CM222, Linear Algebra

1. Vector Spaces

Definition 1.1. A **real vector space** (or a **vector space over** $\mathbb{R}$) is a set V together with two operations '+' and '.' called *addition* and *scalar multiplication* respectively, such that:
(i) addition is a binary operation on V which makes V into an abelian group and
(ii) the operation of scalar multiplying an element $v \in V$ by an element $\alpha \in \mathbb{R}$ gives an element $\alpha.v$ of V.

In addition, the following axioms must be satisfied:
(a) $1.v = v$ for all $v \in V$
(b) If $\alpha, \beta \in \mathbb{R}$ and $v \in V$ then $\alpha.(\beta.v) = (\alpha\beta).v$
(c) If $\alpha \in \mathbb{R}$ and $v, w \in V$ then $\alpha.(v + w) = \alpha.v + \alpha.w$
(d) If $\alpha, \beta \in \mathbb{R}$ and $v \in V$ then $(\alpha + \beta).v = \alpha.v + \beta.v$

We refer to the elements of V as *vectors*.

The definition of a **complex vector space** is exactly the same except that field $\mathbb{R}$ of scalars is replaced by the complex field $\mathbb{C}$.

In fact one can have vector spaces over any field; in this course we will not consider any other fields (but most of what we do will hold with no change for any field).

Informally, a vector space is a set of elements that can be added and multiplied by scalars and obey the usual rules.

Example 1.2. $\mathbb{R}^n = \{ (x_1, x_2, \ldots, x_n) \,|\, x_1, x_2, \ldots, x_n \in \mathbb{R} \}$ with its usual operations of vector addition and scalar multiplication is a real vector space. (We could just as well use "column vector" notation.)

Example 1.3. $\{0\}$ (with the output of any operation being 0 of course) is a real vector space.

Example 1.4. $\mathbb{C}$ is a vector space over $\mathbb{R}$ (with the usual operations, so addition is defined by $(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2)$ and scalar multiplication is defined by $\alpha.(x + iy) = (\alpha x) + i(\alpha y)$.

We usually omit the symbol '.' in the notation for scalar multiplication. Note also that since V is an abelian group under $+$, it has an additive identity (or zero) element, which we'll denote as 0_V or simply $\mathbf{0}$. As usual, the additive inverse of a vector $v \in V$ is denoted $-v$.

Example 1.5. If m and n are positive integers, we let $M_{m,n}(\mathbb{R})$ denote the set of $m \times n$ real matrices. We can add two such matrices, the sum being defined by the formula

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$
$$= \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & & & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix}.$$

We can abbreviate the above notation by writing the first matrix above as $A = (a_{ij})_{i,j}$, meaning that A is the matrix whose entry in row i, column j (or '(i,j)-entry') is a_{ij} for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Similarly writing $B = (b_{ij})_{i,j}$, the formula becomes $A + B = (a_{ij} + b_{ij})_{i,j}$. Scalar multiplication of matrices is defined by $\alpha A = (\alpha a_{ij})_{i,j}$ for $\alpha \in \mathbb{R}$ and $A = (a_{ij})_{i,j} \in M_{m,n}(\mathbb{R})$.

Recall that $M_{m,n}(\mathbb{R})$ is an abelian group by standard properties of matrix addition. It is also easy to see that $M_{m,n}(\mathbb{R})$ is a vector space: a) It is clear that $1 \cdot A = A$ from the definition of scalar multiplication. b) If $\alpha, \beta \in \mathbb{R}$ and $A = (a_{ij})_{i,j}$, then

$$\alpha(\beta A) = \alpha(\beta a_{ij})_{i,j} = (\alpha(\beta a_{ij}))_{i,j} = ((\alpha\beta)a_{ij})_{i,j} = (\alpha\beta)A$$

(the third equality follows from associativity of multiplication on $\mathbb{R}$, the rest from the definition of scalar multiplication). The proofs of axioms c) and d) are similar, but using the distributive law on $\mathbb{R}$.

Example 1.6. Let $\mathcal{F}$ denote the set of functions $f : \mathbb{R} \to \mathbb{R}$. Recall that we can add two function by adding their values; i.e., if $f, g \in \mathcal{F}$, then $f + g$ is the function defined by $(f + g)(x) = f(x) + g(x)$. Similarly if $\alpha \in \mathbb{R}$ and $f \in \mathcal{F}$, then we define the scalar multiple of f by α by multiplying its values by α; i.e., $(\alpha f)(x) = \alpha(f(x))$. The proof that $\mathcal{F}$ is a real vector space is left as an exercise. Some interesting subsets of $\mathcal{F}$ are also vector spaces, e.g.,

- $\mathcal{D} = \{\, f : \mathbb{R} \to \mathbb{R} \mid f \text{ is differentiable} \,\}$;
- $\mathcal{P} = \{\, f : \mathbb{R} \to \mathbb{R} \mid f \text{ is defined by a polynomial} \,\}$.

We have the following basic properties, which we leave as an exercise to deduce from the axioms:

Proposition 1.7. *Suppose that V is a real vector space with zero element $\mathbf{0}$. Then*

(1) $0.v = \mathbf{0}$ *for all* $v \in V$*;*
(2) $\alpha.\mathbf{0} = \mathbf{0}$ *for all* $\alpha \in \mathbb{R}$*;*
(3) $(-1).v = -v$ *for all* $v \in V$*.*

Definition 1.8. If S is a subset of a vector space V, then the **span** of S, denoted by $\operatorname{span} S$, is the set of all finite sums of the form $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$ where $\alpha_i \in \mathbb{R}$ (some may be zero) and $v_i \in S$. We call a sum of the form $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$ a **linear combination** of elements of S; thus $\operatorname{span} S$ is the set of all linear combinations of elements of S. If the span of a set S is the whole vector space V (i.e. if $\operatorname{span} S = V$) then S is called a **spanning set**.

In practice, the set S will usually be finite, say $S = \{s_1, s_2, \ldots, s_m\}$. In that case

$$\operatorname{span} S = \{\, \alpha_1 s_1 + \alpha_2 s_2 + \cdots + \alpha_m s_m \mid \alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}\}.$$

(Note that every vector of the form on the right is in $\operatorname{span} S$, by the definition of $\operatorname{span} S$, since each $s_i \in S$. On the other hand, if $v \in \operatorname{span} S$, then $v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$ for some $\alpha_i \in \mathbb{R}$ and $v_i \in S$. Since $S = \{s_1, \ldots, s_m\}$, we know that each $v_i = s_{j_i}$ for some $j_i \in \{1, 2, \ldots, m\}$. Some of the s_j might appear mutliple times, others not all, but we can combine like terms, add $0.s_j$'s where necessary and apply the axioms and Prop. 1.7(a) to write v in the required form.)

By convention, the span of the empty set is $\{\mathbf{0}\}$.

Example 1.9. For $i = 1, \ldots, n$, let $s_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (with a 1 in the ith place, and 0's elsewhere). Then $S = \{s_1, \ldots, s_n\}$ spans $\mathbb{R}^n$.

Example 1.10. $\{1, i\}$ spans $\mathbb{C}$ (as a real vector space).

Example 1.11. Let $E_{ij} \in M_{m,n}(\mathbb{R})$ be the matrix in which the (i,j)-entry is 1 and the rest of the entries are 0. Then $\{\, E_{i,j} \mid i = 1, \ldots, m,\ j = 1, \ldots, n\,\}$ spans $M_{m,n}(\mathbb{R})$, since any matrix $A = (a_{ij})_{i,j}$ can be written as a linear combination of the $E'_{ij}s$:

$$A = \sum_{i,j} a_{ij} E_{ij}.$$

Example 1.12. The set of functions $\{1, x, x^2, x^3, \ldots\}$ spans the space $\mathcal{P}$ of polynomial functions mentioned in Example 1.6 (since, by definition, a polynomial is a linear combination of such functions).

