u_2 \in U$ (closure under addition);

SS3 $u \in U \implies \lambda\,u \in U \ \ \forall \lambda \in F$ (closure under scalar multiplication).

Remarks

(1) Letting $\lambda = -1$ in **SS3**, and using Proposition 1.3.1(4), shows that $u \in U \implies -u \in U$. Letting $\lambda = 0$ (with any $u \in U$, which exists by **SS1**), and using Proposition 1.3.1(2), we find that $\underline{0} \in U$. Both these facts also follow directly from the definition. Note that we would usually verify **SS1** by checking that $\underline{0} \in U$.

(2) **SS1** and **SS2** can be replaced by the single condition $\lambda u + \mu v \in U \ \ \forall \lambda, \mu \in F, \ u, v \in U$.

(3) $\{\underline{0}\}$ and V are always subspaces of V.

Definition 1.4.2. *Let V be an F-vector space, $S = \{v_1, v_2, \ldots, v_r\}$ a finite subset of V. The* **span** *of S,*

$$\operatorname{Span} S := \{\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_r v_r : \ \lambda_1, \ldots, \lambda_r \in F\}.$$

$\operatorname{Span} S$ is the set of all **linear combinations** of $v_1, \ldots, v_r$.

Theorem 1.4.3. $\operatorname{Span} S$ *is always a subspace of V.*

Proof. We check the subspace criterion. If $S \neq \emptyset$ (as the notation suggests) then $\operatorname{Span} S \neq \emptyset$, and $\operatorname{Span} S$ is clearly closed under addition and scalar multiplication. Even if $S = \emptyset$ then by convention $\operatorname{Span} S = \{\underline{0}\}$, which is still a subspace. □

Example 1.4.4. *In $\mathbb{R}^3$, which we can visualise as the set of points in 3-dimensional space, having chosen an origin and coordinate axes, the subspace* $\operatorname{Span}\{v\} = \{\lambda v : \ \lambda \in \mathbb{R}\}$ *is a line through the origin, in the direction of v (unless $v = \underline{0}$, in which case* $\operatorname{Span}\{v\} = \{\underline{0}\}$*).*

Example 1.4.5. *In $\mathbb{R}^3$, the subspace* $\operatorname{Span}\{v, w\} = \{\lambda v + \mu w : \ \lambda, \mu \in \mathbb{R}\}$ *is typically a plane through the origin, but if w is a multiple of v it will only be a line through the origin, unless $v = w = \underline{0}$, in which case it is just the origin, i.e. $\{\underline{0}\}$.*

Example 1.4.6. *Given a fixed natural number n, the set of polynomials (in one variable, with real coefficients) of degree less than or equal to n is* $\operatorname{Span}\{1, x, \ldots, x^n\}$*, so is an $\mathbb{R}$-vector subspace of $\mathbb{R}[x]$, sometimes denoted $\mathbb{R}[x]_{\leq n}$.*

Definition 1.4.7. *Let $A \in M_{m,n}(F)$ be an m-by-n matrix with entries in F. Let $r_1, \ldots, r_m \in F_n$ be the rows of A. Then the subspace* $\operatorname{Span}\{r_1, \ldots, r_m\}$ *of F_n is called the* **row space** *of A, denoted* $\operatorname{Row}(A)$.

Definition 1.4.8. *Given $A \in M_{m,n}(F)$, let $c_1, \ldots, c_n \in F^m$ be the columns of A. The subspace* $\operatorname{Span}\{c_1, \ldots, c_n\}$ *of F^m is called the* **column space** *of A, denoted* $\operatorname{Col}(A)$.

1.5. **ROW AND COLUMN REDUCTION.** .

We have just seen, given a matrix $A \in M_{m,n}(F)$, the row space $\mathrm{Row}(A)$ and the column space $\mathrm{Col}(A)$. Given a row vector $v \in F_n$, how do we tell whether or not it belongs to $\mathrm{Row}(A)$? Similarly, how do we recognise members of $\mathrm{Col}(A)$ among elements of F^m?

In MAS111 you saw three types of elementary row operation on a matrix $A \in M_{m,n}(F)$:

ERO1 $r_k \leftarrow r_k - \lambda r_l$ for some distinct $1 \leq k, l \leq m$ and $\lambda \in F$.

ERO2 $r_k \leftarrow \lambda r_k$ for some $1 \leq k \leq m$ and non-zero $\lambda \in F$.

ERO3 Swap r_k and r_l, for some distinct $1 \leq k, l \leq m$.

Given two matrices $A, B \in M_{m,n}(F)$, B is said to be **row-equivalent** to A if B can be obtained from A by a finite sequence of elementary row operations. Since each ERO can by inverted by another of the same type, it is easy to see that row-equivalence is an equivalence relation on $M_{m,n}(F)$. (For reflexivity, doing nothing counts as a sequence of EROs, or we could do an ERO2 with $\lambda = 1$.)

Similarly we have three types of elementary column operation:

ECO1 $c_k \leftarrow c_k - \lambda c_k$ for some distinct $1 \leq k, l \leq n$ and $\lambda \in F$.

ECO2 $c_k \leftarrow \lambda c_k$ for some $1 \leq k \leq n$ and non-zero $\lambda \in F$.

ECO3 Swap c_k and c_l, for some distinct $1 \leq k, l \leq n$.

Two matrices $A, B \in M_{m,n}(F)$ are said to be **column-equivalent** if one can be obtained from the other by a finite sequence of elementary column operations.

Proposition 1.5.1. *If $A, B \in M_{m,n}(F)$ are row-equivalent then* $\mathrm{Row}(B) = \mathrm{Row}(A)$.

Proof. It suffices to consider the case that B is obtained from A by a single ERO. Let $S = \{r_1, \ldots, r_m\}$ be the rows of A and $S' = \{r'_1, \ldots, r'_m\}$ be the rows of B.

(1) If B is obtained from A by the application of some ERO1, say $r'_k = r_k - \lambda r_l$ but $r'_i = r_i$ whenever $i \neq k$. Given any $\sum_{i=1}^m \alpha_i r'_i \in \mathrm{Row}(B)$, we can re-write it as $\alpha_k(r_k - \lambda r_l) + \sum_{1 \leq i \leq m, i \neq k} \alpha_i r_i \in \mathrm{Row}(A)$, so $\mathrm{Row}(B) \subseteq \mathrm{Row}(A)$. Similarly, substituting $r_k = r'_k + \lambda r'_l$ we see that $\mathrm{Row}(A) \subseteq \mathrm{Row}(B)$. Hence $\mathrm{Row}(B) = \mathrm{Row}(A)$.

(2) If B is obtained from A by the application of some ERO2, say $r'_k = \lambda r_k$ with $\lambda \neq 0$, but $r'_i = r_i$ whenever $i \neq k$. Any $\sum_{i=1}^m \alpha_i r'_i \in \mathrm{Row}(B)$ may now be rewritten as $\alpha_k \lambda r_k + \sum_{1 \leq i \leq m, i \neq k} \alpha_i r_i \in \mathrm{Row}(A)$, so $\mathrm{Row}(B) \subseteq \mathrm{Row}(A)$. Similarly, substituting $r_k = \lambda^{-1} r'_k$ we see that $\mathrm{Row}(A) \subseteq \mathrm{Row}(B)$. Hence $\mathrm{Row}(B) = \mathrm{Row}(A)$.

(3) If B is obtained from A by the application of some ERO3 then $S' = S$ so $\mathrm{Span}\, S' = \mathrm{Span}\, S$, i.e. $\mathrm{Row}(B) = \mathrm{Row}(A)$.

□

Definition 1.5.2. *A matrix $A \in M_{m,n}(F)$ is in* **reduced row echelon form** *(RREF) if*

(1) Any zero rows are at the bottom of the matrix.

(2) In any non-zero row, the leftmost non-zero entry is 1*. (This is called a* **pivot** *or* **leading** 1*.)*

(3) The leading 1 *in any row is strictly further to the right than the leading* 1 *in any row above it.*

(4) The column containing any leading 1 *has* 0 *in all the other entries (below* and *above).*

In MAS111 you saw how, given any matrix $A \in M_{m,n}(\mathbb{R})$, one can apply a sequence of EROs (equivalently multiply A on the left by a sequence of elementary matrices) to convert it to a matrix in reduced row echelon form. The same works with $\mathbb{R}$ replaced by any **field** F. (It is important to be able to divide by non-zero elements when scaling leading elements to 1.) We record this as follows.

Theorem 1.5.3. *Any matrix $A \in M_{m,n}(F)$ is row-equivalent to some B in reduced row echelon form.*

Now to check whether $v \in F_n$ belongs to $\mathrm{Row}(A)$, apply Theorem 1.5.3 to replace A by a row-equivalent B in RREF, which by Proposition 1.5.1 has the same row space as A. Then it suffices to check whether or not $v \in \mathrm{Row}(B)$. But this is easy since an element of $\mathrm{Row}(B)$ is of the form $v = \sum_{i=1}^{t} \alpha_i r_i$, where $r_1, \ldots, r_t$ are the non-zero rows of B. If the leading 1 in r_i is in column $j(i)$ (a position in which the other rows have a 0) then α_i can be read off from $v_{j(i)}$, then once the α_i are known, $v = \sum_{i=1}^{t} \alpha_i r_i$ determines what all the other entries of v (in non-pivotal positions) must be for $v \in \mathrm{Row}(B)$.

Similarly, we could show that column-equivalent matrices have the same column space, and check whether $v \in F^m$ belongs to $\mathrm{Col}(A)$ by replacing A by a column-equivalent B in "reduced column echelon form" (multiplying on the right by a suitable sequence of elementary matrices), then easily checking whether $v \in \mathrm{Col}(B)$.

1.6. LINEAR EQUATIONS AND NULL SPACES. .

A system of m linear equations in n unknowns is of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \vdots \quad \vdots \quad \vdots \quad \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m, \end{aligned}$$

where the a_{ij} and the b_i all belong to a given field F. In matrix form

$$A\mathbf{x} = \mathbf{b},$$

with $A = (a_{ij}) \in M_{m,n}(F)$, $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in F^m$, and we seek solutions $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in F^n$. We call A the **coefficient matrix** of the system of linear equations.

If $\mathbf{b} = \underline{0}$, the system of linear equations is said to be **homogeneous**. The set of solutions to a system of homogeneous linear equations has a special name.

Definition 1.6.1. *Given a field F, and a matrix $A \in M_{m,n}(F)$, the* **null space** *of A is defined by*

$$\mathrm{Null}(A) := \{\mathbf{x} \in F^n : \ A\mathbf{x} = \underline{0}\}.$$

The name is justified by the following theorem.

Theorem 1.6.2. *Given a field F, and a matrix $A \in M_{m,n}(F)$, $\mathrm{Null}(A)$ is an F-vector subspace of F^n.*

Proof. We check the subspace criterion. First, $\underline{0} \in \mathrm{Null}(A)$, since $A\underline{0} = \underline{0}$. Next, if $\mathbf{x} \in \mathrm{Null}(A)$ and $\lambda \in F$ then $A(\lambda\mathbf{x}) = \lambda(A\mathbf{x}) = \lambda\,\underline{0} = \underline{0}$, so $\lambda\mathbf{x} \in \mathrm{Null}(A)$. Finally, if $\mathbf{x}, \mathbf{y} \in \mathrm{Null}(A)$ then $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \underline{0} + \underline{0} = \underline{0}$, so $\mathbf{x}+\mathbf{y} \in \mathrm{Null}(A)$. □

We have used here familiar properties of matrix algebra. For example, to see in more detail why $A(\lambda\mathbf{x}) = \lambda(A\mathbf{x})$, for the i^{th} row it is

$$a_{i1}(\lambda x_1) + \ldots + a_{in}(\lambda x_n) = \lambda(a_{i1}x_1 + \ldots + a_{in}x_n).$$

Note that this uses the commutativity of multiplication in F.

The elementary row operations applied to a matrix A correspond to various ways of manipulating the homogeneous system of linear equations $A\mathbf{x} = \underline{0}$. An ERO1 corresponds to subtracting a multiple of one equation from another. An ERO2 corresponds to rescaling an equation, and an ERO3 corresponds to swapping two equations. The whole point of elementary row operations is to replace a system of linear equations by an equivalent but simpler system that can readily be solved.

Proposition 1.6.3. *If $A, B \in M_{m,n}(F)$ and A is row equivalent to B, then* $\mathrm{Null}(A) = \mathrm{Null}(B)$.

Proof. It suffices to show this when B is obtained from A by a single ERO, and it suffices to show that $\mathrm{Null}(A) \subseteq \mathrm{Null}(B)$, since the same argument applied to the ERO getting us back from B to A will show that $\mathrm{Null}(B) \subseteq \mathrm{Null}(A)$, hence that $\mathrm{Null}(B) = \mathrm{Null}(A)$. But if $\mathbf{x}$ satisfies the equations $A\mathbf{x} = \underline{0}$ then it also satisfies the equations $B\mathbf{x} = \underline{0}$, since we have just added a multiple of one equation to another, rescaled an equation, or swapped a pair of equations. Alternatively, if E is the elementary matrix that applies the row operation then $B = EA$, so

$$\mathbf{x} \in \mathrm{Null}(A) \implies A\mathbf{x} = \underline{0} \implies E(A\mathbf{x}) = \underline{0} \implies (EA)\mathbf{x} = \underline{0} \implies B\mathbf{x} = \underline{0} \implies \mathbf{x} \in \mathrm{Null}(B).$$

□

In other words, the system of equations $B\mathbf{x} = \underline{0}$ is equivalent to the original system $A\mathbf{x} = \underline{0}$. (For non-homogeneous systems of equations, we may similarly apply EROs to the *augmented* coefficient matrix $(A \mid \mathbf{b})$.) Now, to solve $A\mathbf{x} = \underline{0}$, we use Theorem 1.5.3 to replace it by the equivalent system $B\mathbf{x} = \underline{0}$, where B is in RREF. This is now easy to solve. The numbers $1 \leq j \leq n$ are divided into two classes: $j_1, j_2, \ldots, j_r$, the positions of the leading 1s, and $k_1, \ldots, k_{n-r}$ the rest. For $1 \leq j \leq n$ let $e_j \in F^m$ have a 1 in the j^{th} entry and 0 elsewhere. Then for $1 \leq t \leq r$ the j_t^{th} column of B is e_{j_t}, and the other $n - r$ columns are just whatever they are. Necessarily $r \leq m$ (as well as $r \leq n$), and if $r < m$ then the last $m - r$ rows of B are zero. We may think of the variables $x_{k_1}, \ldots, x_{k_{n-r}}$ as independent variables, with the dependent variables $x_{j_1}, \ldots, x_{j_r}$ determined by the equations, the i^{th} of which (for $1 \leq i \leq r$) is

$$x_{j_i} + \sum_{t=1}^{n-r} b_{ik_t} x_{k_t} = 0,$$

so

$$x_{j_i} = -\sum_{t=1}^{n-r} b_{ik_t} x_{k_t}.$$

We can now obtain the general solution by assigning arbitrary values $x_{k_t} = \lambda_t$ to the independent variables, and $x_{j_i} = -\sum_{t=1}^{n-r} b_{ik_t}\lambda_t$ for the dependent variables. This can be written in vector form as

$$\mathbf{x} = \sum_{t=1}^{n-r} \lambda_t v_t,$$

where $v_t \in F^n$ has a 1 in the k_t position, 0 in the k_s position for $1 \leq s \leq n - r$ with $s \neq t$, and $-b_{ik_t}$ in the j_i position for $1 \leq i \leq r$. There is a lot of impenetrable notation kicking around here. To follow what is going on you really need to work through a numerical example of solving a homogeneous system of linear equations by reducing the coefficient matrix to reduced row echelon form. Again the null space of the matrix is exactly the same thing as the set of solutions. We have arrived at

Theorem 1.6.4. *If the matrix $A \in M_{m,n}(F)$ is row-equivalent to a matrix B in RREF then* $\mathrm{Null}(A) = \mathrm{Span}\{v_1, \ldots, v_{n-r}\}$, *with the v_t as above.*

(Note that if $r = n$ then $\mathrm{Null}(A) = \mathrm{Span}\,\emptyset = \{\underline{0}\}$, but if $r < n$ then $A\mathbf{x} = \underline{0}$ has non-zero solutions.) Thus the set of all solutions to the system of linear equations $A\mathbf{x} = \underline{0}$ (which may be infinite) can be obtained from the finite subset $\{v_1, \ldots, v_{n-r}\}$ by taking linear combinations, and such a finite generating (or **spanning**) set may be found explicitly by reduction of A to RREF. (Recall also that r is the number of pivots, i.e. of non-zero rows, in RREF.) This is a remarkable fact about linear equations, which marks them out as something very special and accommodating. The solution of systems of linear equations is **the** motivating problem behind the theory of vector spaces, and in both the form of the general solution as a linear combination, and the EROs involved in obtaining that solution, we see the importance of the two operations of addition and scalar multiplication (in both F^n and F_n). This explains the choice of operations in the definition of a vector space.

If $\mathbf{x} \in \mathrm{Null}(A)$ then $a_1x_1 + \ldots + a_nx_n = 0$ for any $(a_1, \ldots, a_n) \in \mathrm{Row}(A)$, not just the rows of A. There is a whole *space* of linear equations satisfied by $\mathbf{x}$.