Lemma 1.13. *Suppose that S, S' are subsets of a vector space V. If $S' \subset \operatorname{span} S$, then $\operatorname{span} S' \subset \operatorname{span} S$.*

(The proof is straightforward: if $v \in \operatorname{span} S'$, then v is a linear combination of elements of S'. But each element of S' is a linear combination of elements of S; substituting and expanding gives v as a linear combination of elements of S.)

Corollary 1.14. *If $S' \subset \operatorname{span} S$, then $\operatorname{span}(S \cup S') = \operatorname{span} S$.*

Definition 1.15. A subset S of a vector space V is **linearly dependent** if if there exist distinct elements $s_1, s_2, \ldots, s_k \in S$ and scalars $\alpha_1, \alpha_2, \ldots, \alpha_k$ *not all equal to* 0 such that $\alpha_1 s_1 + \alpha_2 s_2 + \cdots + \alpha_k s_k = \mathbf{0}$. If S is not linearly dependent we say that S is **linearly independent**.

By convention, the empty set is linearly independent.

Again in practice S will usually be finite, say $S = \{s_1, s_2, \dots, s_m\}$ with $s_1, \dots, s_m$ distinct. Then S is linearly independent if

$$(*) \qquad \alpha_1 s_1 + \alpha_2 s_2 + \cdots + \alpha_m s_m = \mathbf{0} \text{ implies that } \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0\,.$$

Example 1.16. The subset $S = \{s_1, s_2, \dots, s_n\}$ of $\mathbb{R}^n$ in Example 1.9 is linearly independent.

Example 1.17. The subset $S = \{(0,1), (1,0), (1,1)\}$ of $\mathbb{R}^2$ is linearly dependent since $1(0,1) + 1(1,0) + (-1)(1,1) = (0,0)$.

Example 1.18. Let $\mathcal{F}$ be the set of functions as in Example 1.6). The subset $S = \{f, g, h\}$ where $f(x) = x+1$, $g(x) = x+2$, $h(x) = x+3$ is linearly dependent (exercise).

It is sometimes convenient to be able to use the characterization of linear independence in $(*)$ above without demanding that $s_1, s_2, \dots, s_m$ be distinct. Note that if they are not distinct, i.e., $s_i = s_j$ for some $i \neq j$, then $1.s_i + (-1)s_j = 0$, so we automatically view $\{s_1, s_2, \dots, s_m\}$ as being *linearly dependent* if there is any repetition among the s_i. For example, we view $\{(0,1), (0,1)\}$ as linearly dependent, even though the S with single element $(0,1)$ is linearly independent. It should be clear from the context whether we mean the sequence of elements or the set.

Lemma 1.19. *Suppose that $s_1, s_2, \dots, s_{m+1}$ are elements of a vector space V and that $S = \{s_1, s_2, \dots, s_m\}$ is linearly independent. Then $s_{m+1} \in \operatorname{span} S$ if and only if $\{s_1, s_2, \dots, s_m, s_{m+1}\}$ is linearly dependent.*

Proof. Suppose first that $s_{m+1} \in \operatorname{span} S$. Then $s_{m+1} = \alpha_1 s_1 + \cdots + \alpha_m s_m$ for some $\alpha_i \in \mathbb{R}$. Therefore

$$\alpha_1 s_1 + \cdots + \alpha_m s_m + (-1)s_{m+1} = \mathbf{0},$$

so $\{s_1, s_2, \dots, s_m, s_{m+1}\}$ is linerly dependent. (Note that the coefficient of s_{m+1} is $-1 \neq 0$.)

Conversely suppose that $\{s_1, s_2, \dots, s_m, s_{m+1}\}$ is linearly dependent. This means that

$$\alpha_1 s_1 + \cdots + \alpha_m s_m + \alpha_{m+1} s_{m+1} = \mathbf{0}$$

for some $\alpha_1, \dots, \alpha_{m+1} \in \mathbb{R}$ with some $\alpha_i \neq 0$. Note that $\alpha_{m+1} \neq 0$ (otherwise we'd have $\alpha_1 s_1 + \cdots + \alpha_m s_m = \mathbf{0}$ contradicting that S is linearly independent). It follows that

$$s_{m+1} = (-\alpha_{m+1}^{-1}\alpha_1)s_1 + \cdots + (-\alpha_{m+1}^{-1}\alpha_m)s_m,$$

so $s_{m+1} \in \operatorname{span} S$. $\square$

Definition 1.20. A subset S of a vector space V is a **basis** of V if it is linearly independent and spans V. If V has a finite basis we say that V is **finite-dimensional.**

Example 1.21. The sets in Examples 1.9, 1.10 and 1.11 are all bases of the vector spaces being considered; they are all therefore finite-dimensional.

Example 1.22. The set $\{1, x, x^2, \ldots\}$ is a basis for $\mathcal{P}$ (which is not finite-dimensional).

Example 1.23. The set in Example 1.17 spans $\mathbb{R}^2$ but is not a basis of $\mathbb{R}^2$ since it is not linearly independent.

Theorem 1.24. *If $S = \{v_1, v_2, \ldots, v_m\}$ spans V then there is a basis of V which is a subset of S.*

Proof. (Sketch) If S is linearly independent, then it is a basis and we are done, so suppose S is linearly dependent. This means that

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = \mathbf{0}$$

for some $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ with some $\alpha_i \neq 0$. We can reorder the v_i and so assume that $\alpha_m \neq 0$. As in the proof of Lemma 1.19, we get that $v_m \in \operatorname{span} S'$ where $S' = \{v_1, \ldots, v_{m-1}\}$. By Corollary 1.14, we see that $\operatorname{span} S' = \operatorname{span}(S' \cup \{v_m\}) = \operatorname{span} S = V$, so S' spans V.

If S' is linearly independent, then we are done. Otherwise repeat the process, which must eventually terminate since S is finite. □

Corollary 1.25. *If V is spanned by a finite set, then V is finite-dimensional.*

Theorem 1.26. *If V is finite-dimensional and $S = \{v_1, v_2, \ldots, v_k\}$ is linearly independent, then there is a basis of V which contains S.*

Proof. Since V is finite-dimensional, V is spanned by a finite set $S' = \{s_1, \ldots, s_n\}$ for some $s_1, \ldots, s_n \in V$. The idea of the proof is to show that we can choose elements from S' to add to S to form a basis.

If S spans V, then S is itself is a basis, so there is nothing to prove. So suppose S does not span V. Then $S' \not\subset \operatorname{span} S$ (since if $S' \subset \operatorname{span} S$, then Lemma 1.13 would show that $V = \operatorname{span} S' \subset \operatorname{span} S$, contradicting the assumption that S does not span V). So we can choose some $s_{i_1} \in S'$ such that $s_{i_1} \notin \operatorname{span} S$. By Lemma 1.19, $S_1 = \{v_1, \ldots, v_k, s_{i_1}\}$ is linearly independent.

If $S_1 = \{v_1, \ldots, v_k, s_{i_1}\}$ spans V, then S_1 is a basis containing S, and we are done. If S_1 does not span V, then repeat the process using S_1 instead of S to get $s_{i_2} \in S'$ such that $S_2 = \{v_1, \ldots, v_k, s_{i_1}, s_{i_2}\}$ is linearly indpendent. (Note that $s_{i_2} \neq s_{i_1}$ since $s_{i_2} \notin \operatorname{span} S_1$.) If S_2 spans we are done; if not then repeat the process, which must yield a spanning set in at most n steps since we keep choosing distinct s_i from S' and S' itself spans V. □

Theorem 1.27. *Suppose that $S = \{v_1, \ldots, v_n\}$ spans V and $S' = \{w_1, \ldots, w_m\}$ is a set of m linearly independent vectors in V. Then $m \leq n$.*

Proof. The idea of the proof is to show that we can replace each w_i with v_i (reordering the v_i if necessary), leading to a contradiction if $m > n$.

If each $v_i \in \operatorname{span}\{w_2, \ldots, w_m\}$, then $S \subset \operatorname{span}\{w_2, \ldots, w_m\}$ implies that $V = \operatorname{span} S \subset \operatorname{span}\{w_2, \ldots, w_m\}$ (Lemma 1.13) and therefore that $w_1 \in \operatorname{span}\{w_2, \ldots, w_m\}$, contradicting that S' is linearly independent (Lemma 1.19). Therefore $v_i \notin \operatorname{span}\{w_2, \ldots, w_m\}$ for some i; reordering we can assume $v_i = v_1$. Applying Lemma 1.19 again, we see that $\{v_1, w_2, \ldots, w_n\}$ is linearly independent.

Now repeat the process with $\operatorname{span}\{v_1, w_3, \ldots, w_n\}$ instead of $\{w_2, \ldots, w_m\}$ to get that $v_i \notin \operatorname{span}\{v_1, w_3, \ldots, w_n\}$ for some $i \neq 1$; reordering we can assume $i = 2$, so $\{v_1, v_2, w_3, \ldots, w_n\}$ is linearly independent.

Continuing in this way, we get that if $m > n$, then $\{v_1, \ldots, v_n, w_{m+1}, \ldots, w_m\}$ is linearly independent, but $w_{m+1} \in V = \operatorname{span}\{v_1, \ldots, v_n\}$, contradicting Lemma 1.19 $\square$

This easily gives:

Theorem 1.28. [BASIS THEOREM] *Every basis of a finite-dimensional vector space has the same number of elements.*

Definition 1.29. If V is a finite-dimensional vector space the **dimension** of V (denoted $\dim V$) is the number of elements in any basis of V.

For example, $\mathbb{R}^n$ has dimension n, $\mathbb{C}$ has dimension 2 (as a vector space over $\mathbb{R}$), and $M_{m,n}(\mathbb{R})$ has dimension mn.

Let V be a finite-dimensional vector space with basis $S = \{v_1, v_2, \ldots, v_n\}$. Then any element $v \in V$ can be written uniquely in the form

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots \alpha_n v_n$$

where α_i are scalars. (It can be written this way since S spans, uniquely since S is linearly independent.) The n-tuple $\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$ is called the n-tuple of **coordinates** of v with respect to the basis $\{v_1, v_2, \ldots, v_n\}$. (Coordinates of vectors are normally regarded as column vectors but sometimes written in rows, or using "transpose notation," to save space.)

Note that if $V = \mathbb{R}^n$ and $S = \{s_1, \ldots, s_n\}$ is its "standard basis" (Example 1.9), then the n-tuple of coordinates of a vector is given by its usual coordinates (since $v = (\alpha_1, \ldots, \alpha_n = \alpha_1 s_1 + \cdots + \alpha_n s_n)$. The coordinates with respect to a different basis will be different.

Example 1.30. Let $v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ in $\mathbb{R}^2$. Then $\{v_1, v_2\}$ is linearly independent and span, and therefore form a basis. The standard basis vector $s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ has coordinates $s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ with respect to the standard basis, but has coordinates $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ with respect to $\{v_1, v_2\}$ (since

$s_1 = -v_1 + v_2$). The vector v_1 has coordinates $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ with respect to the standard basis, and has coordinates $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ with respect to $\{v_1, v_2\}$.

Suppose that $S = \{v_1, \dots, v_n\}$ and $S' = \{v'_1, \dots, v'_n\}$ are bases for V, and that $v \in V$ coordinates $(\alpha_1, \dots, \alpha_n)^t$ with respect to S and $(\alpha'_1, \dots, \alpha'_n)^t$ with respect to S' (where ^{t} denotes the transpose). We can relate the coordinate vectors with respect to the different bases as follows: Note that each $v_j \in S$ has an n-tuple of coordinates $(t_{1j}, \dots, t_{nj})$ with respect to S'. Define the **transition matrix** from S to S' as the $n \times n$-matrix $T = (t_{ij})_{i,j}$. Then

$$T \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \alpha'_1 \\ \vdots \\ \alpha'_n \end{pmatrix},$$

i.e., multiplication by T converts the coordinates with respect to S into the coordinates with respect to S'. (Proof:

$$v = \sum_{j=1}^{n} \alpha_j v_j = \sum_{j=1}^{n} \alpha_j \sum_{i=1}^{n} t_{ij} v'_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} t_{ij} \alpha_j \right) v'_i,$$

so $\alpha'_i = \sum_{j=1}^{n} t_{ij} \alpha_j$ for $i = 1, \dots, n$.)

Definition 1.31. A subset W of a vector space V is called a **subspace** if it is a vector space (with the operations as in V).

Equivalently a subset W of V is a subspace if it is non-empty and closed under the operations of addition and scalar multiplication. (See the exercises.)

For example,

- The set of solutions in $\mathbb{R}^n$ to a system of homogeneous linear equations is a subspace of $\mathbb{R}^n$.
- The set $\{ f \in \mathcal{P} \mid f \text{ has degree} \leq n \}$ is a subspace of $\mathcal{P}$, which is a subspace of $\mathcal{D}$, which is a subspace of $\mathcal{F}$ (see Example 1.6).
- If S is any subset of V, then $\operatorname{span} S$ is a subspace of V (exercise).

Lemma 1.32. *If W is a subspace of an n-dimensional vector space V then W is finite-dimensional with dimension $m \leq n$. The case $m = n$ holds if and only if $W = V$.*

Proof. We need to show that W has a basis with at most n elements. If $W = \{\mathbf{0}\}$, then $\dim W = 0$ (the empty set is a basis) and the lemma holds. If $W \neq \{\mathbf{0}\}$, then W contains some non-zero vector w_1, and $S_1 = \{w_1\}$ is linearly independent. If S_1 spans W, then S_1 is a basis, so $\dim W = 1$. If S_1 does not span W, then there is some $w_2 \notin \operatorname{span} S_1$, and by Lemma 1.19, $S_2 = \{w_1, w_2\}$ is linearly independent. Continuing in this way, we either get that $S_m = \{w_1, \dots, w_m\}$ is a basis for W for some $m < n$, or that

$S_n = \{w_1, \dots, w_n\}$ is linearly independent, in which case S_n is a basis for V, so $V = \operatorname{span} S_n \subset W$, so $W = V$ and has dimension n. □

2. Linear Maps

Definition 2.1. Let V and W be vector spaces. A map (or 'function' or 'transformation') $f : V \to W$ is said to be **linear** if preserves (or 'respects') addition and scalar multiplication, that is
(i) $f(v_1 + v_2) = f(v_1) + f(v_2)$ for all $v_1, v_2 \in V$ and
(ii)$f(\alpha v) = \alpha f(v)$ for all scalars α and $v \in V$.

Example 2.2. Let $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$ (viewed as column vectors). If A is an $m \times n$-matrix (over $\mathbb{R}$), then $f(v) = Av$ defines a linear map $f : V \to W$.

Example 2.3. Differentiation $\frac{d}{dx}$ defines a linear map from $\mathcal{D}$ to $\mathcal{F}$.

Definition 2.4. A map $f : V \to W$ is said to be an **isomorphism** (of real vector spaces) if f is linear and bijective. If there is an isomorphism from V to W, then we say V is **isomorphic** to W.

Proposition 2.5. *Suppose U, V, W are vector spaces and $f : V \to W$ and $g : U \to V$ are linear maps. Then*

(1) *The composite $f \circ g$ from U to W is linear.*
(2) *If f is an isomorphism, then so is its inverse map from W to V.*

The proof is left as an exercise.