1.7. **LINEAR DEPENDENCE, BASES.** .

When we give a subspace of F^n as a null space or as a span, there may be some redundancy in our description.

Example 1.7.1. *In $\mathbb{R}^3$, let $U = \mathrm{Null}(A)$, where $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{pmatrix}$. The third equation $3x + 5y + 7z = 0$ is redundant because it is just the sum of the first two, $x + 2y + 3z = 0$ and $2x + 3y + 4z = 0$, so $U = \mathrm{Null}\left(\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix}\right)$.*

Example 1.7.2. *In $\mathbb{R}^3$, let $U = \mathrm{Span}\{v_1, v_2\}$, where $v_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 3 \\ 3 \\ 6 \end{pmatrix}$. The second vector v_2 is redundant because it is a multiple of the first one. So $U = \mathrm{Span}\{v_1\}$ is just a line through the origin, not a plane.*

Both these are examples of linear dependence.

Definition 1.7.3. *Let V be an F-vector space. A finite subset $\{v_1, v_2, \ldots, v_r\} \subseteq V$ is* **linearly dependent** *if there exist scalars $\alpha_1, \ldots, \alpha_r$, not all zero, such that*

$$\alpha_1v_1 + \alpha_2v_2 + \ldots + \alpha_rv_r = \underline{0}.$$

If $\alpha_t \neq 0$ then we may rearrange this as

$$v_t = -\alpha_t^{-1}\left(\sum_{1\leq i\leq r, i\neq t} \alpha_i v_i\right).$$

So equivalently a finite set of vectors is linearly dependent if and only if at least one of them can be written as a linear combination of the others.

Example 1.7.4. *In* $\mathbb{R}_3$, $\{(1,2,3),(2,3,4),(3,5,7)\}$ *is linearly dependent, since* $(3,5,7) = (1,2,3)+(2,3,4)$, *i.e.* $(1,2,3)+(2,3,4)+(-1)(3,5,7) = \underline{0}$.

Example 1.7.5. *In* $\mathbb{R}^3$, $\left\{\begin{pmatrix}1\\1\\2\end{pmatrix}, \begin{pmatrix}3\\3\\6\end{pmatrix}\right\}$ *is linearly dependent, since* $\begin{pmatrix}3\\3\\6\end{pmatrix} = 3\begin{pmatrix}1\\1\\2\end{pmatrix}$, *i.e.* $(-3)\begin{pmatrix}1\\1\\2\end{pmatrix} + \begin{pmatrix}3\\3\\6\end{pmatrix} = \underline{0}$.

Definition 1.7.6. *Let* V *be an* F*-vector space. A finite subset* $\{v_1, v_2, \ldots, v_r\} \subseteq V$ *is* **linearly independent** *if the only scalars* $\alpha_1, \ldots, \alpha_r \in F$ *such that* $\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_r v_r = \underline{0}$ *are* $\alpha_1 = \ldots = \alpha_r = 0$.

In other words, $\{v_1, v_2, \ldots, v_r\} \subseteq V$ is linearly independent if and only if it is not linearly dependent.

Definition 1.7.7. *Let* V *be an* F*-vector space. A finite subset* $S = \{v_1, \ldots, v_n\}$ *of* V *is a* **basis** *for* V *if every* $v \in V$ *can be expressed in a unique way as a linear combination of the elements of* S.

Example 1.7.8. $\{e_1, e_2, \ldots, e_n\}$ *is a basis for* F^n *(the "standard" basis). Each* $\mathbf{x} = \begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix} \in F^n$ *can be expressed uniquely as* $\sum_{i=1}^n x_i e_i$.

Example 1.7.9. *Special case* $n = 1$ *(cf. Example 1.2.4):* $\{1\}$ *is a basis for the* F*-vector space* F.

Example 1.7.10. $\mathbb{C}$ *is an* $\mathbb{R}$*-vector space with basis* $\{1, i\}$.

Example 1.7.11. $\mathbb{H}$ *(Hamilton's quaternions) is an* $\mathbb{R}$*-vector space with basis* $\{1, i, j, k\}$.

Example 1.7.12. *The finite field $\mathbb{F}_4$ is an $\mathbb{F}_2$-vector space with basis $\{1, \alpha\}$, where $\alpha^2+\alpha+\overline{1} = \overline{0}$ (cf. Semester 1, Example 3.6.12).*

Example 1.7.13. *(cf. Example 1.4.6.) Fixing a natural number n, $\{1, x, \ldots, x^n\}$ is a basis for the subspace $\mathbb{R}[x]_{\leq n}$ of $\mathbb{R}[x]$, each element of $\mathbb{R}[x]_{\leq n}$ being uniquely of the form $a_0 1 + a_1 x + \ldots + a_n x^n$.*

Proposition 1.7.14. *Let V be an F-vector space. A finite subset $S = \{v_1, \ldots, v_n\}$ of V is a basis for V if and only if*

(1) $V = \operatorname{Span} S$;

(2) S is linearly independent.

Proof. First suppose that the two conditions hold. We will show that S is then a basis for V. Given any $v \in V$, since $V = \operatorname{Span} S$, there exist $\lambda_1, \ldots, \lambda_n \in F$ such that $v = \sum_{i=1}^n \lambda_i v_i$. If this expression is not unique then also $v = \sum_{i=1}^n \mu_i v_i$, with $\mu_t \neq \lambda_t$ (for at least one t). But then $\sum_{i=1}^n (\lambda_i - \mu_i) v_i = \underline{0}$, and since $\lambda_t - \mu_t \neq 0$ this is a linear dependence relation, contrary to the linear independence of S.

Conversely, suppose that S is a basis for V. Since every $v \in V$ is a linear combination of the elements of S, $V = \operatorname{Span} S$. Now we deduce linear independence from the uniqueness of such linear combinations. Suppose to the contrary that S is linearly dependent, say $v_t = \sum_{1 \leq i \leq n, i \neq t} \alpha_i v_i$. Then $v_t = v_t$ and $v_t = \sum_{1 \leq i \leq n, i \neq t} \alpha_i v_i$ are two different expressions for v_t as a linear combination of elements of S, contrary to S being a basis. Therefore S must be linearly independent. □

Lemma 1.7.15. *Let V be an F-vector space, $S = \{v_1, \ldots, v_r\}$ a finite subset, U the subspace $\operatorname{Span} S$. If S is linearly dependent, say v_t is a linear combination of the others, then $U = \operatorname{Span}(S - \{v_t\})$.*

Proof. Given any $u = \sum_{i=1}^r \alpha_i v_i \in \operatorname{Span} S$, we may substitute for v_t in terms of the others, to rewrite u as a linear combination of the elements of $S - \{v_t\}$. □

Proposition 1.7.16. *Let V be an F-vector space, and suppose that $V = \operatorname{Span} S$ with S finite. Then V has a basis contained in S.*

Proof. If S is linearly independent then S is already a basis for V, by Proposition 1.7.14. If S is not linearly independent then by Lemma 1.7.15 we have $V = \operatorname{Span} S'$ for a subset S' of S with one element fewer. If S' is linearly independent then S' is a basis for V. If not, discard another element. Keep doing this, and eventually before we run out of elements we must arrive at a subset of S that still spans V but is also linearly independent. □

Example 1.7.17. *(cf. Example 1.7.1.) If* $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{pmatrix} \in M_3(\mathbb{R})$ *then* $\{(1,2,3),(2,3,4),(3,5,7)\}$ *is a finite spanning set for* $\mathrm{Row}(A)$*, and discarding* $(3,5,7)$ *produces a basis* $\{(1,2,3),(2,3,4)\}$ *for* $\mathrm{Row}(A)$*.*

Example 1.7.18. *In* $\mathbb{R}^3$*, let* $U = \mathrm{Span}\{v_1, v_2\}$*, where* $v_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ *and* $v_2 = \begin{pmatrix} 3 \\ 3 \\ 6 \end{pmatrix}$*. Then* $\{v_1, v_2\}$ *is a finite spanning set for* U*, and we can discard* v_2 *to get a basis* $\{v_1\}$ *for* U*.*

Proposition 1.7.19. *Given* $A \in M_{m,n}(F)$*, row-equivalent to* B *in RREF, the set* $S = \{r_1, r_2, \ldots, r_r\}$ *of non-zero rows of* B *is a basis for* $\mathrm{Row}(A)$*.*

Proof. By Proposition 1.5.1, $\mathrm{Row}(A) = \mathrm{Row}(B)$, and since (almost by definition) $\mathrm{Row}(B) = \mathrm{Span}\, S$, we have $\mathrm{Row}(A) = \mathrm{Span}\, S$. By Proposition 1.7.14, it remains to show that S is linearly independent. But if $\alpha_1 r_1 + \alpha_2 r_2 + \ldots + \alpha_r r_r = \underline{0}$ then for each $1 \leq i \leq r$, if r_i has a leading 1 in the $j(i)$ position (where all the other r_t have a 0), reading off the $j(i)$ entry on both sides of $\alpha_1 r_1 + \alpha_2 r_2 + \ldots + \alpha_r r_r = \underline{0}$ shows that $\alpha_i = 0$, as required. □

Proposition 1.7.20. *Given* $A \in M_{m,n}(F)$*, the spanning set* $S = \{v_1, v_2, \ldots, v_{n-r}\}$ *for* $\mathrm{Null}(A)$ *given by Theorem 1.6.4 is a basis.*

Proof. Since Theorem 1.6.4 already gives us $\mathrm{Null}(A) = \mathrm{Span}\, S$, it remains (by Proposition 1.7.14) to show that S is linearly independent. But if $\alpha_1 v_1 + \ldots + \alpha_{n-r} v_{n-r} = \underline{0}$ then for each $1 \leq t \leq n-r$, since v_t has a 1 in the k_t position, where the other v_s have a 0, we can read off from the k_t entry of both sides of $\alpha_1 v_1 + \ldots + \alpha_{n-r} v_{n-r} = \underline{0}$ that $\alpha_t = 0$, as required. □

Any subspace of F^n may be described in one of two ways:

(1) **implicitly**, by a set of equations $A\mathbf{x} = \underline{0}$, i.e. as a null space;
(2) **explicitly**, by a parametrisation $\mathbf{x} = \sum_{t=1}^{n-r} \lambda_t v_t$, i.e. in the form $\mathrm{Span}\{v_1, \ldots, v_{n-r}\}$.

1.8. **LINEAR MAPS.** .

Recall that given two groups G and H, a group homomorphism between them is a map $f : G \to H$ such that

$$f(ab) = f(a)f(b)\ \forall a, b \in G.$$

Given two rings R and S, a ring homomorphism between them is a map $f : R \to S$ such that

(1) $f(a+b) = f(a) + f(b)\ \ \forall a, b \in R$;
(2) $f(ab) = f(a)f(b)\ \ \forall a, b \in R$;

(3) $f(1_R) = 1_S$.

Similarly, given two F-vector spaces V and W, we should define a notion of "F-vector space homomorphism" between them. But that is not what we will call it.

Definition 1.8.1. *Let F be a field and V, W be F-vector spaces. A map $\ell : V \to W$ is said to be an* **F-linear map** *(or just "linear" if F is understood) if*

(1) $\ell(u+v) = \ell(u) + \ell(v) \ \ \forall u, v \in V$;
(2) $\ell(\lambda v) = \lambda \ell(v) \qquad \forall \lambda \in F,\ v \in V$.

An equivalent single condition is that $\ell(\lambda u + \mu v) = \lambda\ell(u) + \mu\ell(v) \ \ \forall \lambda, \mu \in F, u, v \in V$. In fact, for any linear combination, repeatedly applying the conditions shows that if $\ell : V \to W$ is linear, if $v_1, \ldots, v_r \in V$ and $\lambda_1, \ldots, \lambda_r \in F$ then $\ell\left(\sum_{i=1}^r \lambda_i v_i\right) = \sum_{i=1}^r \lambda_i \ell(v_i)$.

Note that (1) says that $\ell : V \to W$ is a homomorphism of additive groups, from which follows $\ell(\underline{0}_V) = \underline{0}_W$, and $\ell(-v) = -\ell(v) \ \ \forall v \in V$.

Let us now see the reason for the name.

Proposition 1.8.2. *A map $\ell : F^n \to F$ is F-linear if and only if it is of the form $\ell_{\mathbf{a}}$ for some $\mathbf{a} = (a_1, \ldots, a_n) \in F_n$, where*

$$\ell_{\mathbf{a}}(\mathbf{x}) := \mathbf{a}\mathbf{x} = a_1x_1 + a_2x_2 + \ldots + a_nx_n \ \ \forall \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in F^n.$$

Proof. First, given a linear map $\ell : F^n \to F$, $\ell(\mathbf{x}) = \ell\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n x_i \ell(e_i)$, where $\{e_1, \ldots, e_n\}$ is the standard basis for F^n (cf. Example 1.7.8). Thus $\ell(\mathbf{x}) = \ell_{\mathbf{a}}(\mathbf{x})$, where $a_i = \ell(e_i)$ for all $1 \leq i \leq n$.

Conversely, to show that $\ell_{\mathbf{a}}$ is linear for any $\mathbf{a} \in F_n$,

$$\ell_{\mathbf{a}}(\mathbf{x}+\mathbf{y}) = \sum_{i=1}^n a_i(x_i + y_i) = \sum_{i=1}^n a_i x_i + \sum_{i=1}^n a_i y_i = \ell_{\mathbf{a}}(\mathbf{x}) + \ell_{\mathbf{a}}(\mathbf{y}),$$

and

$$\ell_{\mathbf{a}}(\lambda\mathbf{x}) = \sum_{i=1}^n a_i(\lambda x_i) = \lambda\left(\sum_{i=1}^n a_i x_i\right) = \lambda\ell_{\mathbf{a}}(\mathbf{x}).$$

□

More generally

Theorem 1.8.3. *A map $\ell : F^n \to F^m$ is F-linear if and only if it is of the form ℓ_A for some matrix $A \in M_{m,n}(F)$, where*

$$\ell_A(\mathbf{x}) := A\mathbf{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{pmatrix}.$$

Proof. First, given a linear map $\ell : F^n \to F^m$, $\ell(\mathbf{x}) = \ell\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n x_i \ell(e_i)$, where $\{e_1, \ldots, e_n\}$ is the standard basis for F^n. Thus $\ell(\mathbf{x}) = \ell_A(\mathbf{x})$, where $A = (\ell(e_1) \mid \ell(e_2) \mid \cdots \mid \ell(e_n))$ for all $1 \leq i \leq n$.

Conversely, to show that ℓ_A is linear for any $A \in M_{m,n}(F)$,

$$\ell_A(\mathbf{x}+\mathbf{y}) = A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \ell_A(\mathbf{x}) + \ell_A(\mathbf{y}),$$

and

$$\ell_A(\lambda\mathbf{x}) = A(\lambda\mathbf{x}) = \lambda(A\mathbf{x}) = \lambda\ell_A(\mathbf{x}).$$

□

Example 1.8.4. *If $A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ then $\ell_A : \mathbb{R}^2 \to \mathbb{R}^2$ is rot_α, and if $B = \begin{pmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{pmatrix}$ then ℓ_B is ref_β (cf. Semester 1, Section 1.9).*

Example 1.8.5. *If $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ then ℓ_A is a shear fixing e_1 but taking e_2 to $e_1 + e_2$.*

Let us now look at some examples of linear maps on other vector spaces.

Example 1.8.6. *The map $z \to \overline{z}$ from $\mathbb{C}$ to itself is $\mathbb{R}$-linear (but not $\mathbb{C}$-linear), because $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ and $\overline{\lambda\, z} = \overline{\lambda}\ \overline{z}$, which equals $\lambda\,\overline{z}$ for $\lambda \in \mathbb{R}$ but not for other $\lambda \in \mathbb{C}$.*

Example 1.8.7. *The differentiation map $\frac{d}{dx} : \mathbb{R}[x] \to \mathbb{R}[x]$ is $\mathbb{R}$-linear, since $\frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx}$ and $\frac{d}{dx}(\lambda\, f) = \lambda\, \frac{df}{dx}$.*

Proposition 1.8.8. *Let F be a field, U, V, W be F-vector spaces, and $f : U \to V$, $g : V \to W$ be F-linear maps. Then the composition $g \circ f : U \to W$ is also F-linear.*

Proof. Given $u_1, u_2 \in U$,

$$(g\circ f)(u_1+u_2) = g(f(u_1+u_2)) \overset{f \text{ linear}}{=} g(f(u_1)+f(u_2)) \overset{g \text{ linear}}{=} g(f(u_1))+g(f(u_2)) = (g\circ f)(u_1)+(g\circ f)(u_2).$$

Given $u \in U, \lambda \in F$,

$$(g \circ f)(\lambda u) = g(f(\lambda u)) \overset{f \text{ linear}}{=} g(\lambda\, f(u)) \overset{g \text{ linear}}{=} \lambda\, g(f(u)) = \lambda\ (g \circ f)(u).$$

□

Remark. We could have proved similar propositions about compositions of group homomorphisms, or compositions of ring homomorphisms.