Example 2.6. Suppose that V has dimension n and $\{v_1, \dots, v_n\}$ is a basis for V. Define $f : \mathbb{R}^n \to V$ by $f((\alpha_1, \dots, \alpha_n)^t) = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Then f is an isomorphism, the inverse being the map sending v to its coordinates. This shows that every vector space of dimension n is isomorphic to $\mathbb{R}^n$.

Definition 2.7. If $f : V \to W$ is a linear map, the **kernel** (or **null space**) $\ker(f)$ and the **image** (or **range**) $\operatorname{im}(f)$ of f are defined by
(i) $\ker(f) = \{v \in V \; f(v) = 0\}$,
(ii) $\operatorname{im}(f) = \{w \mid w = f(v) \text{ for some } v \in V\}$.

Proposition 2.8. *If $f : V \to W$ is a linear map, then*

(1) $\ker(f)$ *is a subspace of V;*
(2) $\operatorname{im}(f)$ *is a subspace of W;*
(3) $\ker(f) = \{0\}$ *if and only if f is injective.*

The proof is left as an exercise.

Example 2.9. Let $f : V \to W$ be as in Example 2.2. Then $\ker(f)$ is the null space of A (the set of solutions to $Av = \mathbf{0}$) and $\operatorname{im}(f)$ is the span of the columns of A.

Lemma 2.10. *Suppose that $f : V \to W$ is a linear map, $\{w_1, \dots, w_r\}$ is a basis for* $\operatorname{im}(f)$ *and $\{v_1, \dots, v_k\}$ is a basis for* $\ker(f)$. *For $i = 1, \dots, r$, let $u_i \in V$ be such that $f(u_i) = w_i$. Then $S = \{u_1, \dots, u_r, v_1, \dots, v_k\}$ is a basis for V.*

Proof. We must show that 1) S is linearly independent, and 2) S spans V.

1) Suppose that $\alpha_1 u_1 + \cdots \alpha_r u_r + \beta_1 v_1 + \cdots + \beta_k v_k = \mathbf{0}$. We must show that $\alpha_1 = \cdots = \alpha_r = \beta_1 = \cdots = \beta_k = 0$. Applying f to the given equation gives $f(\alpha_1 u_1 + \cdots \alpha_r u_r + \beta_1 v_1 + \cdots \beta_k v_k) = f(\mathbf{0}) = \mathbf{0}$. Since f is linear, we have

$$\begin{aligned}&f(\alpha_1 u_1 + \cdots \alpha_r u_r + \beta_1 v_1 + \cdots + \beta_k v_k)\\ &= \alpha_1 f(u_1) + \cdots + \alpha_r f(u_r) + \beta_1 f(v_1) + \cdots + \beta_k f(v_k)\\ &= \alpha_1 w_1 + \cdots + \alpha_r w_r = \mathbf{0}.\end{aligned}$$

Since $\{w_1, \ldots, w_r\}$ is linearly independent, we get $\alpha_1 = \cdots = \alpha_r = 0$. Therefore the given equation becomes $\beta_1 v_1 + \cdots + \beta_k v_k = \mathbf{0}$. Since $\{v_1, \ldots, v_k\}$ is linearly independent, it follows that $\beta_1 = \cdots = \beta_k = 0$ as well.

2) Suppose that $v \in V$. Since $f(v) \in \mathrm{im}(f) = \mathrm{span}\{w_1, \ldots, w_r\}$, we have $f(v) = \alpha_1 w_1 + \cdots + \alpha_r w_r$ for some $\alpha_1, \ldots, \alpha_r \in \mathbb{R}$. Letting $v' = v - (\alpha_1 u_1 + \cdots + \alpha_r u_r)$, we have

$$\begin{aligned}f(v') &= f(v) - (\alpha_1 f(u_1) + \cdots + \alpha_r f(u_r))\\ &= f(v) - \alpha_1 w_1 + \cdots + \alpha_r w_r = \mathbf{0}.\end{aligned}$$

Therefore $v' \in \ker(f) = \mathrm{span}\{v_1, \ldots, v_k\}$, so we have $v' = \beta_1 v_1 + \cdots + \beta_k v_k$ for some $\beta_1, \ldots, \beta_k \in \mathbb{R}$. It follows that

$$v = \alpha_1 u_1 + \cdots \alpha_r u_r + \beta_1 v_1 + \cdots + \beta_k v_k$$

is in the span of S. □

Definition 2.11. If V and W are finite-dimensional and $f : V \to W$ is a linear map, we define the **rank** of f (denoted $\mathrm{rank}(f)$) to be the dimension of $\mathrm{im}(f)$, and the **nullity** of f (denoted $\mathrm{nullity}(f)$) to be the dimension of $\ker(f)$.

It follows immediately from the lemma that

Theorem 2.12. *If V and W are finite-dimensional and $f : V \to W$ is linear, then*

$$\mathrm{rank}(f) + \mathrm{nullity}(f) = \dim(V)\,.$$

Suppose that $f : V \to W$ is a linear map, and we are given bases $S = \{v_1, \ldots, v_n\}$ for V and $S' = \{w_1, \ldots, w_m\}$ for W. For $j = 1, \ldots, n$, let $(a_{1j}, \ldots, a_{mj})^t$ be the coordinates of $f(v_j) \in W$. Define the **matrix** of f with respect to S and S' to be the $m \times n$-matrix $A_f = (a_{ij})_{i,j}$. Then the map f can be described on coordinates as "multiplication (on the left) by A_f," in the following sense:

Lemma 2.13. *If $v \in V$ has coordinates $(\alpha_1, \ldots, \alpha_n)^t$, then $f(v)$ has coordinates $A_f(\alpha_1, \ldots, \alpha_n)^t$.*

The proof is straightforward: A_f is defined so that $f(v_j) = \sum_{i=1}^m a_{ij} w_i$. That v has coordinates $(\alpha_1, \ldots, \alpha_n)^t$ means that $v = \sum_{j=1}^n \alpha_j v_j$. Therefore

$$f(v) = \sum_{j=1}^n \alpha_j f(v_j) = \sum_{i,j} a_{ij} \alpha_j w_i$$

has ith coordinate $\sum_{j=1}^{n} a_{ij}\alpha_j$.

Example 2.14. For $f : V \to W$ as in Example 2.2, the matrix of f with respect to the standard basis is just A.

If we fix bases for V and W, then each matrix $A \in M_{m,n}(\mathbb{R})$ uniquely determines a linear map $f : V \to W$ whose matrix is A (the proof is an exercise). Recall that the choices of bases determine isomorphisms $\phi : \mathbb{R}^n \to V$ and $\psi : \mathbb{R}^m \to W$ (where $\phi(\mathbf{x})$ is the vector in V with coordinates are $\mathbf{x}$, and $\phi(\mathbf{y})$ is the vector in W with coordinates $\mathbf{y}$). If f has matrix A, then $\psi^{-1} \circ f \circ \phi$ is the map sending $\mathbf{x}$ to $A\mathbf{x}$.

Proposition 2.15. *If U, V and W are vector spaces with bases $\{u_1, \ldots, u_k\}$, $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$, and $f : V \to W$ and $g : U \to V$ are linear maps, then $A_{f\circ g} = A_f A_g$.*

Proof. If $u \in U$ has coordinates $\mathbf{x} = (\alpha_1, \ldots, \alpha_k)^t$ (with respect to the chosen basis), then $g(u)$ has coordinates $A_g\mathbf{x}$, $f(g(u))$ has coordinates $A_f(A_g\mathbf{x}) = (A_f A_g)\mathbf{x}$. Therefore $A_f A_g$ is the matrix of $f \circ g$. $\square$

Note that in the proposition, we are assuming the matrices are always with the respect to the chosen bases for U, V and W. Choosing different bases will give different matrices:

Proposition 2.16. *Suppose that S and S' are bases for V, and T and T' are bases for W. If a linear map $f : V \to W$ has matrix A_f with respect to S and T, then the matrix of f with respect to S' and T' is $A'_f = QAP^{-1}$, where P is the transition matrix from S to S' and Q is the transition matrix from T to T'.*

Proof. If $v \in V$ has coordinates $\mathbf{x} = (\alpha_1, \ldots, \alpha_k)^t$ with respect to S', then v has coordinates $P^{-1}\mathbf{x}$ with respect to S, so $f(v)$ has coordinates $A_f P^{-1}\mathbf{x}$ with respect to T, and QA_fP^{-1} with respect to T'. $\square$

Example 2.17. Let $A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$ and define $f : \mathbb{R}^2 \to R^2$ by $f(\mathbf{x}) = A\mathbf{x}$. Then the matrix of A with respect to the standard basis S is just A. The matrix of A with respect to the basis $S' = \{(2,1)^t, (3,1)^t\}$ is given by $A' = P^{-1}AP$, where $P = \begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix}$ is the transition matrix *from S' to S*, so

$$A' = \begin{pmatrix} -1 & 3 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 7 \\ -3 & -4 \end{pmatrix}.$$

Recall that if $f : V \to W$ is a linear map, then $\ker(f) = \{\, v \in V \mid f(v) = 0 \,\}$ is a subspace of V of dimension nullity(f), and $\mathrm{im}(f) = \{\, f(v) \in W \mid v \in V \}$ is a subspace of W of dimension rank(f). If $f : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $f(\mathbf{x}) = A\mathbf{x}$ (see Example 2.9), then we also write nullity(A) for the nullity of f (the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$) and rank(A) for the rank of f (the dimension of the span in $\mathbb{R}^m$ of the columns of A, so this is the "column-rank" of A, but we'll see shortly this is the same as the "row-rank").

Theorem 2.12 says that

$$\dim(V) = \operatorname{rank}(f) + \operatorname{nullity}(f).$$

(assuming V and W are finite-dimensional). We know that:

- f is injective if and only if $\operatorname{nullity}(f) = 0$;
- f is surjective if and only if $\operatorname{rank}(f) = \dim(W)$;
- so f is an isomorphism if and only if $\operatorname{nullity}(f) = 0$ and $\operatorname{rank}(f) = \dim(W)$.

If follows easily that:

Proposition 2.18. *If $f : V \to W$ is an isomorphism then* $\dim(V) = \dim(W)$. *If* $\dim(V) = \dim(W)$, *then the following are equivalent:*

(1) *f is injective;*
(2) *f is surjective;*
(3) *f is an isomorphism.*

An $n \times n$-matrix A is invertible if and only there is an $n \times n$-matrix B such that $AB = BA = I_n$. Because of the correspondence between matrices and linear maps, A is invertible if and only if the associated linear map $f : \mathbb{R}^n \to \mathbb{R}^n$ (sending $\mathbf{x}$ to $A\mathbf{x}$) is an isomorphism. So in that context Prop. 2.18 says that:

$$A \text{ is invertible } \Leftrightarrow \operatorname{nullity}(A) = 0 \Leftrightarrow \operatorname{rank}(A) = n.$$

(Such an A is also called **non-singular**.)

Composing with an isomorphism doesn't change rank or nullity:

Lemma 2.19. *Suppose that $f : V \to W$ and $g : U \to V$ are linear maps of finite-dimensioanl vector spaces.*

(1) *If f is an isomorphism, then*

$$\operatorname{nullity}(f \circ g) = \operatorname{nullity}(g) \quad \textit{and} \quad \operatorname{rank}(f \circ g) = \operatorname{rank}(g).$$

(2) *If g is an isomorphism, then*

$$\operatorname{nullity}(f \circ g) = \operatorname{nullity}(f) \quad \textit{and} \quad \operatorname{rank}(f \circ g) = \operatorname{rank}(f).$$

Proof. 1) We show that $\ker(f \circ g) = \ker(g)$. If $u \in \ker(f \circ g)$, then $f(g(u)) = 0$. Since f is injective, this implies that $g(u) = 0$, so $u \in \ker(g)$. On the other hand if $u \in \ker(g)$, then $g(u) = 0$ so $f(g(u)) = 0$ and $u \in \ker(f \circ g)$. It follows that $\operatorname{nullity}(f \circ g) = \operatorname{nullity}(g)$. The assertion about ranks follows from Thm. 2.12. (Note that in this part we only needed that f is injective.)

2) Similarly to 1), we see that since g is surjective, $\operatorname{im}(f \circ g) = \operatorname{im}(f)$, and we get the asserion about ranks. The assertion about nullity then follows from Thm. 2.12 (and the fact that $\dim(U) = \dim(V)$ since g is an isomorphism). □

Applying this to multiplication by matrices, we see that if $A \in M_{m,n}(\mathbb{R})$, and P is an invertible $n \times n$-matrix and Q is an invertible $m \times m$-matrix, then A has the same rank and nullity as QAP. Note also that the rank and nullity of $f : V \to W$ are the same as those of its matrix A_f (for any choice of bases for V and W). Lemma 2.10 shows that we can choose the bases so that A_f has a very simple form:

Lemma 2.20. *Suppose that $f : V \to W$ is a linear map of finite-dimensional vector spaces. Then there are bases $\{v_1, v_2, \ldots, v_n\}$ of V and $\{w_1, w_2, \ldots, w_m\}$ of W such that the matrix of f with respect to the chosen bases has the partitioned matrix form*

$$A_f = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

where $r = \mathrm{rank}(f)$, I_r is the $r \times r$ identity matrix and $\mathbf{0}$ denote zero matrices of appropriate sizes. (If $r = m$ then the bottom row is absent; if $r = n$ then the right hand column is absent; if $r = 0$ then A_f is the zero matrix.)

Proof. Let $S = \{u_1, \ldots, u_r, v_1, \ldots, v_k\}$ be the basis for V in Lemma 2.10. So $\{w_1, \ldots, w_r\}$ (where each $w_i = f(u_i)$) is a basis for $\mathrm{im}(f)$, and $\{v_1, \ldots, v_k\}$ is a basis for $\ker(f)$. Since $\{w_1, \ldots, w_r\}$ is linearly independent, it can be extended to a basis $T = \{w_1, \ldots, w_m\}$ for W (by Thm. 1.26). Now $f(u_1)$ has coordinates $(1, 0, \ldots, 0)^t$ with respect to T, $f(u_2)$ has coordinates $(0, 1, 0, \ldots, 0)^t \ldots$ and $f(u_r)$ has coordinates $(0, 0, \ldots, 0, 1, 0, \ldots, 0)^t$ (with a 1 in the rth coordinate). Also, $f(v_1) = \ldots = f(v_k)$ have coordinates $(0, \ldots, 0)$, so the matrix of f is exactly what we want. $\square$

Corollary 2.21. *If $A \in M_{m,n}(\mathbb{R})$, then there are invertible matrices $P \in M_{n,n}(\mathbb{R})$ and $Q \in M_{m,m}(\mathbb{R})$ such that*

$$A = Q \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P.$$

Recall the definition and a few basic facts about transpose matrices: If $A = (a_{ij})_{i,j}$ is an $m \times n$-matrix, then A^t denotes its $n \times m$ **transpose** matrix whose (j, i)-entry is $a_{i,j}$. I.e.,

$$\text{If } A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \text{ then } A^t = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix},$$

In other words the rows of A^t are the columns of A and vice-versa.

If $A \in M_{m,n}(\mathbb{R})$ and $B \in M_{n,k}(\mathbb{R})$, then $(AB)^t = B^t A^t$. If P is an invertible $n\times$-matrix, then so is P^t (since if $PQ = QP = I_n$, then $P^t Q^t = Q^t P^t = I_n$).