Proposition 1.8.9. *Given matrices $B \in M_{m,n}(F), A \in M_{k,m}(F)$, $\ell_A \circ \ell_B = \ell_{AB}$.*

Proof. For any $\mathbf{x} \in F^n$,

$$(\ell_A \circ \ell_B)(\mathbf{x}) = \ell_A(\ell_B(\mathbf{x})) = A(B\mathbf{x}) \overset{\text{matrix mult associative}}{=} (AB)\mathbf{x} = \ell_{AB}(\mathbf{x}).$$

□

Note that the proof of associativity of multiplication for *square* matrices, in Section 2.6 in Semester 1, works just as well in the generality we need here.

1.9. LINEAR ISOMORPHISM.

Definition 1.9.1. *If V, W are F-vector spaces, a map $\ell : V \to W$ is a* **linear isomorphism** *if it is a linear map and also a bijection.*

We write $V \simeq W$ and say that V is isomorphic to W.

Example 1.9.2. *If $A \in M_n(F)$, $\ell_A : F^n \to F^n$ is a linear isomorphism if and only if A is an invertible matrix (equivalent to $\det A \neq 0$.)*

Proposition 1.9.3. *If $\ell : V \to W$ is a linear isomorphism of F-vector spaces, then so is the inverse map $\ell^{-1} : W \to V$.*

Proof. Given $w_1, w_2 \in W$, let $v_1 = \ell^{-1}(w_1)$ and $v_2 = \ell^{-1}(w_2)$, so $\ell(v_1) = w_1$ and $\ell(v_2) = w_2$. Then

$$\ell^{-1}(w_1 + w_2) = \ell^{-1}(\ell(v_1) + \ell(v_2)) \overset{\text{linearity of } \ell}{=} \ell^{-1}(\ell(v_1 + v_2)) = v_1 + v_2 = \ell^{-1}(w_1) + \ell^{-1}(w_2),$$

and for $w \in W, \lambda \in F$, say $\ell^{-1}(w) = v$, so $\ell(v) = w$. Then

$$\ell^{-1}(\lambda w) = \ell^{-1}(\lambda(\ell(v))) \overset{\text{linearity of } \ell}{=} \ell^{-1}(\ell(\lambda v)) = \lambda v = \lambda\, \ell^{-1}(w).$$

□

Remark. We could have proved similar propositions about inverses of group isomorphisms, or inverses of ring isomorphisms.

Theorem 1.9.4. *If the F-vector space V has a basis $B = \{v_1, v_2, \ldots, v_n\}$ then the map $\iota_B : F^n \to V$ given by*

$$\iota_B\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = \sum_{i=1}^{n} x_i v_i$$

is a linear isomorphism.

Proof. By definition of a basis (Definition 1.7.7), ι_B is a bijection. It suffices then to show that it is linear. But

$$\iota_B(\mathbf{x}+\mathbf{y}) = \sum_{i=1}^{n}(x_i+y_i)v_i = \sum_{i=1}^{n} x_i v_i + \sum_{i=1}^{n} y_i v_i = \iota_B(\mathbf{x}) + \iota_B(\mathbf{y})$$

and

$$\iota_B(\lambda\mathbf{x}) = \sum_{i=1}^{n}(\lambda x_i)v_i = \lambda(\sum_{i=1}^{n} x_i v_i) = \lambda\,\iota_B(\mathbf{x}).$$

□

Definition 1.9.5. *Given $v \in V$, we may think of $\iota_B^{-1}(v) \in F^n$ as a coordinate vector for v, with respect to the choice of basis B. Suppose that V is an F-vector space with basis $B = \{v_1, \ldots, v_n\}$ and that W is an F-vector space with basis $C = \{w_1, \ldots, w_m\}$, and let $\ell : V \to W$ be an F-linear map. By Theorem 1.8.3, the map $\iota_C^{-1} \circ \ell \circ \iota_B : F^n \to F^m$, which is linear by Propositions 1.8.8 and 1.9.3, is of the form ℓ_A for some matrix $A \in M_{m,n}(F)$. This is called the matrix* **representing ℓ with respect to the bases B and C.**

In other words,

$$\ell\left(\sum_{j=1}^{n} x_j v_j\right) = \sum_{i=1}^{m} y_i w_i, \text{ where } \mathbf{y} = A\mathbf{x}.$$

For example, if $A \in M_{m,n}(F)$ then $\ell_A : F^n \to F^m$ (given by $\mathbf{x} \mapsto A\mathbf{x}$) is represented by A with respect to the standard bases of F^n and F^m.

1.10. **DIMENSION.** .

We would like to say that if an F-vector space V has a basis $\{v_1, \ldots, v_n\}$ with n elements (so that $V \simeq F^n$, by Theorem 1.9.4), then V is n-dimensional. But to know that this is well-defined, we must show that any two bases for V have the same number of elements.

Proposition 1.10.1. *If $\{v_1, \ldots, v_n\}$ is a linearly independent subset of F^m then $n \leq m$.*

Proof. Let $A = (v_1|v_2|\ldots|v_n) \in M_{m,n}(F)$. Let r be the number of non-zero rows in an RREF matrix row-equivalent to A. If $n > m$ then $r < n$ (since $r \leq m$, the number of rows of A). By the comment immediately following Theorem 1.6.4, there exists at least one *non-zero* solution $\mathbf{x}$ to $A\mathbf{x} = \underline{0}$, given that $r < n$. This is a linear dependence relation $\sum_{j=1}^{n} x_j v_i = \underline{0}$, so if $\{v_1, \ldots, v_n\}$ is linearly independent then we must have $n \leq m$. □

Theorem 1.10.2. *Let V be an F-vector space. If $B = \{v_1, \ldots, v_n\}$ and $C = \{w_1, \ldots, w_m\}$ are bases for V then $n = m$.*

Proof. Let $\iota_C : F^m \to V$ be the linear isomorphism appearing in Theorem 1.9.4. Since B is linearly independent (by Proposition 1.7.14), $\iota_C^{-1}(B)$ is a linearly independent subset of F^m, of size n. Hence $n \leq m$, by Proposition 1.10.1. Reversing the roles of B and C, also $m \leq n$, hence $n = m$, as required. □

This allows us to make the following definition.

Definition 1.10.3. *An F-vector space is said to be* **finite-dimensional** *if it has a finite spanning set. By Proposition 1.7.16 this is equivalent to having a finite basis. The* **dimension** *of V, denoted* $\dim V$ *or* $\dim_F(V)$*, is the number of elements in any basis for V.*

Looking at Examples 1.7.8, 1.7.10, 1.7.11, 1.7.12 and 1.7.13, we can now say that $\dim_{\mathbb{R}}(\mathbb{C}) = 2$, $\dim_{\mathbb{R}}(\mathbb{H}) = 4$, $\dim_{\mathbb{F}_2}(\mathbb{F}_4) = 2$ and $\dim_{\mathbb{R}}(\mathbb{R}[x]_{\leq n}) = n + 1$.

Proposition 1.10.4. *Let V be a finite-dimensional F-vector space, with* $\dim V = m$.

(1) If $S = \{v_1, \ldots, v_n\}$ is a linearly independent subset of V then $n \leq m$.

(2) If $n = m$ then $\{v_1, \ldots, v_n\}$ (still assumed linearly independent) is a basis of V.

Proof. (1) Let B be any basis for V, and $\iota_B : F^m \to V$ the isomorphism given by Theorem 1.9.4. Then $\iota_B^{-1}(S)$ is a linearly independent subset of F^m, of size n, so $n \leq m$ by Proposition 1.10.1.

(2) By Proposition 1.7.14, it remains to prove that $V = \operatorname{Span} S$. If not, $\exists v_{n+1} \in V$ with $v_{n+1} \notin \operatorname{Span} S$. If $\alpha_1 v_1 + \ldots + \alpha_n v_n + \alpha_{n+1} v_{n+1} = \underline{0}$ then $\alpha_{n+1} = 0$, otherwise $v_{n+1} = -\alpha_{n+1}^{-1} \sum_{i=1}^{n} \alpha_i v_i \in \operatorname{Span} S$. But now also $\alpha_1 = \ldots = \alpha_n = 0$, by linear independence of S. Hence $S \cup \{v_{n+1}\}$ is linearly independent, but by (1) it would be too big to be linearly independent. Hence it must be the case that $V = \operatorname{Span} S$.

□

Proposition 1.10.5. *If V is a finite-dimensional F-vector space and U is a subspace of V, then U is finite-dimensional, and* $\dim U \leq \dim V$*, with equality if and only if $U = V$.*

Proof. By Proposition 1.10.4(1), we may select a linearly independent subset S of U of maximal size, since that size is at most $\dim V$. Arguing as in the proof of Proposition 1.10.4(2), if $U \neq \operatorname{Span} S$ then we would be able to create an even larger linearly independent subset of U. So $U = \operatorname{Span} S$, therefore U is finite-dimensional. The proposition now follows immediately by applying Proposition 1.10.4 to a basis for U, considered as a linearly independent subset of V. □

By convention, $\emptyset$ is a basis for $\{\underline{0}\}$, and $\dim\{\underline{0}\} = 0$.

Proposition 1.10.6. *Let V be a finite-dimensional F-vector space, S a linearly independent subset of V. Then S can be extended to a basis of V.*

Proof. Arguing as in the proof of Proposition 1.10.4, if S does not already span V, it may be enlarged by one element to create a larger linearly independent subset. Keep repeating this as long as the subset does not span V, and since by Proposition 1.10.4 the size of a linearly independent subset is bounded above (by $\dim V$), we must eventually reach a linearly independent subset that *does* span V, i.e. a basis (cf. Proposition 1.7.14). □

Definition 1.10.7. *An F-vector space V is said to be* **infinite-dimensional** *if it is not finite-dimensional.*

Example 1.10.8. *$\mathbb{R}[x]$ is infinite-dimensional, by Proposition 1.10.5, since it contains subspaces $\mathbb{R}[x]_{\leq n}$ of arbitrarily large dimension.*

1.11. **RANK, NULLITY.** .

We return to the subject of linear equations, armed with the concept of dimension. The basic question is, given a homogeneous system of linear equations, what should be the dimension of the solution space? This is actually a question we have already answered. If we have m equations in n unknowns, $x_1, x_2, \ldots, x_n$, with coefficient matrix A (i.e. equations $A\mathbf{x} = \underline{0}$), we might think that each equation should cut down the dimension of the solution space by 1, starting from the n-dimensional space F^n of all possibilities available before we started imposing equations. But it may be the case that the equations are not independent, so instead we should subtract from n the "number of independent equations". Row reduction produces a set of r independent equations, equivalent to the original m equations, so we should expect dimension $n - r$ for the solution space. Indeed this is exactly what Proposition 1.7.20 tells us, since it exhibits a basis $\{v_1, \ldots, v_{n-r}\}$ for $\mathrm{Null}(A)$, so the dimension of $\mathrm{Null}(A)$ is $n - r$, where r is the number of non-zero rows in an RREF matrix B row-equivalent to A. Meanwhile, Proposition 1.7.19 gives us an interpretation of this number r as the dimension of the row space of A.

Definition 1.11.1. *Given $A \in M_{m,n}(F)$, the* **row rank** *of A is defined by* $\mathrm{RRank}(A) := \dim(\mathrm{Row}(A))$.

Definition 1.11.2. *Given $A \in M_{m,n}(F)$, the* **nullity** *of A is defined by* $\mathrm{Nullity}(A) := \dim(\mathrm{Null}(A))$.

These definitions allow us to express what we have arrived at as follows.

Proposition 1.11.3. *Given $A \in M_{m,n}(F)$,* $\mathrm{Nullity}(A) = n - \mathrm{RRank}(A)$.

Proposition 1.11.4. *Given $A \in M_n(F)$, A is invertible if and only if* $\mathrm{RRank}(A) = n$.

Proof. Recall from MAS111 the algorithm for finding the inverse of a square matrix A. One attempts to row reduce $(A|I)$ to $(I|A^{-1})$, where I is the n-by-n identity matrix. This succeeds if and only if reducing A to RREF produces I, which has n non-zero rows, i.e. $r = n$. □

Proposition 1.11.3 now tells us that if A is invertible then $\text{Nullity}(A) = n - n = 0$, i.e. $\text{Null}(A) = \{\underline{0}\}$, but we already know this. The equivalent set of equations coming from the row-equivalent matrix I is $x_1 = 0, x_2 = 0, \ldots, x_n = 0$. Looking at the same thing a different way, if A is invertible then

$$A\mathbf{x} = \underline{0} \implies A^{-1}A\mathbf{x} = \underline{0} \implies \mathbf{x} = \underline{0}.$$

Definition 1.11.5. *Given $A \in M_{m,n}(F)$, the* **column rank** *of A is defined by* $\text{CRank}(A) := \dim(\text{Col}(A))$.

Proposition 1.11.6. *If $A \in M_{m,n}(F)$ then* $\text{CRank}(A) = \text{RRank}(A)$.

Proof. Let B be a RREF matrix row-equivalent to A. We will prove the proposition in the three steps

(1) $\text{RRank}(A) = \text{RRank}(B)$;
(2) $\text{RRank}(B) = \text{CRank}(B)$;
(3) $\text{CRank}(B) = \text{CRank}(A)$.

(1) $\text{RRank}(A) = \text{RRank}(B)$ follows trivially from the fact that $\text{Row}(B) = \text{Row}(A)$, by Proposition 1.5.1.

(2) Let $r = \text{RRank}(B)$, i.e. the number of non-zero rows in B. In B, the columns containing the pivots form the subset $\{e_1, \ldots, e_r\}$ of the standard basis for F^m. These r columns are linearly independent. They also span $\text{Col}(A)$, since any column of B (and hence anything spanned by the columns of B) is of the form

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \sum_{j=1}^{r} \alpha_j e_j.$$

(Recall that the last $m-r$ rows of B are zero.) Hence $\{e_1, \ldots, e_r\}$ is a basis for $\text{Col}(B)$, so $\text{CRank}(B) = r = \text{RRank}(B)$.

(3) By Proposition 1.6.3, $\text{Null}(A) = \text{Null}(B)$, so $A\mathbf{x} = \underline{0} \iff B\mathbf{x} = \underline{0}$. This means that the columns of A and the columns of B satisfy exactly the same linear relations. The

columns $e_1, \ldots, e_r$ of B, in positions $j_1, \ldots, j_r$ say, form a basis for $\mathrm{Col}(B)$, as in the proof of step (2). It follows that if $A = (c_1|c_2|\ldots|c_n)$ then $\{c_{j_1}, \ldots, c_{j_r}\}$ is a basis for $\mathrm{Col}(A)$, in particular $\mathrm{CRank}(A) = r = \mathrm{CRank}(B)$. In a bit more detail, if $\lambda_1 c_{j_1} + \ldots + \lambda_r c_{j_r} = \underline{0}$ then, looking at the same relation among columns of B we get $\lambda_1 e_1 + \ldots + \lambda_r e_r = \underline{0}$, so $\lambda_1 = \ldots = \lambda_r = 0$, showing that $\{c_{j_1}, \ldots, c_{j_r}\}$ is linearly independent. Similarly, since any column of B is a linear combination of the $\{e_1, \ldots, e_r\}$, the corresponding column of A will be a linear combination (with the same coefficients) of the $\{c_{j_1}, \ldots, c_{j_r}\}$, so $\{c_{j_1}, \ldots, c_{j_r}\}$ spans $\mathrm{Col}(A)$.

□

Definition 1.11.7. *Given $A \in M_{m,n}(F)$, the* **rank** *of A is defined by* $\mathrm{rank}(A) = \mathrm{RRank}(A) = \mathrm{CRank}(A)$ *(they are the same by Proposition 1.11.6).*

Note that $\mathrm{RRank}(A) \leq m$ and $\mathrm{CRank}(A) \leq n$ (since $\mathrm{Row}(A)$ and $\mathrm{Col}(A)$ are subspaces of F_n and F^m respectively), so $\mathrm{rank}(A) \leq \min\{m, n\}$.

We may now reformulate Proposition 1.11.3 as

Theorem 1.11.8. *Given $A \in M_{m,n}(F)$,*

$$\mathrm{Nullity}(A) = n - \mathrm{rank}(A).$$

Example 1.11.9. *The subspace of $\mathbb{R}^3$ defined by the equation $x + y + z = 0$ has dimension $3 - 1 = 2$ (a plane through the origin), while the subspace defined by the equations*

$$x + y + z = 0$$

$$x + 2y + 3z = 0$$

has dimension $3 - 2 = 1$ (a line through the origin).