Note that the span (in $\mathbb{R}^n$) of the rows of A is the same as the span of the columns of A^t, so the row-rank of A is the same as the (column-)rank of A^t.

Corollary 2.22. $\mathrm{rank}(A) = \mathrm{rank}(A^t)$*, so the dimension of the span of the columns of A is the same as the dimension of the span of the rows of A.*

Proof. Write $A = Q \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P$ as in Corollary 2.21. Then $A^t = P^t \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^t$. Since P^t and Q^t are invertible, we get $\mathrm{rank}(A^t) = r = \mathrm{rank}(A)$. □

3. Equations and Matrices

Recall that if A is an $m \times n$-matrix, then $A\mathbf{x} = \mathbf{0}$ defines a homogeneous system of linear equations (with m equations in n unknowns). The set of solutions is precisely the kernel of the corresponding linear map from $f : \mathbb{R}^n \to \mathbb{R}^m$. So Theorem 2.12 gives:

Theorem 3.1. [Homogeneous Linear Equations.] *Let A be an $m \times n$ matrix. The homogenous system of linear equations $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution if and only if* $\mathrm{rank}(A) < n$. *The dimension of the solution space is* $n - \mathrm{rank}(A)$.

Consequently, if $m < n$ (i.e. if there are more unknowns than equations) then there is always a non-trivial solution.

Recall the usual method for finding the space of solutions is to apply elementary row operations to put A into Row Echelon Form, from which the solution space can be read off. Recall the three types of elementary row operations:

(I) Interchange two rows.
(II) Multiply one row by a non-zero scalar.
(III) Add a scalar multiple of one row to another row.

Example 3.2. Let $A = \begin{pmatrix} 0 & 1 & -1 & -1 & -1 \\ -1 & 1 & -3 & 0 & 2 \\ 1 & 0 & 2 & -1 & -3 \\ 2 & 1 & 3 & 0 & -1 \end{pmatrix}$. Denoting the ith row by R_i, apply the row operation exchanging R_1 with R_2, and then multiply (the new) R_1 by -1 to get $\begin{pmatrix} 1 & -1 & 3 & 0 & -2 \\ 0 & 1 & -1 & -1 & -1 \\ 1 & 0 & 2 & -1 & -3 \\ 2 & 1 & 3 & 0 & -1 \end{pmatrix}$. Then (sometimes combining row elementary operations to save writing...)

$$\xrightarrow[R_4 \to R_4 - 2R_1]{R_3 \to R_3 - R_1,} \begin{pmatrix} 1 & -1 & 3 & 0 & -2 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 3 & -3 & 0 & 3 \end{pmatrix} \xrightarrow[R_4 \to R_4 - 3R_2]{R_3 \to R_3 - R_2,} \begin{pmatrix} 1 & -1 & 3 & 0 & -2 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 6 \end{pmatrix}$$

$$\xrightarrow[R_4 \to \frac{1}{3}R_4]{R_3 \leftrightarrow R_4,} \begin{pmatrix} 1 & -1 & 3 & 0 & -2 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow[R_1 \to R_1 + R_2]{R_2 \to R_2 + R_3} \begin{pmatrix} 1 & 0 & 2 & 0 & -1 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The resulting matrix defines a system of equations equivalent to the original one, whose solutions have x_3 and x_5 arbitrary, with

$$x_1 = -2x_3 + x_5, \quad x_2 = x_3 - x_5, \quad x_4 = -2x_5.$$

We can therefore describe the general solution in the form

$$\mathbf{x} = \alpha \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ -1 \\ 0 \\ -2 \\ 1 \end{pmatrix}.$$

The solution space therefore has these two vectors as a basis. (Note that this procedure computes the rank and nullity as well. In this case the rank is 3 and nullity is 2.)

To see that row operations yield an equivalent system of equations, we can make the following observation: the effect of applying a row operation to a matrix A is the same as that multiplying A on the left by the corresponding **elementary matrix**:

(I) $E_I(p,q)$ *(interchanging row p and row q)*: all the diagonal entries of $E_I(p,q)$ except the p^{th} and the q^{th} are 1, the $(p,q)^{\text{th}}$ and the $(q,p)^{\text{th}}$ entries are 1 and all the remaining entries are 0.

(II) $E_{II}(p,\lambda)$ *(multiplying row p by a non-zero scalar λ)*: all the diagonal entries of $E_{II}(p,\lambda)$ except the p^{th} are 1, the $(p,p)^{\text{th}}$ entry is λ and all the remaining entries are 0.

(III) $E_{III}(p,q,\lambda)$ *(adding λ times row p to row q)*: all the diagonal entries of $E_{III}(p,q,\lambda)$ are 1, the $(q,p)^{\text{th}}$ entry is λ and all the remaining entries are 0.

Note that each of these matrices is invertible: The inverses of $E_I(p,q)$, $E_{II}(p,\lambda)$ and $E_{III}(p,q,\lambda)$ are $E_I(p,q)$, $E_{II}(p,\lambda^{-1})$ and $E_{III}(p,q,-\lambda)$ respectively.

Therefore the effect of applying a sequence of elementary row operations to A is to replace A by $E_kE_{k-1}\cdots E_2E_1A$ where $E_1,\ldots,E_k$ are the corresponding elementary matrices. Thus we are replacing A by QA where $Q = E_kE_{k-1}\cdots E_2E_1$ is invertible. Note that if $A\mathbf{x} = \mathbf{0}$ then $QA\mathbf{x} = \mathbf{0}$, and if $QA\mathbf{x} = \mathbf{0}$ then $A\mathbf{x} = Q^{-1}QA\mathbf{x} = \mathbf{0}$. So the two systems of equations have the same solutions.

Similarly column operations correspond to multiplying A on the *right* by elementary matrices:

(I) $E_I(p,q)$ interchanges column p and column q;

(II) $E_{II}(p,\lambda)$ column p by λ;

(III) $E_{III}(p,q,\lambda)$ adds λ times column q to column p.

Recall that the image of the map $f : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(\mathbf{x}) = A\mathbf{x}$ is given by the span of the columns of A, and this is the set of $\mathbf{b} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{b}$ has solutions. Applying column operations amounts to multiplying A on the right by an invertible matrix P, and $A\mathbf{x} = \mathbf{b}$ has solutions if and only if $AP\mathbf{y} = \mathbf{b}$ has solutions (set $\mathbf{x} = P\mathbf{y}$ and $\mathbf{y} = P^{-1}\mathbf{x}$), This can also be thought of as saying that column operations don't change the span of the columns. So for example, you could use column operations to find a basis for the image of a linear map.

Example 3.3. To find a basis for the image of the map $f : \mathbb{R}^3 \to \mathbb{R}^4$ defined by the matrix $\begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & 2 \\ 2 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}$ applying *column* operations gives

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 3 & 3 \\ 2 & -3 & -3 \\ -1 & 2 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ -1 & 3 & 0 \\ 2 & -3 & 0 \\ -1 & 2 & 0 \end{pmatrix},$$

so the rank of f is two, and a basis for the image is $\{(1, -1, 2, -1)^t, (0, 3, -3, 2)^t\}$. Note that *row* operations could be used to compute the rank of A, but that would change the span of the columns of A (while preserving the span of its rows).

Note that $A\mathbf{x} = \mathbf{b}$ has solutions (i.e., $\mathbf{b}$ is in the span of the columns of A) if and only if the rank of the augmented matrix $(A|\mathbf{b})$ is the same as the rank of A. One can use row operations on this augmented matrix to find the set of solutions to the system. We use the augmented matrix $(\,A\,|\mathbf{b})$ to represent the system, and apply row operations to obtain an augmented matrix $(\,A'\,|\mathbf{b}')$ with A' in row echelon form. Note that the new matrix represents an equivalent system since applying the same series of row operations to both A and $\mathbf{b}$ amounts to multiplying them both on the left by the same invertible matrix, say Q, and the equation $A\mathbf{x} = \mathbf{b}$ is equivalent to the equation $QA\mathbf{x} = Q\mathbf{b}$. In particular if A' is in row echelon form, one can immediately tell whether $A'\mathbf{x} = \mathbf{b}' = (b'_1, \dots, b'_m)$ has a solution, according to whether $b'_i = 0$ for all $i > r$, where r is the rank of A'. Furthermore if there are any solutions, they can easily be read off from $(\,A'\,|\mathbf{b}')$.

Example 3.4. Consider the system $A\mathbf{x} = (1, 2, -1, -1)^t$ where A is as in Example 3.2. The same sequence of row operations applied now to the augmented matrix

$$\left(\begin{array}{ccccc|c} 1 & -1 & 3 & 0 & -2 & 1 \\ 0 & 1 & -1 & -1 & -1 & 2 \\ 1 & 0 & 2 & -1 & -3 & -1 \\ 2 & 1 & 3 & 0 & -1 & -1 \end{array}\right) \to \cdots \to \left(\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & -1 & -1 \\ 0 & 1 & -1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The solutions are therefore given by

$$x_1 = -1 - 2x_3 + x_5, \quad x_2 = 1 + x_3 - x_5, \quad x_4 = -2x_5$$

with x_3, x_5 arbitrary. We can therefore describe the general solution as

$$\mathbf{x} = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ -1 \\ 0 \\ -2 \\ 1 \end{pmatrix}.$$

Note the structure of the set of solutions of a non-homogeneous system. If $\mathbf{x}_0$ is a fixed solution, then $A\mathbf{x} = \mathbf{b} = A\mathbf{x}_0$ if and only if $A(\mathbf{x} - \mathbf{x}_0) = \mathbf{0}$, if and only if $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$ for some solution $\mathbf{y}$ of the *homogeneous* system $A\mathbf{y} = \mathbf{0}$. We sum this up as follows:

Theorem 3.5. [Non-Homogeneous Linear Equations.] *Let A be an $m \times n$ matrix and $\mathbf{b}$ be a non-zero vector in $\mathbb{R}^m$. The non-homogenous system of linear equations $A\mathbf{x} = \mathbf{b}$ has a solution if and only if* $\operatorname{rank}(A) = \operatorname{rank}(A|\mathbf{b})$ *where $(A|\mathbf{b})$ is the $m \times (n+1)$ augmented matrix obtained by adjoining $\mathbf{b}$ to A as the $(n+1)^{\text{th}}$ column. If a solution $\mathbf{x_0}$ to the system exists then every solution is of the form $\mathbf{x_0} + \mathbf{y}$ where y is any solution of the homogeneous system $A\mathbf{y} = \mathbf{0}$. In particular, the system has a unique solution if and only if* $\operatorname{rank}(A) = \operatorname{rank}(A|\mathbf{b}) = n$.

Row operations can also be used to find the inverse of a matrix of an invertible $n \times n$-matrix A. Since A has rank n, we get it into the row echelon form I_n by row operations. So starting with the augmented matrix $(A|I_n)$ and applying row operations yields $(I_n|B)$, where $QA = I_n$ and $QI_n = B$. Therefore $Q = B$, and so $BA = I_n$ and $B = A^{-1}$.

Example 3.6. To find the inverse of the matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, apply row operations to the augmented matrix $(\, A \,|\, I_n \,)$ giving

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & -1 & -2 & -1 & 1 & 0 \end{array}\right)$$

$$\to \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -2 & -1 & 1 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array}\right)$$

$$\to \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 1 & 0 & -2 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{2} & \frac{3}{2} & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array}\right)$$

giving $A^{-1} = \begin{pmatrix} -\frac{1}{2} & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & 1 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$.

For another algorithmic application of row and column operations, suppose we are given a linear map $f : V \to W$ of finite-dimensional vector spaces, and we want to find bases for V and W with respect to which the matrix of f is as in Lemma 2.20. If A is the matrix of f with respect to given bases for V and W, then this amounts to finding invertible matrices Q and P so that

$$QAP = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

(the transition matrices to the desired bases being P^{-1} and Q). To achieve this, first apply row operations to the augmented matrix $(A|I_n)$ to obtain a matrix $(A'|Q)$ where $A' = QA$ is in row echelon form. Next consider the vertically augmented matrix $\left(\frac{A'}{I_m}\right)$. Applying *column* operations, it is easy to see we can get the top matrix into the desired form $A'' = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. Since applying column operations amounts to multiplying the top and bottom parts of the augmented matrix by the same invertible matrix P, we get that the resulting augmented matrix is $\left(\frac{A''}{P}\right)$ where $A'' = A'P = QAP$. Note that there are (infinitely) many possible P and Q such that QAP has the desired form. For example, if $A = \mathbf{0}$ then any invertible P and Q work; if A is square and invertible, then given a P and Q which work (so $QAP = I_n$), we can choose any invertible R and replace Q by RQ and P by PR^{-1}.

Example 3.7. Suppose $A = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix}$. Subtracting $2R_1$ from R_2 gives

$$\left(\begin{array}{ccc|cc} 1 & 0 & -1 & 1 & 0 \\ 2 & 1 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|cc} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 2 & -2 & 1 \end{array}\right)$$

so we can let $A' = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}$ and $Q = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$. Then applying column operations gives

$$\left(\begin{array}{ccc} 1 & 0 & -1 \\ 0 & 1 & 2 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 2 \\ \hline 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{array}\right),$$

so we have $P = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$.

4. Determinants

We recall the definition of the determinant and review its main properties. If A is a square matrix, then the determinant of A is a certain scalar associated to A, denoted $\det(A)$ (or simply $|A|$). We can define the determinant inductively by first defining the determinant of any 1×1-matrix. Then we define the determinant of an $n \times n$-matrix assuming we have already defined the determinant of any $(n-1) \times (n-1)$-matrix.

Definition 4.1. The **determinant** of an $n \times n$-matrix $A = (a_{ij})$ is defined as follows.

(1) The **determinant** of the 1×1 matrix (a) is a.
(2) Suppose that determinants have been defined for $(n-1) \times (n-1)$-matrices. For integers i, j with $1 \le i \le n$, $1 \le i \le n$ the (i,j)-**minor** of A is defined to be the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the i^{th} row and the j^{th} column of A.
(3) The **determinant** of A is defined by

$$\det(A) = a_{11}M_{11} - a_{12}M_{12} + \ldots + (-1)^{n+1}a_{1n}M_{1n} = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} M_{1j} \,,$$

where M_{1j} denotes the $(1,j)$-minor of A. We also write $|A|$ for $\det(A)$.

For example if $n = 2$, then $M_{11} = \det(a_{22}) = a_{22}$ and $M_{22} = \det(a_{21}) = a_{21}$, so we get the usual formula $\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}$.

For $n = 3$, we have (using the $|\ |$ notation)

$$M_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{23}a_{32},$$

$$M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = a_{21}a_{33} - a_{23}a_{31}, \; ,$$

$$M_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = a_{21}a_{32} - a_{22}a_{31}.$$

so

$$\det(A) = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.$$

Note that this coincides with the description of the determinant of a 3×3-matrix given by repeating the first two columns of A after the matrix:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$$

forming the products along each diagonal and adding them with appropriate signs ($+$ for the three $\searrow$-sloping diagonals, $-$ for the three $\nearrow$-sloping diagonals).

Example 4.2. The determinant of $A = \begin{pmatrix} 2 & -1 & 1 \\ -3 & 4 & -1 \\ 1 & 6 & 2 \end{pmatrix}$ is $16 + 1 - 18 - 4 + 12 - 6 = 1$.