Example 1.11.10. *The subspace U of $\mathbb{F}_2^8$ defined by the equation $x_1 + x_2 + \ldots + x_8 = 0$ (equivalently $x_8 = x_1 + \ldots + x_7$) is 7-dimensional. The map $(x_1, x_2, \ldots, x_7)^t \mapsto (x_1, x_2, \ldots, x_7, x_1 + \ldots + x_7)$ is an isomorphism $\mathbb{F}_2^7 \simeq U$. This is the ASCII code with check digit.*

1.11.1. *Non-homogeneous equations revisited.* The significance of the row space $\mathrm{Row}(A)$ is that it is essentially the space of all linear equations implied by the equations $A\mathbf{x} = \underline{0}$. What about the column space? If we look at a general system of linear equations $A\mathbf{x} = \mathbf{b}$, with $\mathbf{b}$ not necessarily $\underline{0}$, it is not guaranteed that there is any solution. When we reduce $(A|\mathbf{b})$ to echelon form, there may be rows of zeros at the bottom of A that do not extend across the divide, resulting in contradictory equations of the form $0 = c$, with $c \in F$ non-zero. Looking at the equation $A\mathbf{x} = \mathbf{b}$ as $\mathbf{b} = \sum_{i=1}^{n} x_i c_i$, we see that $A\mathbf{x} = \mathbf{b}$ is solvable if and only if $\mathbf{b} \in \mathrm{Col}(A)$. If

it is, the solution found by row reduction has (in general) parameters in it, just like when we solve the homogeneous system of equations with coefficient matrix A. In fact

Proposition 1.11.11. *If $\mathbf{b} \in \mathrm{Col}(A)$ and $\mathbf{x} = \mathbf{x}_p$ is some fixed particular solution to $A\mathbf{x} = \mathbf{b}$, then the general solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{x}_h + \mathbf{x}_p$, where $\mathbf{x}_h$ is the general solution to the homogeneous system $A\mathbf{x} = \underline{0}$, i.e. $\mathbf{x}_h \in \mathrm{Null}(A)$.*

Proof. Given $A\mathbf{x}_p = \mathbf{b}$, we have

$$A\mathbf{x} = \mathbf{b} \iff A(\mathbf{x} - \mathbf{x}_p) = A\mathbf{x} - A\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \underline{0} \iff \mathbf{x} - \mathbf{x}_p \in \mathrm{Null}(A).$$

□

The dimension of the space of $\mathbf{b}$ for which $A\mathbf{x} = \mathbf{b}$ is solvable is the rank r (as column rank), while the number of parameters in the general solution is $n - r$ (with r as row rank), which goes up as r goes down, so the "fewer" $\mathbf{b}$ there are that get hit by $A\mathbf{x}$, the "more times" each such $\mathbf{b}$ gets hit. In the extreme case $r = n$ (possible only for A square, since $r \leq \min\{m, n\}$), $A\mathbf{x} = \mathbf{b}$ is solvable for every $\mathbf{b}$, and since $n - r = 0$ the solution is unique. We can also see this by considering that $r = n$ is the case that A is invertible, and the unique solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$.

1.12. SPACES OF FUNCTIONS. .

We begin by proving a very general theorem, from which we will deduce lots of consequences.

Theorem 1.12.1. *Let V be an F-vector space, and X any non-empty set whatsoever. Then*

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

is also an F-vector space, with operations (given $f, g \in \mathcal{F}(X, V),\ \lambda \in F$)

$$(f + g)(x) := f(x) + g(x) \quad \forall x \in X,$$

$$(\lambda\, f)(x) := \lambda\, f(x) \quad \forall x \in X.$$

Proof. We check the axioms one by one (cf. Definition 1.2.1). Since addition and scalar multiplication in $\mathcal{F}(X, V)$ are defined directly in terms of the corresponding operations in V, each axiom for $\mathcal{F}(X, V)$ follows from the corresponding axiom satisfied by V, which we are given is an F-vector space.

V1 We must show that $(\mathcal{F}(X, V), +)$ is an abelian group. We have a binary operation $+$ on $\mathcal{F}(X, V)$ defined above. It is commutative, since $\forall f, g \in \mathcal{F}(X, V),\ x \in X$,

$$(f + g)(x) \overset{\text{defn. of } +}{=} f(x) + g(x) \overset{\substack{\text{commutativity} \\ \text{of } + \text{ in } V}}{=} g(x) + f(x) = (g + f)(x),$$

so $f + g = g + f$. It remains to check the group axioms.

It is associative since $\forall f, g, h \in \mathcal{F}(X, V),\ x \in X$,

$$((f+g)+h)(x) \overset{\text{defn. of } +}{=} (f+g)(x)+h(x) = [f(x)+g(x)]+h(x) \overset{\substack{\text{associativity} \\ \text{in } V}}{=} f(x)+[g(x)+h(x)]$$

$$= f(x)+(g+h)(x) = (f+(g+h))(x),$$

so $(f+g)+h = f+(g+h)$. If we define $\underline{0}_{\mathcal{F}} \in \mathcal{F}(X, V)$ by $\underline{0}_{\mathcal{F}}(x) = \underline{0}_V\ \forall x \in X$ then $\underline{0}_{\mathcal{F}}$ is a neutral element in $\mathcal{F}(X, V)$, since for any $f \in \mathcal{F}(X, V)$ we have

$$(f+\underline{0}_{\mathcal{F}})(x) = f(x)+\underline{0}_{\mathcal{F}}(x) = f(x)+\underline{0}_V = f(x),$$

so $f+\underline{0}_{\mathcal{F}} = f$. Given $f \in \mathcal{F}(X, V)$, if we define $-f$ by $(-f)(x) := -f(x)\ \forall x \in X$ then $-f$ is an inverse for f, since $(f+(-f))(x) = f(x)+(-f)(x) = f(x)+(-(f(x))) = 0_V = \underline{0}_{\mathcal{F}}(x)\ \forall x \in X$, so $f+(-f) = \underline{0}_{\mathcal{F}}$.

V2 We must show that $\lambda(f+g) = \lambda f + \lambda g,\ \forall \lambda \in F,\ f, g \in \mathcal{F}(X, V)$. But for any $x \in X$,

$$(\lambda(f+g))(x) = \lambda((f+g)(x)) = \lambda(f(x)+g(x)) \overset{\text{V2 in } V}{=} \lambda(f(x))+\lambda(g(x)) = (\lambda f)(x)+(\lambda g)(x)$$

$$= (\lambda f+\lambda g)(x).$$

V3 We must show that $(\lambda+\mu)f = \lambda f+\mu f,\ \forall \lambda, \mu \in F,\ f \in \mathcal{F}(X, V)$. But for any $x \in X$,

$$((\lambda+\mu)f)(x) = (\lambda+\mu)(f(x)) \overset{\text{V3 in } V}{=} \lambda(f(x))+\mu(f(x)) = (\lambda f)(x)+(\mu f)(x) = (\lambda f+\mu f)(x).$$

V4 We must show that $(\lambda\mu)f = \lambda(\mu f),\ \forall \lambda, \mu \in F,\ f \in \mathcal{F}(X, V)$. But for any $x \in X$,

$$((\lambda\mu)f)(x) = (\lambda\mu)(f(x)) \overset{\text{V4 in } V}{=} \lambda(\mu(f(x))) = \lambda((\mu f)(x)) = (\lambda(\mu f))(x).$$

V5 We must show that $1f = f,\ \forall f \in \mathcal{F}(X, V)$. But for any $x \in X$,

$$(1f)(x) = 1(f(x)) \overset{\text{V5 in } V}{=} f(x).$$

□

We are often looking at the special case $V = F$. Choosing an ordered n-tuple of elements of F is equivalent to assigning an element of F to each of the natural numbers from 1 to n, so F^n may be thought of as $\mathcal{F}(X, F)$, where $X = \{1, \ldots, n\}$. (These F-vector spaces are isomorphic.) So $\mathcal{F}(X, F)$ generalises the idea of F^n to arrays of elements of F indexed by arbitrary sets. The indices are the inputs to functions, the elements of F the outputs.

1.12.1. *Some subspaces of the $\mathbb{R}$-vector space $\mathcal{F}(\mathbb{R},\mathbb{R})$.*

Example 1.12.2. $C(\mathbb{R},\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous on } \mathbb{R}\}$. *This is a subspace because*

(1) $0 : \mathbb{R} \to \mathbb{R}$ *is continuous;*

(2) if f and g are continuous then so is $f+g$;

(3) if f is continuous and $\lambda \in \mathbb{R}$ then λf is continuous.

These follow from things proved in MAS221 Analysis (likewise below).

Given a closed interval $[a,b]$, one can define $C([a,b],\mathbb{R})$ similarly (often abbreviated to $C[a,b]$).

Example 1.12.3. $C^1(\mathbb{R},\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuously differentiable on } \mathbb{R}\}$.

Example 1.12.4. $C^\infty(\mathbb{R},\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ has (continuous) derivatives of all orders on } \mathbb{R}\}$.

Example 1.12.5. $\mathbb{R}[x]$, *thought of as functions rather than formal expressions.*

In fact $\mathbb{R}[x] \subseteq C^\infty(\mathbb{R},\mathbb{R}) \subseteq C^1(\mathbb{R},\mathbb{R}) \subseteq C(\mathbb{R},\mathbb{R}) \subseteq \mathcal{F}(\mathbb{R},\mathbb{R})$. These are all infinite-dimensional spaces, since $\mathbb{R}[x]$ is (cf. Example 1.10.8). Identities such as $\cos(3x) = \cos^3 x - 3\cos x \sin^2 x$ may be thought of as linear dependence relations in the $\mathbb{R}$-vector space $C(\mathbb{R},\mathbb{R})$.

1.12.2. *Some linear maps on function spaces.*

Example 1.12.6. *Given $a \in \mathbb{R}$, define* $\mathrm{ev}_a : \mathcal{F}(\mathbb{R},\mathbb{R}) \to \mathbb{R}$ *by* $\mathrm{ev}_a(f) := f(a)$. *This is linear because*

(1) $\mathrm{ev}_a(f+g) = (f+g)(a) = f(a)+g(a) = \mathrm{ev}_a(f) + \mathrm{ev}_a(g)$;

(2) $\mathrm{ev}_a(\lambda f) = (\lambda f)(a) = \lambda(f(a)) = \lambda\,\mathrm{ev}_a(f)$.

This is analogous to a linear map $\mathbf{x} \mapsto x_i$ from $\mathbb{R}^n$ to $\mathbb{R}$.

Example 1.12.7. $\frac{d}{dx} : C^1(\mathbb{R},\mathbb{R}) \to C(\mathbb{R},\mathbb{R})$. *This is linear because*

(1) $(f+g)' = f' + g'$;

(2) $(\lambda f)' = \lambda(f')$.

Example 1.12.8. $\int : C([a,b],\mathbb{R}) \to \mathbb{R}$ *given by* $f \mapsto \int_a^b f(x)\,dx$. *This is linear because*

(1) $\int_a^b (f+g)(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$;

(2) $\int_a^b \lambda f(x)\,dx = \lambda \int_a^b f(x)\,dx$.

This is analogous to the linear map $\mathbf{x} \mapsto x_1 + x_2 + \ldots + x_n$ from $\mathbb{R}^n$ to $\mathbb{R}$. In fact, in numerical integration, a function is approximated by an element of $\mathbb{R}^n$ (for some large n) obtained by sampling the function at n points, and the integral is approximated by some linear combination of these sample values.

1.13. SPACES OF LINEAR MAPS, RINGS OF LINEAR OPERATORS.

Theorem 1.13.1. *Let V, W be F-vector spaces. Then $L(V,W) := \{f : V \to W \ : f \text{ is linear}\}$ is a subspace of $\mathcal{F}(V,W)$.*

Proof. (1) $\underline{0}_{\mathcal{F}} : V \to W$ (defined by $\underline{0}_{\mathcal{F}}(v) = \underline{0}_W \ \forall v \in V$) is linear, since

$$\underline{0}_{\mathcal{F}}(v_1 + v_2) = \underline{0}_W = \underline{0}_W + \underline{0}_W = \underline{0}_{\mathcal{F}}(v_1) + \underline{0}_{\mathcal{F}}(v_2) \ \forall v_1, v_2 \in V$$

and

$$\underline{0}_{\mathcal{F}}(\lambda\, v) = \underline{0}_W \overset{\text{Prop. } 1.3.1(1)}{=} \lambda\, \underline{0}_W = \lambda(\underline{0}_{\mathcal{F}}(v)) \ \forall v \in V, \lambda \in F.$$

(2) Suppose $f, g \in L(V,W)$, i.e. $f, g : V \to W$ are linear maps. We must show that $f + g : V \to W$ is also linear. But

$$(f+g)(v_1+v_2) = f(v_1+v_2) + g(v_1+v_2) \overset{\text{linearity of } f \text{ and } g}{=} [f(v_1) + f(v_2)] + [g(v_1) + g(v_2)]$$

$$= [f(v_1) + g(v_1)] + [f(v_2) + g(v_2)] = (f+g)(v_1) + (f+g)(v_2)$$

and

$$(f+g)(\lambda\, v) = f(\lambda\, v) + g(\lambda\, v) = \lambda(f(v)) + \lambda(g(v))$$

$$= \lambda(f(v) + g(v)) = \lambda\, (f+g)(v).$$

(3) Suppose $f \in L(V,W)$, i.e. $f : V \to W$ is linear, and that $\mu \in F$. We must show that $\mu\, f : V \to W$ is also linear. But

$$(\mu f)(v_1 + v_2) = \mu(f(v_1 + v_2)) \overset{\text{linearity of } f}{=} \mu(f(v_1) + f(v_2))$$

$$\overset{\text{V2 in } W}{=} \mu(f(v_1)) + \mu(f(v_2)) = (\mu\, f)(v_1) + (\mu\, f)(v_2)$$

and

$$(\mu\, f)(\lambda v) = \mu(f(\lambda v)) \overset{\text{linearity of } f}{=} \mu(\lambda(f(v)))$$

$$\overset{V4}{=} (\mu\lambda) f(v) = (\lambda\mu) f(v) \overset{V4}{=} \lambda(\mu(f(v))) = \lambda((\mu\, f)(v)).$$

□

Example 1.13.2. *Let $V = F^n, W = F^m$. Then $L(F^n, F^m)$ is an F-vector space. There is an isomorphism $M_{m,n}(F) \xrightarrow{\sim} L(F^n, F^m)$ given by $A \mapsto \ell_A$ (cf. Theorem 1.8.3)), where in $M_{m,n}(F)$ the operations are the usual addition and scalar multiplication of matrices. Note that $\dim M_{m,n}(F) = mn$, since $\{E_{ij} : 1 \le i \le m, 1 \le j \le n\}$ is a basis, where E_{ij} has 1 in the ij position, 0 elsewhere.*

Example 1.13.3. *Special case $m = 1$, $L(F^n, F) \simeq M_{1,n}(F) = F_n$. To $\mathbf{a} \in F_n$ we associate the linear function $\ell_{\mathbf{a}} \in L(F^n, F)$ (cf. Proposition 1.8.2).*

Example 1.13.4. *More generally, for any f-vector space V, the space $L(V,F)$ of linear functions on V, is called the* **dual space** *of V, denoted V^*. For example $(F^n)^* \simeq F_n$, $\int \in (C[a,b])^*$ and $\mathrm{ev}_a \in (\mathcal{F}(\mathbb{R},\mathbb{R}))^*$ (cf. Examples 1.12.6 and 1.12.8).*

Definition 1.13.5. *Let V be any F-vector space. An element of $L(V) := L(V,V)$ is called a* **linear operator** *on V.*

For example, $\frac{d}{dx}$ is a linear operator on $C^\infty(\mathbb{R},\mathbb{R})$.

Definition 1.13.6. *Suppose V is an F-vector space, and $\ell \in L(V)$ a linear operator on V. If $\ell(v) = \lambda\, v$ with non-zero $v \in V$, then v is an* **eigenvector** *for ℓ, with* **eigenvalue** *λ.*

Example 1.13.7. *In $C^\infty(\mathbb{R},\mathbb{R})$, e^{2x} is an eigenvector of $\frac{d}{dx}$, with eigenvalue 2.*

By Proposition 1.8.8, if U, V, W are F-vector spaces and $f : U \to V$, $g : V \to W$ linear maps then $g \circ f : U \to W$ is also linear. Letting $U = V = W$, if $f, g \in L(V)$ then $g \circ f \in L(V)$.

Theorem 1.13.8. *Given any F-vector space V, $(L(V), +, \circ)$ is a ring, with multiplicative identity* $\mathrm{id} : V \to V$.

Tedious proof omitted. In general, this ring is non-commutative. For example, if $V = F^n$ then $A \mapsto \ell_A$ gives a ring isomorphism $M_n(F) \xrightarrow{\sim} L(F^n)$. That matrix multiplication in $M_n(F)$ corresponds to composition in $L(F^n)$ was Proposition 1.8.9.

Example 1.13.9. *The element $\left(\frac{d}{dx}\right)^2$ of $L(C^\infty(\mathbb{R},\mathbb{R}))$ has eigenvalues $-m^2$ for $m \in \mathbb{N}_0$, with eigenvectors* $\cos mx$, $\sin mx$ *for $m^2 \neq 0$ and 1 for $m^2 = 0$.*

Example 1.13.10. *Recall from Semester 1, Section 2.7, the Weyl algebra W, a ring obtained from $\mathbb{C}$ by adjoining elements P, Q subject to the relation $PQ - QP = 1$. There is a ring homomorphism $\theta : W \to L(C^\infty(\mathbb{R},\mathbb{C}))$ such that*
$\theta(\lambda) = \lambda\,\mathrm{id}$ *(i.e. $f \mapsto \lambda f$) $\forall \lambda \in \mathbb{C}$;*
$\theta(Q) = x$ *(i.e. $f \mapsto xf$);*
$\theta(P) = \frac{d}{dx}$ *(i.e. $f \mapsto \frac{df}{dx}$). To know this is well-defined, we must check that $\theta(PQ - QP)$ and $\theta(1)$ agree, i.e. that $\frac{d}{dx} \circ x - x \circ \frac{d}{dx} = \mathrm{id}$. But $\frac{d}{dx}(xf) = f + x\frac{df}{dx}$, by the Product Rule, so*

$$\frac{d}{dx}(xf) - x\frac{df}{dx} = f, \quad i.e. \quad \frac{d}{dx} \circ x - x \circ \frac{d}{dx} = \mathrm{id},$$

as required. There was a leftover challenge question from Semester 1: is P a unit in W, i.e. does it have a multiplicative inverse in W? The answer is "No", because applying θ, $\theta(P) = \frac{d}{dx}$ would be invertible in $L(C^\infty(\mathbb{R},\mathbb{C}))$. But $\frac{d}{dx}$ is not invertible in $L(C^\infty(\mathbb{R},\mathbb{C}))$, because it maps any constant to 0, so is not surjective.

1.14. KERNEL, IMAGE, QUOTIENT SPACES, FIRST ISOMORPHISM THEOREM. .

A linear map $\ell : V \to W$ of F-vector spaces is in particular a homomorphism of additive groups, so it has a kernel, as defined in Semester 1, Definition 1.8.10, i.e.

$$\ker \ell := \{v \in V \ : \ \ell(v) = \underline{0}_W\}.$$

By Semester 1, Proposition 1.8.11, $\ker \ell$ is an additive subgroup of V. Now remembering that, in a vector space, addition is not the only operation, we might hope for more.

Proposition 1.14.1. *If $\ell : V \to W$ is a linear map of F-vector spaces then* $\ker \ell$ *is an F-vector subspace of V.*

Proof. We already know that $v_1, v_2 \in \ker \ell \implies v_1 + v_2 \in \ker \ell$, since ℓ is a homomorphism of additive groups. it remains to show that $\lambda\, v \in \ker \ell$ whenever $v \in \ker \ell$ and $\lambda \in F$. But if $v \in \ker \ell$ then $\ell(v) = \underline{0}_W$, and now

$$\ell(\lambda\, v) \overset{\text{linearity of } \ell}{=} \lambda(\ell(v)) = \lambda\, \underline{0}_W \overset{\text{Prop. } 1.3.1(1)}{=} \underline{0}_W,$$

so $\lambda\, v \in \ker \ell$, as required. □

Example 1.14.2. *If $A \in M_{m,n}(F)$, recall the linear map $\ell_A : F^n \to F^m$ given by $\mathbf{x} \mapsto A\mathbf{x}$. Then* $\ker \ell_A = \{\mathbf{x} \in F^n \ : \ A\mathbf{x} = \underline{0}\} = \mathrm{Null}(A)$*, and Proposition 1.14.1 recovers Theorem 1.6.2.*

Example 1.14.3. *The kernel of the linear map $\frac{d}{dx} : C^1(\mathbb{R}, \mathbb{R}) \to C(\mathbb{R}, \mathbb{R})$ is the 1-dimensional subspace* $\mathrm{Span}\{1\}$*, the constant functions (cf. Example 1.12.7).*

Example 1.14.4. *If $A \in M_{m,n}(F)$ and λ is an eigenvalue of A, then* $\ker(\ell_A - \lambda\,\mathrm{id}) = \mathrm{Null}(A - \lambda\, I)$ *is the λ-eigenspace of A.*

Example 1.14.5. *Inside $C^\infty(\mathbb{R}, \mathbb{R})$, the kernel of the linear operator $\left(\frac{d}{dx}\right)^2 + m^2\mathrm{id}$ (with $m \in \mathbb{N}_{>0}$) is the 2-dimensional subspace* $\mathrm{Span}\{\cos mx, \sin mx\}$*.*

We also saw in Semester 1 (Proposition 1.8.5) that the image of a group homomorphism is a subgroup of the codomain. Again, we can say more.

Proposition 1.14.6. *If $\ell : V \to W$ is a linear map of F-vector spaces then $\ell(V) = \mathrm{im}(\ell)$ is an F-vector subspace of W.*

Proof. We already know that $\ell(V)$ is non-empty and closed under addition. It remains to prove closure under scalar multiplication. But if $w \in \ell(V)$, say $w = \ell(v)$, then

$$\lambda\, w = \lambda(\ell(v)) \overset{\text{linearity of } \ell}{=} \ell(\lambda\, v) \in \ell(V),$$

as required. □

Example 1.14.7. *If $A \in M_{m,n}(F)$, so $\ell_A : F^n \to F^m$, then* $\mathrm{im}(\ell_A) = \{A\mathbf{x} \ : \ \mathbf{x} \in F^n\} = \mathrm{Col}(A)$ *(cf. Definition 1.4.8).*

Now let V be an F-vector space, U an F-vector subspace of V. Since V is an abelian group under addition, U is automatically a normal subgroup, so we may consider the quotient group V/U under addition: $\overline{v_1} + \overline{v_2} = \overline{v_1 + v_2}$, where $\overline{v} := v + U$.

Proposition 1.14.8. *On the additive group V/U there is also a well-defined operation of scalar multiplication: $\lambda\,\overline{v} := \overline{\lambda\, v}$, and with these two operations V/U becomes an F-vector space.*

Proof. First we must show that scalar multiplication is well-defined, i.e. that if $\overline{v_1} = \overline{v_2}$ then $\overline{\lambda\, v_1} = \overline{\lambda\, v_2}$. But if $\overline{v_1} = \overline{v_2}$ then $v_1 - v_2 \in U$, and now

$$\lambda\, v_1 - \lambda\, v_2 = \lambda(v_1 - v_2) \in U,$$

since U is closed under scalar multiplication, so $\overline{\lambda\, v_1} = \overline{\lambda\, v_2}$, as required. Now, to show that V/U is a vector space is a straightforward check of the list of axioms (details omitted). □

Proposition 1.14.9. *Let V be a finite-dimensional F-vector space, U a subspace of V. Then* $\dim(V/U) = \dim V - \dim U$.

Proof. Let $\dim V = n$, $\dim U = m$, and let $\{v_1, \ldots, v_m\}$ be a basis for U. By Proposition 1.10.6, we may extend this linearly independent subset of V to a basis $\{v_1, \ldots, v_m, v_{m+1}, \ldots, v_n\}$ for V. It suffices to show that $\{\overline{v}_{m+1}, \ldots, \overline{v}_n\}$ is a basis for V/U. Given any $\overline{v} \in V/U$, we have $v = \sum_{i=1}^n \alpha_i v_i$ for unique $\alpha_i \in F$. Then $\overline{v} = \sum_{i=m+1}^n \alpha_i \overline{v}_i$, since the difference $v - \sum_{i=m+1}^n \alpha_i v_i = \sum_{i=1}^m \alpha_i v_i$, which belongs to U. Adjusting v by an element of U, to get a different representative of the same element of V/U, only changes at most the coefficients $\alpha_1, \ldots, \alpha_m$, so the uniqueness of $\alpha_{m+1}, \ldots, \alpha_n$ follows from that of $\alpha_1, \ldots, \alpha_n$. □

Theorem 1.14.10 (First Isomorphism Theorem for Vector Spaces)**.** *Let $\ell : V \to W$ be an F-linear map of F-vector spaces. Then $\overline{\ell}(\overline{v}) := \ell(v)$ gives a well-defined F-linear isomorphism of vector spaces*

$$V/\ker \ell \simeq \mathrm{im}(\ell).$$

Proof. By the First Isomorphism Theorem for groups (Semester 1, Theorem 1.8.18) we know that $\overline{\ell}$ is a well-defined isomorphism of additive groups. It suffices to show that it respects scalar multiplication. But if $\lambda \in F$ and $\overline{v} \in V/\ker \ell$ then

$$\overline{\ell}(\lambda\,\overline{v}) = \overline{\ell}(\overline{\lambda\, v}) = \ell(\lambda\, v) \overset{\text{linearity of } \ell}{=} \lambda(\ell(v)) = \lambda(\overline{\ell}(\overline{v})),$$

as required. □

Example 1.14.11. *Consider the linear map* $\ell : \mathbb{R}^2 \to \mathbb{R}$ *given by* $\ell\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = x$. *Any* $a \in \mathbb{R}$ *is* $\ell\left(\begin{pmatrix} a \\ 0 \end{pmatrix}\right)$, *so* ℓ *is surjective. And* $\ker \ell$ *is the subspace* $L_0 : \ x = 0$, *i.e.* $\mathrm{Span}\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}$. *The general element of* $\mathbb{R}^2/\ker \ell$ *is the coset* $\begin{pmatrix} a \\ b \end{pmatrix} + L_0$, *which is the line* $L_a : \ x = a$. *The bijection* $\mathbb{R}^2/\ker \ell \simeq \mathbb{R}$ *is* $L_a \mapsto a$.

Example 1.14.12. *Consider again the non-homogeneous system of linear equations* $A\mathbf{x} = \mathbf{b}$, *where* $A \in M_{m,n}(F)$. *We seek the general solution to* $A\mathbf{x} = \mathbf{b}$, *i.e.* $\ell_A(\mathbf{x}) = \mathbf{b}$. *This is the subset* $\ell_A^{-1}(\{\mathbf{b}\})$ *of* F^n, *which is a coset of* $\ker \ell_A$, *equal to the element* $\overline{\ell_A}^{-1}(\mathbf{b})$ *of* $F^n/\ker \ell_A$, *where* $\overline{\ell}_A : F^n/\ker \ell_A \simeq F^m$ *is the isomorphism given by the First Isomorphism Theorem. So the general solution is of the form* $\mathbf{x} \in \overline{\mathbf{x}_p} = \mathbf{x}_p + \ker \ell_A$, *where* $\mathbf{x}_p$ *is any particular solution to* $A\mathbf{x} = \mathbf{b}$, *i.e. any element of* F^n *representing the coset* $\overline{\ell_A}^{-1}(\mathbf{b}) \in F^n/\ker \ell_A$. *Looking at* $\ker \ell_A$ *as the set of solutions to the homogeneous system* $A\mathbf{x} = \underline{0}$, *we recover the expression* $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$ *from Proposition 1.11.11.*

Example 1.14.13. *The general solution of the homogeneous equation* $\frac{d^2y}{dx^2} + 4y = 0$ *is* $y_h = A\cos 2x + B\sin 2x$. *A particular solution to the non-homogeneous equation* $\frac{d^2y}{dx^2} + 4y = e^x$ *is* $y_p = (1/5)e^x$, *so (from MAS110) the general solution is* $y = y_p + y_h = (1/5)e^x + A\cos 2x + B\sin 2x$. *Analogous to the previous example, we may view this in the following way. Let* $\ell \in L(C^\infty(\mathbb{R},\mathbb{R}))$ *be the linear operator such that* $\ell(y) = \frac{d^2y}{dx^2} + 4y$. *The general solution we seek is the subset* $\ell^{-1}(\{e^x\})$ *of* $C^\infty(\mathbb{R},\mathbb{R})$. *This subset is a coset, equal to the element* $\overline{(1/5)e^x} = (1/5)e^x + \ker \ell$ *of* $C^\infty(\mathbb{R},\mathbb{R})/\ker \ell$, *where* $\ker \ell = \mathrm{Span}\{\cos 2x, \sin 2x\}$.

Example 1.14.14. *(cf. Example 1.14.3.) The kernel of the linear map* $\frac{d}{dx} : C^1(\mathbb{R},\mathbb{R}) \to C(\mathbb{R},\mathbb{R})$ *is the* 1*-dimensional subspace* $\mathrm{Span}\{1\}$, *the constant functions. Any* $g \in C(\mathbb{R},\mathbb{R})$ *is* $\frac{d}{dx}(G(x))$, *where* $G(x) := \int_0^x g(t)\,dt$ *(Fundamental Theorem of Calculus). So* $\mathrm{im}\left(\frac{d}{dx}\right) = C(\mathbb{R},\mathbb{R})$. *We have an isomorphism* $C^1(\mathbb{R},\mathbb{R})/\mathrm{Span}\{1\} \simeq C(\mathbb{R},\mathbb{R})$ *given by* $\overline{f} \mapsto \frac{df}{dx}$. *Writing the coset* $\overline{f}$ *in the more familiar way* $f(x) + C$, *this is* $f(x) + C \mapsto \frac{df}{dx}$, *and the inverse map is* $g \mapsto \int_0^x g(t)\,dt \ + C$.

Theorem 1.14.15 (Rank-Nullity Theorem)**.** *Let* $\ell : V \to W$ *be a linear map of* F*-vector spaces, with* V *finite-dimensional. Then*

$$\dim(\ker \ell) + \dim(\mathrm{im}\,\ell) = \dim V.$$

Proof. By the First Isomorphism Theorem, $V/\ker\ell \simeq \operatorname{im}\ell$. Equating dimensions, and applying Proposition 1.14.9 to the dimension of the left hand side, $\dim V - \dim(\ker\ell) = \dim(\operatorname{im}\ell)$, as required. □

To see the reason for the name, consider the case of $\ell_A : F^n \to F^m$, with $A \in M_{m,n}(F)$. Then $\dim(\ker\ell_A) = \dim(F^n) - \dim(\operatorname{im}\ell_A)$ becomes $\text{Nullity}(A) = n - \dim(\text{Col}(A)) = n - \text{rank}(A)$ (cf. Examples 1.14.2, 1.14.7), and we recover Theorem 1.11.8.

1.15. **CHANGE OF BASIS.** .

Let V be a finite-dimensional F-vector space. If we choose a basis $\mathcal{B} = \{b_1, \ldots, b_n\}$ for V, where $n = \dim V$, we have an isomorphism $\iota_{\mathcal{B}} : F^n \xrightarrow{\sim} V$ given by

$$\iota_{\mathcal{B}}\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = \sum_{i=1}^{n} x_i b_i.$$

Now given any linear operator $\ell \in L(V)$, i.e. a linear map $\ell : V \to V$, the linear operator $\iota_{\mathcal{B}}^{-1} \circ \ell \circ \iota_{\mathcal{B}} \in L(F^n)$ is necessarily of the form ℓ_A, for some matrix $A \in M_n(F)$. This is the matrix representing ℓ with respect to the basis $\mathcal{B}$. (This is the case $W = V$, $B = C = \mathcal{B}$ of Definition 1.9.5.) Concretely, if

$$\ell\left(\sum_{j=1}^{n} x'_j b_j\right) = \sum_{i=1}^{m} y'_i b_i, \text{ then } \mathbf{y}' = A\mathbf{x}'.$$

In particular, $\ell(b_j) = \sum_{i=1}^{n} a_{ij} b_i$.

One might ask how the matrix representing a linear operator changes if we change the basis. The answer is that we conjugate the matrix by an invertible *transition* matrix that gets us from one basis to the other. To illustrate how this works, we consider the case that $V = F^n$ and we change from any basis $\mathcal{B} = \{b_1, \ldots, b_n\}$ to the standard basis $\{e_1, \ldots, e_n\}$.

Proposition 1.15.1. *Let $\ell \in L(F^n)$ be any linear operator. Let $A \in M_n(F)$ be the matrix representing ℓ with respect to the basis $\mathcal{B} = \{b_1, \ldots, b_n\}$ and let C be the matrix representing ℓ with respect to the standard basis $\{e_1, \ldots, e_n\}$. Then*

$$C = BAB^{-1},$$

where $B = (b_1|b_2|\ldots|b_n)$.

Proof. Given $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^n x_i e_i \in F^n$, if also $\mathbf{x} = \sum_{i=1}^n x'_i b_i$ and we let $\mathbf{x}' = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$ then $\mathbf{x} = B\mathbf{x}'$, because the right hand side is $(b_1|\dots|b_n)\begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = \sum_{i=1}^n x'_i b_i$, which is indeed $\mathbf{x}$. Similarly if $\mathbf{y} := \ell(\mathbf{x})$ and $\mathbf{y} = \sum_{i=1}^n y'_i b_i$ then $\mathbf{y} = B\mathbf{y}'$. We can think of $\mathcal{B}$ as giving us an alternative coordinate system on F^n, and the matrix B takes us from that back to the standard coordinate system.

Now $\mathbf{y} = C\,\mathbf{x}$ and $\mathbf{y}' = A\,\mathbf{x}'$, by the definition of representing matrices. Substituting $\mathbf{x} = B\mathbf{x}'$ and $\mathbf{y} = B\mathbf{y}'$ in the first equation, we get $B\mathbf{y}' = CB\mathbf{x}'$, so $\mathbf{y}' = B^{-1}CB\mathbf{x}'$. Comparing with the second equation, $A\mathbf{x}' = B^{-1}CB\mathbf{x}'$. This is for all $\mathbf{x}'$, from which it follows that $A = B^{-1}CB$ and $C = BAB^{-1}$. Note that B *is* invertible since $\{b_1, \dots, b_n\}$ is linearly independent. □

The matrices C and A are said to be "similar", and one can show that similarity is an equivalence relation.

Example 1.15.2. *The reflection* ref_0 *is represented by the matrix* $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ *with respect to the standard basis, because* $\mathrm{ref}_0(e_1) = e_1$ *and* $\mathrm{ref}_0(e_2) = -e_2$. *Similarly for general* β, *if* $b_1 := \begin{pmatrix} \cos(\beta/2) \\ \sin(\beta/2) \end{pmatrix}$ *and* $b_2 := \begin{pmatrix} -\sin(\beta/2) \\ \cos(\beta/2) \end{pmatrix}$, *then* ref_β *is represented by this same matrix* A *with respect to the basis* $\mathcal{B} = \{b_1, b_2\}$. *This is because* $\mathrm{ref}_\beta(b_1) = b_1$ *and* $\mathrm{ref}_\beta(b_2) = -b_2$. *(Note that* b_1, *along the axis of reflection, is an eigenvector for* ref_β *with eigenvalue* 1, *while* b_2, *perpendicular to the axis of reflection) is an eigenvector for* ref_β *with eigenvalue* -1.*) With respect to the standard basis,* ref_β *is represented by the matrix* $C = BAB^{-1}$, *by Proposition 1.15.1. Carrying out this calculation, one finds the same matrix* $\begin{pmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{pmatrix}$ *that we found in Semester 1. Section 1.9, where it was called* $M(\mathrm{ref}_\beta)$.

Example 1.15.3. *Let* $\ell \in L(\mathbb{R}^3)$ *be the rotation about the z-axis, through an angle* θ, *in the right-handed sense. It is represented by the matrix* $A := \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$. *Now choose a right-handed system* $\{b_1, b_2, b_3\}$ *of mutually orthogonal vectors of unit length, and let* $R \in L(\mathbb{R}^3)$ *be a rotation about the axis* b_3, *through an angle* θ *in the right-handed sense with respect to* b_1 *and* b_2. *Then similarly* R *is represented with respect to the basis* $\mathcal{B} = \{b_1, b_2, b_3\}$ *by this same*

matrix A. With respect to the standard basis $\{e_1, e_2, e_3\}$, R would be represented by the matrix $C = BAB^{-1}$.

Now suppose that $\ell \in L(F^n)$ is represented by matrices A with respect to some basis $\mathcal{B}$ and C with respect to the standard basis, where $C = BAB^{-1}$, as above. Then observe that

$$\det(C) = \det(BAB^{-1}) = \det(B)\det(A)\det(B^{-1}) = \det(B)\det(A)(\det(B))^{-1} = \det(A) \text{ and}$$

$$\operatorname{tr}(C) = \operatorname{tr}(B(AB^{-1})) = \operatorname{tr}((AB^{-1})B) \text{ (cf. Question 34) } = \operatorname{tr}(A).$$

It follows that we may define $\det(\ell)$ and $\operatorname{tr}(\ell)$ for any linear operator $\ell \in L(F^n)$ (or more generally $\ell \in L(V)$ for any finite-dimensional vector space V) as the determinant or trace of any representing matrix, and it doesn't make any difference which representing matrix (i.e. which basis) we choose.

As an application of this, consider a crystal structure whose atoms are positioned at the points of a lattice $L = \{a_1v_1 + a_2v_2 + a_3v_3 : a_1, a_2, a_3 \in \mathbb{Z}\}$, where $\{v_1, v_2, v_3\}$ is a basis for $\mathbb{R}^3$. Let R be a rotation about the origin, through an angle θ, such that $R(L) \subseteq L$, i.e. a rotational symmetry of the crystal. For each $1 \leq j \leq 3$, $v_j \in L \implies R(v_j) \in L$, so each $R(v_j)$ is an integer linear combination of v_1, v_2, v_3. The coefficients go in the j^{th} column of the matrix representing R with respect to the basis $\{v_1, v_2, v_3\}$, so this matrix is in $M_3(\mathbb{Z})$, hence $\operatorname{tr}(R) \in \mathbb{Z}$. But with respect to a basis $\{b_1, b_2, b_3\}$ as in Example 1.15.3 R is represented by the matrix $A = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$, so $\operatorname{tr}(R) = \operatorname{tr}(A) = 1 + 2\cos\theta$. Hence $2\cos\theta \in \mathbb{Z}$, so $\cos\theta = -1, -1/2, 0, 1/2$ or 1, and $\theta = \pm 2\pi/k$, with $k = 1, 2, 3, 4$ or 6 (never $k = 5$ or $k > 6$). That this was observed in nature provided evidence for the atomic hypothesis. Google "images for crystals".

2. INNER PRODUCT SPACES

2.1. INNER PRODUCT SPACES, CAUCHY-SCHWARZ INEQUALITY.

Definition 2.1.1. *Let V be a vector space* over $\mathbb{R}$.
A map $\langle , \rangle : V \times V \to \mathbb{R}$ is called an **inner product** *if and only if*

IP1 $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle \ \forall\, u, v, w \in V$;

IP2 $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle \qquad \forall\, u, v \in V, \lambda \in \mathbb{R}$;

IP3 $\langle u, v\rangle = \langle v, u\rangle \qquad \forall\, u, v \in V$;

IP4 $\forall v \in V$, $\langle v, v\rangle \geq 0$, *with equality if and only if* $v = 0$.

V (strictly speaking $(V, \langle , \rangle)$) is then said to be an **inner product space** *(real).*

Definition 2.1.2. *Let V be an inner product space. For any $v \in V$, the* **length** *or* **norm** *of v is defined by*

$$||v|| := \sqrt{\langle v, v \rangle}.$$

The **distance** *between $u, v \in V$ is defined to be $||u - v||$.*

Definition 2.1.3. *Let V be an inner product space. Vectors $u, v \in V$ are said to be* **orthogonal** *(or perpendicular) if and only if $\langle u, v \rangle = 0$.*

Proposition 2.1.4. *For any natural number $n \geq 1$, $\mathbb{R}^n$ is an inner product space with*

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i \quad \forall \, \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n.$$

Proof. **IP1**: $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \sum_{i=1}^{n} (x_i + y_i) z_i = \sum_{i=1}^{n} x_i z_i + \sum_{i=1}^{n} y_i z_i = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$.
IP2: $(\lambda \mathbf{x}) \cdot \mathbf{y} = \sum_{i=1}^{n} (\lambda x_i) y_i = \lambda \left(\sum_{i=1}^{n} x_i y_i \right) = \lambda (\mathbf{x} \cdot \mathbf{y})$.
IP3: $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} y_i x_i = \mathbf{y} \cdot \mathbf{x}$.
IP4: $\mathbf{x} \cdot \mathbf{x} = \sum_{i=1}^{n} x_i^2 \geq 0$, with equality if and only if all $x_i = 0$. □

Remark. IP1,2,3 can be proved also using $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^t \, \mathbf{y}$.

Theorem 2.1.5 (Cauchy-Schwarz Inequality)**.** *Let V be an inner product space. For any $u, v \in V$,*

$$|\langle u, v \rangle| \leq ||u|| \, ||v||.$$

Proof. Fix $u, v \in V$. Let $t \in \mathbb{R}$ be variable. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(t) := ||tu + v||^2 = \langle tu + v, tu + v \rangle$. Observe that $f(t) \geq 0 \; \forall t \in \mathbb{R}$, by **IP4**. Using **IP1,2,3** (see Question 35 for details), one can show that

$$f(t) = t^2 \langle u, u \rangle + 2t \langle u, v \rangle + \langle v, v \rangle,$$

which is a quadratic function of t, with non-negative coefficient of t^2. It cannot have distinct real roots (since $f(t) \geq 0$ for *all* $t \in \mathbb{R}$), so $(2\langle u, v \rangle)^2 - 4\langle u, u \rangle \langle v, v \rangle \leq 0$. Hence $\langle u, u \rangle^2 \leq \langle u, u \rangle \langle v, v \rangle$. Taking square roots, $|\langle u, v \rangle| \leq ||u|| \, ||v||$, as required. □

Corollary 2.1.6. *Now it is OK to define the angle θ between any non-zero u, v in an inner product space V, by*

$$\theta = \cos^{-1} \left(\frac{\langle u, v \rangle}{||u|| \, ||v||} \right)$$

. (The Cauchy-Schwarz inequality ensures that the number we are trying to take $\cos^{-1}$ of does lie in its domain $[-1, 1]$.)

Corollary 2.1.7 (Triangle Inequality)**.**

$$||u+v|| \leq ||u|| + ||v||.$$

Proof.

$$||u+v||^2 = \langle u+v, u+v\rangle = \langle u,u\rangle + 2\langle u,v\rangle + \langle v,v\rangle = ||u||^2 + 2\langle u,v\rangle + ||v||^2$$

$$\leq ||u||^2 + 2|\langle u,v\rangle| + ||v||^2$$

$$\leq ||u||^2 + 2||u||\,||v|| + ||v||^2 \text{ by Cauchy-Schwarz}$$

$$= (||u|| + ||v||)^2.$$

Now take square roots. □

Example 2.1.8. *In* $\mathbb{R}^4$ *let* $\mathbf{x} = \begin{pmatrix}1\\2\\3\\4\end{pmatrix}$, $\mathbf{y} = \begin{pmatrix}4\\3\\2\\1\end{pmatrix}$. *Then* $\mathbf{x}\cdot\mathbf{y} = 4+6+6+4 = 20$ *and* $||\mathbf{x}|| = ||\mathbf{y}|| = \sqrt{1^2+2^2+3^2+4^2} = \sqrt{30}$. *In accordance with Cauchy-Schwarz,* $|\mathbf{x}\cdot\mathbf{y}| = 20 \leq 30 = ||\mathbf{x}||\,||\mathbf{y}||$. *The angle between* $\mathbf{x}$ *and* $\mathbf{y}$ *is* $\theta = \cos^{-1}(20/30) \approx 0.84 \approx 48.2^o$. *And in accordance with the Triangle Inequality,*

$$||\mathbf{x}+\mathbf{y}|| = ||(5,5,5,5)^t|| = \sqrt{100} = 10 \leq 2\sqrt{30} = ||\mathbf{x}|| + ||\mathbf{y}||.$$

We have introduced inner products to capture the geometrical notions of length and angle, which the theory of vector spaces says nothing about. We now have, for any integer $n \geq 1$, a perfectly well-defined mathematical object $\mathbb{R}^n$ (simply a set of ordered n-tuples of real numbers, with componentwise addition and scalar multiplication) with its dot product, and we can do this purely arithmetically, without any recourse to geometrical intuition. One might now ask, is real physical space actually like this (for $n = 3$)? The answer is actually no (we appear to live in a curved space-time, according to general relativity), but it is nonetheless an extremely good approximation at earthly scales.

2.2. **ORTHOGONALITY.**

Definition 2.2.1. *Let* V *be an inner product space. A set of non-zero vectors* $\{v_1,\ldots,v_r\} \subseteq V$ *is said to be* **orthogonal** *if* $\langle v_i, v_j\rangle = 0$ *whenever* $i \neq j$.

Example 2.2.2. $\left\{\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}3\\0\\-1\end{pmatrix}, \begin{pmatrix}1\\-5\\3\end{pmatrix}\right\}$ *is an orthogonal subset of* $\mathbb{R}^3$ *(with dot product).*

Proposition 2.2.3. *If $\{v_1, \ldots, v_r\}$ is an orthogonal subset of an inner product space V, then $\{v_1, \ldots, v_r\}$ is linearly independent.*

Proof. Suppose that $\lambda_1 v_1 + \ldots + \lambda_r v_r = \underline{0}$ for some $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$. We must show that $\lambda_1 = \ldots = \lambda_r = 0$. To do this, for any fixed i with $1 \leq i \leq r$, take the inner product of both sides with v_i, to get

$$\lambda_i \langle v_i, v_i \rangle + \sum_{j \neq i} \lambda_j \langle v_j, v_i \rangle = \langle \underline{0}, v_i \rangle = 0,$$

but $\langle v_j, v_i \rangle = 0$ for $j \neq i$, and $\langle v_i, v_i \rangle \neq 0$ (by **IP4**, since $v_i \neq 0$), so $\lambda_i = 0$, as required. □

[Exercise: prove carefully from the axioms that in an inner product space, $\langle \underline{0}, v \rangle = 0$ for all v.]

Corollary 2.2.4. *(cf. Proposition 1.10.4) If V is a finite-dimensional inner product space, with $\dim V = n$, and if $\{v_1, \ldots, v_r\}$ is an orthogonal subset of V, then $r \leq n$, and if $r = n$ then $\{v_1, \ldots, v_r\}$ is an orthogonal* **basis** *for V.*

Example 2.2.5. $\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -5 \\ 3 \end{pmatrix} \right\}$ *must be an orthogonal basis for $\mathbb{R}^3$.*

Question: do they always exist?

Definition 2.2.6. *An orthogonal subset $\{v_1, \ldots, v_r\}$ of an inner product space V is said to be* **orthonormal** *if further each $||v_i|| = 1$.*

If $\{v_1, \ldots, v_r\}$ is orthogonal then $\{\frac{v_1}{||v_1||}, \ldots, \frac{v_r}{||v_r||}\}$ is orthonormal.

Example 2.2.7. $\left\{ \frac{1}{\sqrt{14}} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}, \frac{1}{\sqrt{35}} \begin{pmatrix} 1 \\ -5 \\ 3 \end{pmatrix} \right\}$ *is an orthonormal basis for $\mathbb{R}^3$. So is* $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$. *One is obtained by rotating the other.*

Proposition 2.2.8. *Suppose that $\{v_1, \ldots, v_n\}$ is an orthogonal basis for a finite-dimensional inner product space V. Then for any $v \in V$,*

$$v = \sum_{i=1}^{n} \frac{\langle v, v_i \rangle}{||v_i||^2} v_i.$$

Proof. Since $\{v_1, \ldots, v_n\}$ is a basis, $v = \sum_{i=1}^n \alpha_i v_i$ for some $\alpha_i \in \mathbb{R}$. Hence

$$\langle v, v_j \rangle = \sum_{i=1}^n \alpha_i \langle v_i, v_j \rangle = \alpha_j \langle v_j, v_j \rangle,$$

so $\alpha_j = \frac{\langle v, v_j \rangle}{\langle v_j, v_j \rangle} = \frac{\langle v, v_j \rangle}{||v_j||^2}$, as required. □

If $\{v_1, \ldots, v_n\}$ is an **orthonormal** basis then all $||v_i||^2 = 1$, so $v = \sum_{i=1}^n \langle v, v_i \rangle v_i$.

Example 2.2.9. *The standard basis* $\{e_1, \ldots, e_n\}$ *of* $\mathbb{R}^n$ *is orthonormal. If* $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ *then* $\mathbf{a} \cdot e_i = a_i$, *and* $\mathbf{a} = \sum_{i=1}^n a_i e_i = \sum_{i=1}^n (a \cdot e_i) e_i$.

Definition 2.2.10. *Let* V *be an inner product space,* W *a finite-dimensional subspace with an orthogonal basis* $\{w_1, \ldots, w_m\}$. *Given* $v \in V$, *define* $P_W(v)$, *the* **orthogonal projection of** v **onto** W *by*

$$P_W(v) := \sum_{i=1}^m \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} w_i = \sum_{i=1}^m \frac{\langle v, w_i \rangle}{||w_i||^2} w_i.$$

(Note that $P_W(v) \in W$, and that $P_W(w) = w$ for any $w \in W$. Also, it follows from Proposition 2.2.17 below that the definition of $P_W(v)$ is independent of which orthogonal basis for W one chooses.)

Proposition 2.2.11. $v - P_W(v)$ *is orthogonal to (every vector in)* W.

Proof. For any $1 \leq j \leq m$,

$$\langle v - P_W(v), w_j \rangle = \langle v - \sum_{i=1}^m \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} w_i, w_j \rangle$$

$$= \langle v, w_j \rangle - \sum_{i=1}^m \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} \langle w_i, w_j \rangle = \langle v, w_j \rangle - \frac{\langle v, w_j \rangle}{\langle w_j, w_j \rangle} \langle w_j, w_j \rangle = 0.$$

It follows that $\langle v - P_W(v), w \rangle = 0$ for any $w = \sum_{i=1}^m \alpha_i w_i \in W$ □

So $v = P_W(v) + (v - P_W(v))$, a sum of something in W and something orthogonal to W.

Theorem 2.2.12. *Let* V *be a finite-dimensional inner product space. Then* V *has an orthogonal basis.*

Proof. (**Gram-Schmidt orthogonalisation process.**)

Let $\{v_1, \ldots, v_n\}$ be **any** basis for V. We seek, step-by-step, to obtain an orthogonal basis $\{w_1, \ldots, w_k\}$ for $\mathrm{Span}\{v_1, \ldots, v_k\}$, for each $1 \leq k \leq n$.

Base step $k = 1$. Just let $w_1 = v_1$.

"Inductive step" Suppose that we have an orthogonal basis $\{w_1, \ldots, w_k\}$ for $\mathrm{Span}\{v_1, \ldots, v_k\}$, for some $1 \leq k < n$. Our aim is to extend it to an orthogonal basis $\{w_1, \ldots, w_{k+1}\}$ for $\mathrm{Span}\{v_1, \ldots, v_{k+1}\}$. Let

$$w_{k+1} = v_{k+1} - \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} w_i.$$

It is easy to check that $\mathrm{Span}\{w_1, \ldots, w_{k+1}\} = \mathrm{Span}\{v_1, \ldots, v_{k+1}\}$. Moreover, since $\sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} w_i$ is the orthogonal projection of v_{k+1} on the subspace $\mathrm{Span}\{w_1, \ldots, w_k\}$, we see that w_{k+1} is orthogonal to that subspace (in particular to each of $w_1, \ldots, w_k$), by Proposition 2.2.11. Hence $\{w_1, \ldots, w_k, w_{k+1}\}$ is an orthogonal basis for $\mathrm{Span}\{v_1, \ldots, v_{k+1}\}$. □

Example 2.2.13. *We seek an orthogonal basis* $\{w_1, w_2, w_3\}$ *for* $\mathbb{R}^3$*, with* $w_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$*. Apply Gram-Schmidt to the basis* $\{v_1, v_2, v_3\} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$.

Step 1. $w_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Step 2. $w_2 = v_2 - \frac{\langle v_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/3 \\ 2/3 \\ -1/3 \end{pmatrix}$. *It is convenient to re-scale this to a new* $w_2 = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$.

Step 3. $w_3 = v_3 - \frac{\langle v_3, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle v_3, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{-1}{6} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 0 \\ 1/2 \end{pmatrix}$. *Again we re-scale, to a new* $w_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.

The conclusion is that $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$ *is an orthogonal basis for* $\mathbb{R}^3$*, so*

$\left\{ \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$ *is an orthonormal basis for* $\mathbb{R}^3$

As an application of this, let $R \in L(\mathbb{R}^3)$ be the rotation through $\pi/2$, in a right-handed sense, about the axis $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. We can calculate the matrix C representing R with respect to the standard basis of $\mathbb{R}^3$. With respect to the orthonormal basis $\{b_1, b_2, b_3\}$ found above, R is represented by the matrix $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$. (This is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$, with $\theta = \pi/2$, cf. Example 1.15.3.)

Let $B = [b_1|b_2|b_3]$. Now $b_i \cdot b_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$

Equivalently, $B^t B = I$, i.e. B is an orthogonal matrix, so $B^{-1} = B^t$. By Proposition 1.15.1, $C = BAB^{-1} = BAB^t$, which works out to $\begin{pmatrix} \frac{1}{3} & \frac{1}{3}(1-\sqrt{3}) & \frac{1}{3}(1+\sqrt{3}) \\ \frac{1}{3}(1+\sqrt{3}) & \frac{1}{3} & \frac{1}{3}(1-\sqrt{3}) \\ \frac{1}{3}(1-\sqrt{3}) & \frac{1}{3}(1+\sqrt{3}) & \frac{1}{3} \end{pmatrix}$.

Definition 2.2.14. *Let* V *be an inner product space,* W *a subspace. The* **orthogonal complement** *of* W *is*

$$W^\perp := \{v \in V \,:\, \langle v, w\rangle = 0 \ \ \forall w \in W\}.$$

Proposition 2.2.15. $W^\perp$ *is also a subspace of* V*.*

To prove this is Question 43.

Proposition 2.2.16. $W \cap W^\perp = \{\underline{0}\}$.

Proof. If $w \in W \cap W^\perp$ then $\langle w, w\rangle = 0$, so $w = \underline{0}$, by **IP4**. □

Proposition 2.2.17. *If* W *is finite-dimensional then given any* $v \in V$*,* $v = w + w'$ *for unique* $w \in W$*,* $w' \in W^\perp$*.*

Proof. Since W is finite-dimensional, W has an orthogonal basis, by Theorem 2.2.12. Hence we may apply Proposition 2.2.11 (see the definition preceding it), and take $w = P_W(v), w' =$

$v - P_W(v)$. For the uniqueness, suppose also $v = w_1 + w_1'$, with $w_1 \in W$ and $w_1' \in W^\perp$. Then $w + w' = w_1 + w_1' \implies w - w_1 = w_1' - w' \in W \cap W^\perp = \{\underline{0}\}$, so $w = w_1$ and $w' = w_1'$. □

In the language of Questions 10 and 11, V is isomorphic to the **direct sum** $W \oplus W^\perp$, and it follows that $\dim V = \dim W + \dim W^\perp$, if V is finite-dimensional.

We can gain some fresh insight into the subject of homogeneous linear equations, using these ideas. Let $V = \mathbb{R}^n$ with the dot product. For $A \in M_{m,n}(\mathbb{R})$ write $A = \begin{bmatrix} \underline{r_1} \\ \vdots \\ \underline{r_m} \end{bmatrix}$; in other words let r_i be the i^{th} row of A. Now the i^{th} equation in the system $A\mathbf{x} = \underline{0}$ is $r_i\mathbf{x} = 0$, i.e. $r_i^t \cdot \mathbf{x} = 0$, so $\text{Null}(A) = W^\perp$, where $W = \text{Span}\{r_1^t, \ldots, r_m^t\} = \text{Col}(A^t)$. Now the relation $\dim W + \dim W^\perp = \dim V$ becomes $\text{rank}(A) + \text{Nullity}(A) = n$, recovering Theorem 1.11.8 (or Proposition 1.11.3). Note that the relationship between W and $W^\perp$ is symmetrical; they are orthogonal complements of each other, of complementary dimensions adding up to n.

Example 2.2.18. *In* $\mathbb{R}^3$, *if* $A = [1,1,1]$ *then* $W = \text{Span}\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}$ *and* $W^\perp = \text{Null}(A) = \text{Span}\left\{\begin{pmatrix}-1\\1\\0\end{pmatrix}, \begin{pmatrix}-1\\0\\1\end{pmatrix}\right\}$, *but also* $W = (W^\perp)^\perp = \text{Null}\left(\begin{pmatrix}-1 & 1 & 0\\ -1 & 0 & 1\end{pmatrix}\right)$.

2.3. **INNER PRODUCTS ON FUNCTION SPACES.** .

Recall (following Example 1.12.2) the $\mathbb{R}$-rector space $C[a,b] = C([a,b], \mathbb{R})$ of continuous, real-valued functions on the closed interval $[a,b]$. This is an infinite-dimensional vector space, and we can think of each value $f(x)$ of the function $f \in C[a,b]$ as being analogous to an entry a_i in a column vector $\mathbf{a}$. So the infinite indexing set of $x \in [a,b]$ replaces the finite indexing set of $i \in \{1, 2, \ldots, n\}$, but the choices of function values have to be made in such a way that the function is continuous overall. Anyway, the dot product on $\mathbb{R}^n$ then inspires the following way of defining an inner product on $C[a,b]$, with integration replacing summation.

Proposition 2.3.1. $\langle f, g\rangle := \int_a^b f(t)g(t)\,dt$ *defines an inner product on* $C[a,b]$.

Proof. .

IP1: $\langle f_1 + f_2, g\rangle = \int_a^b (f_1(t) + f_2(t))g(t)\,dt = \int_a^b f_1(t)g(t)\,dt + \int_a^b f_2(t)g(t)\,dt = \langle f_1, g\rangle + \langle f_2, g\rangle$.

IP2: $\langle \lambda f, g\rangle = \int_a^b \lambda f(t)g(t)\,dt = \lambda \int_a^b f(t)g(t)\,dt = \lambda\langle f, g\rangle$.

IP3: $\langle f, g\rangle = \int_a^b f(t)g(t)\,dt = \int_a^b g(t)f(t)\,dt = \langle g, f\rangle$.

IP4: $\langle f, f\rangle = \int_a^b (f(t))^2\,dt \geq 0$. To prove strict inequality when $f \neq 0$ requires careful use of the continuity of f. Those doing MAS221 may enjoy tackling this in Question 38. □

Example 2.3.2. *Inside* $C[-\pi,\pi]$ *(with inner product as above) consider the subspace (prove it is)* $\mathrm{CP}[-\pi,\pi] := \{f \in C[-\pi,\pi] : f(-\pi) = f(\pi)\}$. *(Note that such an* f *can be extended to a continuous* 2π*-periodic function* $f : \mathbb{R} \to \mathbb{R}$*.) Then*

$$\{1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots\}$$

is an orthogonal subset of $\mathrm{CP}[-\pi,\pi]$. *For example*

$$\langle \cos x, \sin x\rangle = \int_{-\pi}^{\pi} \cos x \sin x\, dx = \int_{-\pi}^{\pi} \frac{1}{2}\sin 2x\, dx = [(-1/4)\cos 2x]_{-\pi}^{\pi} = (-1/4)(1-1) = 0.$$

Another way to see this is that the integrand is an odd function. For the full set of integrals see MAS110, Chapter 7, Question 10. We have

$$||1||^2 = \langle 1, 1\rangle = \int_{-\pi}^{\pi} 1^2\, dx = 2\pi.$$

For $m > 0$,

$$||\cos mx||^2 = \langle \cos mx, \cos mx\rangle = \int_{-\pi}^{\pi} \cos^2 mx\, dx = \frac{1}{2}\int_{-\pi}^{\pi} 1 + \cos 2mx\, dx$$

$$= \frac{1}{2}[x + (1/2m)\sin 2mx]_{-\pi}^{\pi} = (1/2)(2\pi) = \pi,$$

and similarly $||\sin mx||^2 = \pi$.

Recall from Proposition 2.2.8 that if $\{v_1, \ldots, v_n\}$ is an orthogonal basis for a finite-dimensional inner product space V, then for any $v = \sum_{i=1}^{n} \alpha_i v_i \in V$, $\alpha_i = \frac{\langle v, v_i\rangle}{||v_i||^2}$.

By analogy, if $f \in \mathrm{CP}[-\pi,\pi]$ and $f(x) = a_0 + \sum_{m=1}^{\infty}(a_m \cos mx + b_m \sin mx)$ (the meaning of this infinite sum is a matter for Analysis), we might expect that

$$a_0 = \frac{\langle f, 1\rangle}{||1||^2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\, dx,$$

$$a_m = \frac{\langle f, \cos mx\rangle}{||\cos mx||^2} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos mx\, dx$$

(for $m > 0$), and

$$b_m = \frac{\langle f, \sin mx\rangle}{||\sin mx||^2} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin mx\, dx.$$

These are the formulas for Fourier coefficients, mentioned at the end of MAS211, and more familiar to those who have returned to Fourier series in MAS222.

Example 2.3.3. *If in* $C[-1,1]$ *we apply the Gram-Schmidt orthogonalisation process to* $\{1, x, x^2, x^3, \ldots\}$, *we get an orthogonal set* $\{L_0(x), L_1(x), L_2(x), \ldots\}$ *of polynomials.*

To work out the first few, we will need the general formula

$$\langle x^m, x^n\rangle = \int_{-1}^{1} x^{m+n}\,dx = \left[\frac{x^{m+n+1}}{m+n+1}\right] = \begin{cases} 0 & m \not\equiv n \pmod 2; \\ \frac{2}{m+n+1} & m \equiv n \pmod 2. \end{cases}$$

Step 1. $L_0 = 1$.

Step 2. $L_1 = x - \frac{\langle x,1\rangle}{\langle 1,1\rangle}\,1 = x - \frac{0}{2}\,1 = x$.

Step 3. $L_2 = x^2 - \frac{\langle x^2,x\rangle}{\langle x,x\rangle}\,x - \frac{\langle x^2,1\rangle}{\langle 1,1\rangle}\,1 = x^2 - 0.x - \frac{2/3}{2}\,1 = x^2 - 1/3$.

Step 4. $L_3 = x^3 - \frac{\langle x^3,x^2-1/3\rangle}{\langle x^2-1/3,x^2-1/3\rangle}\,(x^2-1/3) - \frac{\langle x^3,x\rangle}{\langle x,x\rangle}\,x - \frac{\langle x^3,1\rangle}{\langle 1,1\rangle}\,1 = x^3 - 0(x^2-1/3) - \frac{2/5}{2/3}x - 0(1) = x^3 - (3/5)x$, etc.

The $L_n(x)$ are usually re-scaled to $P_n(x)$ such that $P_n(1) = 1$, then these are called Legendre polynomials (Legendre 1752-1833). The first few are

$$P_0(x) = 1, P_1(x) = x, P_2(x) = (1/2)(3x^2-1), P_3(x) = (1/2)(5x^3-3x), P_4(x) = (1/8)(35x^4-30x^2+3).$$

Look at pictures on Wikipedia. For any fixed $n \geq 1$, the finite set $\{P_0(x), P_1(x), \ldots, P_n(x)\}$ is an orthogonal basis for $\mathbb{R}[x]_{\leq n}$, the subspace of polynomials of degree up to n.

2.4. ADJOINTS AND SELF-ADJOINT OPERATORS.

Definition 2.4.1. *Let V be an inner product space, $T \in L(V)$ a linear operator on V. An* **adjoint** *to T is any $T^* \in L(V)$ such that*

$$\langle Tv, w\rangle = \langle v, T^*w\rangle \quad \forall\, v, w \in V.$$

One may show that if it exists then it is unique (see Question 46).

Proposition 2.4.2. *Consider $\mathbb{R}^n$ with its dot product, a matrix $A \in M_n(\mathbb{R})$ and the associated linear operator $\ell_A \in L(\mathbb{R}^n)$ given by $\ell_A(\mathbf{x}) = A\mathbf{x}$ $\forall\, \mathbf{x} \in \mathbb{R}^n$. Then we have an adjoint $(\ell_A)^* = \ell_{A^t}$.*

Proof. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\langle A\mathbf{x}, \mathbf{y}\rangle = (A\mathbf{x})^t\,\mathbf{y} \overset{\text{Q1(b)}}{=} (\mathbf{x}^tA^t)\mathbf{y} = \mathbf{x}^t(A^t\,\mathbf{y}) = \langle \mathbf{x}, A^t\,\mathbf{y}\rangle.$$

□

Definition 2.4.3. *Let V be an inner product space, $T \in L(V)$. We say that T is* **self-adjoint** *if*

$$\langle Tv, w\rangle = \langle v, Tw\rangle \quad \forall\, v, w \in V,$$

i.e. if T^* exists and is equal to T.

Example 2.4.4. *$\ell_A \in L(\mathbb{R}^n)$ is self-adjoint (with respect to dot product) $\iff A^t = A \iff A$ is symmetric.*

Proposition 2.4.5. *Let $(V, \langle , \rangle)$ be a finite-dimensional inner product space, with $\dim V = n$. Then there exists a linear bijection $\ell : V \to \mathbb{R}^n$ such that*

$$\langle u, v\rangle = \ell(u) \cdot \ell(v) \quad \forall\, u, v \in V.$$

(In other words, $\ell : (V, \langle , \rangle) \to (\mathbb{R}^n, \cdot)$ is an isomorphism of inner product spaces.)

Proof. By Theorem 2.2.12, V has an orthonormal basis $B = \{w_1, \ldots, w_n\}$.

Recall that $\langle w_i, w_j\rangle = \begin{cases} 0 & i \neq j; \\ 1 & i = j. \end{cases}$.

We may define a linear isomorphism $\ell : V \to \mathbb{R}^n$ by $\ell(\sum_{i=1}^n a_i w_i) = \sum_{i=1}^n a_i e_i = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$.

This is just ι_B^{-1}, where the linear isomorphism $\iota_B : \mathbb{R}^n \to V$ is as in Theorem 1.9.4. Now, if $u = \sum_{i=1}^n a_i w_i$ and $v = \sum_{i=1}^n b_i w_i$ then

$$\langle u, v\rangle = \left\langle \sum_{i=1}^n a_i w_i, \sum_{j=1}^n b_j w_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n a_i b_j \langle w_i, w_j\rangle = \sum_{i=1}^n a_i b_i = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \ell(u) \cdot \ell(v).$$

□

Corollary 2.4.6. *If V is a finite-dimensional inner product space, any $T \in L(V)$ has an adjoint.*

Theorem 2.4.7. *Let V be an inner-product space, $T \in L(V)$ a self-adjoint operator.*

(1) If $T(v_1) = \lambda_1 v_1$ and $T(v_2) = \lambda_2 v_2$, with $v_1, v_2 \in V - \{\underline{0}\}$ and $\lambda_1 \neq \lambda_2$, then v_1 and v_2 are orthogonal. (Eigenvectors for distinct eigenvalues of a **self-adjoint** *operator are orthogonal.)*

(2) If W is a finite-dimensional subspace of V, and if $T(W) \subseteq W$, then $T(W^\perp) \subseteq W^\perp$, where $W^\perp$ is the orthogonal complement of W (Definition 2.2.14).

Proof. (1) $\lambda_1\langle v_1, v_2\rangle \overset{\textbf{IP2}}{=} \langle \lambda_1 v_1, v_2\rangle = \langle Tv_1, v_2\rangle \overset{T \text{ self-adjoint}}{=} \langle v_1, Tv_2\rangle = \langle v_1, \lambda_2 v_2\rangle = \lambda_2\langle v_1, v_2\rangle$. But $\lambda_1 \neq \lambda_2$, so $\langle v_1, v_2\rangle = 0$.

(2) Given $u \in W^\perp$, we want to show that $Tu \in W^\perp$, i.e. that $\langle w, Tu\rangle = 0 \ \forall\, w \in W$. But $\langle w, Tu\rangle = \langle Tw, u\rangle = 0$, since $Tw \in W$ and $u \in W^\perp$.

□

Example 2.4.8. *$V = \mathbb{R}^2$, $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, which is symmetric, so ℓ_A is self-adjoint. The characteristic polynomial* $\det(xI - A) = (x-1)^2 - 4$ *has roots $\lambda = 1 \pm 2$, i.e. $3, -1$. These are the eigenvalues of A (or of ℓ_A).*
$A - 3I = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}$. Its null space (the "eigenspace" for $\lambda = 3$) is spanned by $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
$A - (-1)I = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$. Its null space (the eigenspace for $\lambda = -1$) is spanned by $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. These are orthogonal eigenvectors for distinct eigenvalues, in keeping with Theorem 2.4.7(i). (The eigenspaces also illustrate (ii).)

Example 2.4.9. *Let $T \in L(\mathbb{R}^3)$ be reflection in the plane $P : x + y + z = 0$. With respect to the orthonormal basis $\{b_1, b_2, b_3\}$ in Example 2.2.13, T is represented by the symmetric matrix $\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. It follows that T is self-adjoint. We have $T(P) \subseteq P$, with T acting as $+1$ on P, and (in accord with Theorem 2.4.7(ii)), also $T(P^\perp) \subseteq P^\perp$ (with T acting as -1 on the line $P^\perp$). Within P, there are lots of eigenvectors that are not orthogonal to each other. But this does not contradict Theorem 2.4.7(i), since they all belong to the* same *eigenvalue $\lambda = +1$.*

Recall that for V an inner product space, $T \in L(V)$, we say that $T^* \in L(V)$ is adjoint to T if $\langle Tv, w\rangle = \langle v, T^*w\rangle \ \forall\, v, w \in V$.

Proposition 2.4.10. *Consider the inner product space $V = \mathrm{CP}^\infty[-\pi, \pi] := \{f \in C^\infty([-\pi, \pi], \mathbb{R}) : f^{(n)}(-\pi) = f^{(n)}(\pi) \ \forall\, n \geq 0\}$ ($f^{(n)}$ being the n^{th} derivative of f), with $\langle f, g\rangle := \int_{-\pi}^{\pi} f(t)g(t)\,dt$.*

(1) $\frac{d}{dx} \in L(V)$ has adjoint $\left(\frac{d}{dx}\right)^* = -\frac{d}{dx}$.
(2) $\left(\frac{d}{dx}\right)^2$ is self-adjoint.

Proof. (1) $\langle \frac{df}{dx}, g\rangle + \langle f, \frac{dg}{dx}\rangle = \int_{-\pi}^{\pi} f'(x)g(x) + f(x)g'(x)\,dx \overset{\text{Product Rule}}{=} \int_{-\pi}^{\pi} (fg)'(x)\,dx = [(fg)(x)]_{-\pi}^{\pi} = 0$, since $(fg)(-\pi) = (fg)(\pi)$. So $\langle \frac{df}{dx}, g\rangle = \langle f, -\frac{dg}{dx}\rangle$, and $-\frac{d}{dx}$ is adjoint to $\frac{d}{dx}$.

(2) Using $(ST)^* = T^*S^*$ (see Question 45) with $S = T = \frac{d}{dx}$, $\left(\left(\frac{d}{dx}\right)^2\right)^* = \left(-\frac{d}{dx}\right)^2 = \left(\frac{d}{dx}\right)^2$, i.e. $\left(\frac{d}{dx}\right)^2$ is self-adjoint.

□

Example 2.4.11. *For $m \in \mathbb{Z}_{\geq 0}$,*

$$\left(\frac{d}{dx}\right)^2 \cos mx = -m^2 \cos mx,$$

and for $n \in \mathbb{Z}_{>0}$,

$$\left(\frac{d}{dx}\right)^2 \sin nx = -n^2 \sin nx.$$

Thus Theorem 2.4.7(i) explains why, for $m \neq n$,

$$\langle \cos mx, \sin nx \rangle = \langle \cos mx, \cos nx \rangle = \langle \sin mx, \sin nx \rangle = 0.$$

(They are eigenvectors, with distinct eigenvalues, for the self-adjoint operator $\left(\frac{d}{dx}\right)^2$.) But it does not explain why $\langle \cos nx, \sin nx \rangle = 0$, since they have the same eigenvalue $-n^2$.

Proposition 2.4.12. *Let $V = C^\infty[-1,1]$ (the space of real-valued functions on $[-1,1]$ with derivatives of all orders), with $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\,dx$. If we define $\mathcal{L} \in L(V)$ by*

$$\mathcal{L}(y) := ((1-x^2)y')' = (1-x^2)y'' - 2xy'$$

then $\mathcal{L}$ is self-adjoint.

Proof. $\langle \mathcal{L}(f), g \rangle$

$$= \langle ((1-x^2)f')', g \rangle = \int_{-1}^{1} ((1-x^2)f')'g\,dx = [(1-x^2)f'g]_{-1}^{1} - \int_{-1}^{1} (1-x^2)f'g'\,dx = -\int_{-1}^{1} (1-x^2)f'g'\,dx,$$

integrating by parts and using $1 - x^2 = 0$ at $x = \pm 1$. In this expression, f and g appear symmetrically, so it must be the same as $\langle \mathcal{L}(g), f \rangle$, which (by **IP3**) is $\langle f, \mathcal{L}(g) \rangle$, as required. □

Recall from Example 2.3.3 the Legendre polynomials

$$\{P_0(x), P_1(x), \ldots\} = \{1, x, (1/2)(3x^2-1), (1/2)(5x^3-3x), \ldots\}$$

obtained by applying Gram-Schmidt orthogonalisation to $\{1, x, x^2, \ldots\}$ (and re-scaling). By construction, $\{P_0(x), \ldots, P_n(x)\}$ is an orthogonal basis for $\mathbb{R}[x]_{\leq n}$. Note that $\mathcal{L}(\mathbb{R}[x]_{\leq n}) \subseteq \mathbb{R}[x]_{\leq n}$, using $\frac{d}{dx}(\mathbb{R}[x]_{\leq n}) \subseteq \mathbb{R}[x]_{\leq n-1}$. Now, fixing n, let $V = \mathbb{R}[x]_{\leq n}$ (a finite-dimensional subspace of $C^\infty[-1,1]$), and let W be the subspace $\mathbb{R}[x]_{\leq n-1}$. Then $W^\perp = \mathrm{Span}\{P_n(x)\}$. Now $\mathcal{L}$ is self-adjoint, and $\mathcal{L}(W) \subseteq W$, so by Theorem 2.4.7(ii), $\mathcal{L}(W^\perp) \subseteq W^\perp$. Hence $P_n(x)$ is an eigenvector of $\mathcal{L}$. Let λ be the eigenvalue, so $\mathcal{L}(P_n(x)) = \lambda P_n(x)$. Comparing coefficients of x^n, $-n(n-1) - 2n = \lambda$, so $\lambda = -n(n+1)$.

We have discovered that $P_n(x)$ is a solution to Legendre's differential equation

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0,$$

which is important in physical applied mathematics (e.g. Laplace's equation in spherical polar coordinates). It is actually $P_n(\cos\phi)$ which appears, explaining the domain $[-1,1]$.

2.5. COMPLEX INNER-PRODUCT SPACES, THE SPECTRAL THEOREM.

Definition 2.5.1. *Let V be a $\mathbb{C}$-vector space.*

A map $\langle , \rangle : V \times V \to \mathbb{C}$ is called an **inner product** *if and only if*

IP1 $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle \ \ \forall\, u, v, w \in V$;

IP2 $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle \qquad \forall\, u, v \in V, \lambda \in \mathbb{R}$;

IP3 $\langle u, v\rangle = \overline{\langle v, u\rangle} \qquad \forall\, u, v \in V$;

IP4 $\forall v \in V,\ \langle v, v\rangle \geq 0$, *with equality if and only if* $v = 0$.

V (strictly speaking $(V, \langle , \rangle)$) is then said to be a **complex inner product space**.

Note that in a complex inner product space, $\langle u, \lambda v\rangle = \overline{\lambda}\langle u, v\rangle$, as can be proved using **IP2** and **IP3** (exercise).

Example 2.5.2. *$V = \mathbb{C}^n$. For* $\mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$, *define*

$$\mathbf{z} \cdot \mathbf{w} := \mathbf{z}^t\, \overline{\mathbf{w}} = \sum_{i=1}^{n} z_i \overline{w_i}.$$

This is an inner product. For **IP4**, *note that*

$$||\mathbf{z}||^2 = \mathbf{z} \cdot \mathbf{z} = \sum_{i=1}^{n} z_i \overline{z_i} = \sum_{i=1}^{n} |z_i|^2.$$

We may define adjoints in exactly the same way as for real inner product spaces.

Definition 2.5.3. *Let V be a complex inner product space, $T \in L(V)$ a linear operator on V. An* **adjoint** *to T is any $T^* \in L(V)$ such that*

$$\langle Tv, w\rangle = \langle v, T^* w\rangle \ \ \forall\, v, w \in V.$$

Proposition 2.5.4. *Consider $\mathbb{C}^n$ with its dot product, a matrix $A \in M_n(\mathbb{C})$ and the associated linear operator $\ell_A \in L(\mathbb{C}^n)$ given by $\ell_A(\mathbf{x}) = A\mathbf{x}\ \ \forall\, \mathbf{x} \in \mathbb{C}^n$. Then we have an adjoint $(\ell_A)^* = \ell_{\overline{A}^t}$.*

Proof. For any $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$,

$$\langle A\mathbf{z}, \mathbf{w}\rangle = (A\mathbf{z})^t\, \overline{\mathbf{w}} \overset{\text{Q1(b)}}{=} (\mathbf{z}^t A^t)\overline{\mathbf{w}} = \mathbf{z}^t \overline{(\overline{A}^t \mathbf{w})} = \langle \mathbf{z}, \overline{A}^t \mathbf{w}\rangle.$$

□

Then ℓ_A is self-adjoint if and only if $\overline{A}^t = A$ (i.e. A is "Hermitian").

Example 2.5.5. $A = \begin{pmatrix} 1 & 1+i & 3i \\ 1-i & 1 & 3-2i \\ -3i & 3+2i & 2 \end{pmatrix}$ *is Hermitian.*

For $A \in M_n(\mathbb{R})$, Hermitian is the same as symmetric.

Theorem 2.5.6. *Let V be a complex inner product space, $T \in L(V)$ self-adjoint. If $Tv = \lambda v$, with $v \in V - \{\underline{0}\}$, $\lambda \in \mathbb{C}$, then in fact $\lambda \in \mathbb{R}$. In other words, eigenvalues of self-adjoint operators are real.*

Corollary 2.5.7. *The eigenvalues of a Hermitian (e.g. real, symmetric) matrix are real.*

Proof. $\lambda\langle v, v\rangle \overset{\mathbf{IP2}}{=} \langle \lambda v, v\rangle = \langle Tv, v\rangle \overset{T \text{ self-adjoint}}{=} \langle v, Tv\rangle \overset{\mathbf{IP3}}{=} \overline{\langle Tv, v\rangle} = \overline{\lambda\langle v, v\rangle} = \overline{\lambda}\,\overline{\langle v, v\rangle} \overset{\mathbf{IP3}}{=} \overline{\lambda}\langle v, v\rangle$. Since (by **IP4**) $\langle v, v\rangle \neq 0$, $\lambda = \overline{\lambda}$, i.e. $\lambda \in \mathbb{R}$. □

Theorem 2.5.8 (Spectral Theorem)**.** *Let V be a finite-dimensional inner product space, $T \in L(V)$ self-adjoint. Then V has an orthogonal basis $\{w_1, \ldots, w_n\}$ of eigenvectors for T.*

Example 2.5.9. $V = \mathbb{C}^2$, $T = \ell_A$, $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ *(which is Hermitian). Then* $\{\mathbf{w}_1, \mathbf{w}_2\} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$ *is an orthogonal basis for V, with $A\mathbf{w}_1 = 3\mathbf{w}_1, A\mathbf{w}_2 = -\mathbf{w}_2$. (cf. Example 2.4.8.)*

Proof. As in Proposition 2.4.5, $(V, \langle , \rangle)$ is isomorphic as an inner product space to $(\mathbb{C}^n, \cdot)$, where $n = \dim V$, so we may assume that we are dealing with the latter. (It follows that $T = \ell_A$ for some Hermitian matrix $A \in M_n(\mathbb{C})$.) Let $\lambda_1 \in \mathbb{C}$ be any root of the characteristic polynomial of T (i.e. of A). (There *is* a root in $\mathbb{C}$, by the Fundamental Theorem of Algebra, cf. Example 3.3.4 in Semester 1. In fact, by Theorem 2.5.6, $\lambda_1 \in \mathbb{R}$.) Let $\mathbf{w}_1 \in \mathbb{C}^n$ be an eigenvector, with $T\mathbf{w}_1 = \lambda_1\mathbf{w}_1$. If $\dim V = 1$ then $\{\mathbf{w}_1\}$ is the desired orthogonal basis. (This is the base step for a proof by induction.) Now suppose that $\dim V > 1$, and let $W = \mathrm{Span}\{\mathbf{w}_1\}$. Then $T(W) \subseteq W$. As in Theorem 2.4.7(ii), $T(W^\perp) \subseteq W^\perp$, so we may view the restriction $T|_{W^\perp}$ as a self-adjoint operator on $W^\perp$. Now $\dim W^\perp = n - 1$ (by the paragraph following Proposition 2.2.17), so by induction on the dimension we may assume that $W^\perp$ has an orthogonal basis $\{\mathbf{w}_2, \ldots, \mathbf{w}_n\}$ of eigenvectors for $T|_{W^\perp}$. Then V has an orthogonal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ of eigenvectors for T. □

Remark. With respect to the basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$, T is represented by the diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. By Proposition 1.15.1, $A = PDP^{-1}$, where $P = (\mathbf{w}_1 | \ldots | \mathbf{w}_n)$. Equivalently, $D = P^{-1}AP$. If we normalise to make the basis orthonormal, then P is an orthogonal

matrix, i.e. $P^t P = I$, so $P^{-1} = P^t$, and we get $D = P^t AP$, or $A = PDP^t$. We have *diagonalised* the matrix A. For real symmetric matrices you have seen this before, in Section 5.2 of MAS211, but a proof was given only in the special case that the eigenvalues are distinct. The diagonalisation of real symmetric matrices (and the associated quadratic forms) has many applications.

Proposition 2.5.10. *Consider the complex inner product space* $V = \mathrm{CP}^\infty([-\pi,\pi],\mathbb{C}) := \{f \in C^\infty([-\pi,\pi],\mathbb{C}) : f^{(n)}(-\pi) = f^{(n)}(\pi)\ \forall n \geq 0\}$, *with* $\langle f,g\rangle := \int_{-\pi}^{\pi} f(t)\overline{g(t)}\,dt$. *Then* $\frac{1}{i}\frac{d}{dx} \in L(V)$ *is self-adjoint.*

Proof. Just as in the proof of Proposition 2.4.10, $\left(\frac{d}{dx}\right)^* = -\frac{d}{dx}$. Using $(\alpha T)^* = \overline{\alpha}\,T^*$ for $\alpha \in C$ (Question 47), with $T = \frac{d}{dx}$ and $\alpha = \frac{1}{i}$,

$$\left(\frac{1}{i}\frac{d}{dx}\right)^* = \frac{1}{\overline{i}}\left(\frac{d}{dx}\right)^* = -\frac{1}{i}\left(-\frac{d}{dx}\right) = \frac{1}{i}\frac{d}{dx},$$

as required. □

Now, for any $m \in \mathbb{Z}$, $\left(\frac{1}{i}\frac{d}{dx}\right)(e^{imx}) = \frac{1}{i}\,ime^{imx} = me^{imx}$, so e^{imx} is an eigenvector, with eigenvalue m. As In Theorem 2.4.7(i), for $m > 0$, e^{imx} and e^{-imx} are orthogonal, since they are eigenvectors for distinct eigenvalues (m and $-m$) of the self-adjoint operator $\left(\frac{1}{i}\frac{d}{dx}\right)$. Note that

$$||e^{inx}||^2 = \int_{-\pi}^{\pi} e^{inx}\overline{e^{inx}}\,dx = \int_{-\pi}^{\pi} 1\,dx = 2\pi.$$

Then

$$\langle \cos mx, \sin mx\rangle = \left\langle \frac{e^{imx}+e^{-imx}}{2}, \frac{e^{imx}-e^{-imx}}{2i}\right\rangle = \frac{1}{2}\overline{\left(\frac{1}{2i}\right)}(||e^{imx}||^2 - ||e^{-imx}||^2)$$

(the cross terms vanish, by orthogonality of e^{imx} and e^{-imx}), which is equal to $(i/4)(2\pi - 2\pi) = 0$, thus explaining the orthogonality of $\cos mx$ and $\sin mx$ (cf. Example 2.4.11).

A wonderful application of complex inner product spaces is to quantum mechanics, where complex-valued wave functions are vectors in a complex inner product space, representing physical states. Observable quantities are represented by self-adjoint operators, but only if the state vector is an eigenvector does such an observable have a definite value, the eigenvalue. (Now we see the significance of this eigenvalue being real.) For example, the x-component of momentum is represented by the self-adjoint operator $\frac{h}{2\pi i}\frac{\partial}{\partial x}$, where h is Planck's constant. MAS324 Quantum Theory highly recommended.