For $n = 4$, the definition of the determinant yields an expression with 24 terms, but fortunately there's a more practical way to compute the determinant of a large matrix using the effect of row operations on the determinant, summed up as follows:

Proposition 4.3. *Suppose that A is an $n \times n$-matrix.*

(1) *Exchanging two rows of A multiplies* $\det A$ *by* -1.
(2) *Multiplying a row of A by $\lambda \in \mathbb{R}^\times$ multiplies* $\det A$ *by* λ.
(3) *Adding a multiple of one row to another does not change* $\det A$.

We will prove the proposition later, after we explain some of its many consequences. First we recall how it can be used in practice to compute determinants. Simply apply row operations, keeping track of their effect on the determinant, until the matrix is upper-triangular, and we know the determinant of an upper-triangular matrix is the product of its diagonal entries (see the exercises).

Example 4.4. We give an example comuting the determinant of a 4×4-matrix:

$$\begin{vmatrix} 1 & 2 & -1 & 4 \\ 2 & 0 & -1 & 3 \\ -1 & 1 & 0 & 1 \\ 4 & -1 & 2 & -1 \end{vmatrix} = \begin{vmatrix} 1 & 2 & -1 & 4 \\ 0 & -4 & 1 & -5 \\ 0 & 3 & -1 & 5 \\ 0 & -9 & 6 & -17 \end{vmatrix} = \begin{vmatrix} 1 & 2 & -1 & 4 \\ 0 & -1 & 0 & 0 \\ 0 & 3 & -1 & 5 \\ 0 & -9 & 6 & -17 \end{vmatrix}$$

(by a series of row operations of type III, the last one being $R_2 \to R_2 + R_3$)

$$= \begin{vmatrix} 1 & 2 & -1 & 4 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 5 \\ 0 & 0 & 6 & -17 \end{vmatrix} = \begin{vmatrix} 1 & 2 & -1 & 4 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 5 \\ 0 & 0 & 0 & 13 \end{vmatrix} = (-1)(-1)13 = 13.$$

We now deduce some general consequences of the proposition (which also follows easily from the definition of the determinant).

Lemma 4.5. *If A has a row of zeroes, then* $\det(A) = 0$.

Proof. Multiplying the zero row by any λ doesn't change A, so part (2) of Prop. 4.3 gives $\lambda \det A = \det A$ for all $\lambda \in \mathbb{R}^\times$, so $\det A = 0$. ☐

The following is not at all easily seen from the definition, but follows fairly quickly from Lemma 4.3.

Theorem 4.6. *If A and B are $n \times n$-matrices, then* $\det(AB) = \det(A)\det(B)$

Proof. Recall that applying a row operation to B gives the matrix EB where E is the corresponding elementary matrix, so the proposition says that

(I) if $E = E_I(p, q)$, then $\det(EB) = -\det(B)$;
(II) if $E = E_{II}(p, \lambda)$, then $\det(EB) = \lambda \det(B)$;
(III) if $E = E_{III}(p, q, \lambda)$, then $\det(EB) = \det(B)$.

In particular, taking $B = I_n$ gives $\det(E) = -1$, λ or 1, according to whether is of type I, II or III. It follows that $\det(EB) = \det(E)\det(B)$ for all B.

Recall that if A is invertible, then A can be written as a product $E_1E_2\cdots E_k$ of elementary matrices. It follows then that

$$\begin{aligned}\det(A) &= \det(E_1E_2\cdots E_k) = \det(E_1)\det(E_2\cdots E_k)\\ &= \cdots = \det(E_1)\det(E_2)\cdots\det(E_k).\end{aligned}$$

(In particular, note that $\det(A)\neq 0$ if A is invertible.) Similarly

$$\begin{aligned}\det(AB) &= \det(E_1E_2\cdots E_kB) = \det(E_1)\det(E_2\cdots E_kB)\\ &= \cdots = \det(E_1)\det(E_2)\cdots\det(E_k)\det(B) = \det(A)\det(B).\end{aligned}$$

If A is not invertible, then the associated row echelon matrix can be written as QA for some invertible matrix Q. Since $\mathrm{rank}(A) < n$, we know that QA has a zero row, so by Prop. 4.5 we have $\det(QA) = 0$. Since Q is invertible $\det(Q)\neq 0$ and $\det(QA) = \det(Q)\det(A)$, so $\det(A) = 0$. Note also that AB is not invertible, so the same argument shows $\det(AB) = 0$, and therefore $\det(AB) = \det(A)\det(B)$ in this case as well. $\square$

In the course of the proof of Theorem 4.6, we saw also:

Theorem 4.7. $\det A \neq 0$ *if and only if* A *is invertible.*

Corollary 4.8. $\det(A^t) = \det(A)$.

Proof. If A is invertible, then $A = E_1\cdots E_k$ for some invertible matrices $E_1\cdots E_k$. If E is an elementary matrix, then E^t is an elementary matrix of the same type and $\det(E) = \det(E^t)$. Since $A^t = E_k^t\cdots E_1^t$, it follows that

$$\det(A) = \det(E_1)\cdots\det(E_k) = \det(E_k)^t\cdots\det(E_1^t) = \det(A^t).$$

If A is not invertible, then neither is $\det(A^t)$ so $\det(A) = \det(A^t) = 0$. $\square$

Since applying column operations amounts to multiplying (on the right) by an elementary matrix, we get:

Corollary 4.9. *Suppose that* A *is an* $n\times n$*-matrix.*

(1) *Exchanging two columns of* A *multiplies* $\det A$ *by* -1.
(2) *Multiplying a column of* A *by* $\lambda\in\mathbb{R}^\times$ *multiplies* $\det A$ *by* λ.
(3) *Adding a multiple of one column to another does not change* $\det A$.

Corollary 4.10. *If* $A = (a_{ij})$ *is upper-triangular or lower-triangular, then* $\det A = a_{11}a_{22}\cdots a_{nn}$.

Proof. This was an exercise if A is upper-trangular. If A is lower-triangular, then A^t is upper-triangular with the same diagonal entries as A. The formula then follows from Prop. 4.8. $\square$

The rest of this section consists of the proof of Prop. 4.3, and an alternate definition of the determinant; these are included for the sake of completeness, but are not required for the exam.

Rather than prove the three properties in Prop. 4.3 directly, we will show they follow from three related properties, and prove those instead:

Proposition 4.11. *Suppose that* A *is an* $n\times n$*-matrix.*

(I') *If A has two identical rows, then* $\det A = 0$.
(II') *Multiplying a row of A by $\lambda \in \mathbb{R}$ multiplies* $\det A$ *by λ.*
(III') *Suppose that*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad A' = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a'_{i1} & a'_{i2} & \cdots & a'_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$
$$\textit{and} \quad A'' = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a''_{i1} & a''_{i2} & \cdots & a''_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$

are identical except in ith row, and that the ith row of A'' is the sum of the ith rows of A and A' (i.e., $a''_{ij} = a_{ij} + a'_{ij}$ for $j = 1, \dots, n$). Then $\det(A'') = \det(A) + \det(A')$.

Note that in (III'), A'' is *not* the sum of A and A'; it is only the ith row that is being described as a sum. Parts (II') and (III') taken together can be thought of as saying that det is *linear on each row.*

Before proving Prop. 4.11, we show how Prop. 4.3 follows from it. Clearly (II') implies (II).

Next we prove (III). Let A' be the matrix identical to A, except that its qth row is replaced λ times the pth row of A. Then by (II') and (I'), we have

$$|A'| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \\ \vdots & \vdots & & \vdots \\ \lambda a_{p1} & \lambda a_{p2} & \cdots & \lambda a_{pn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \lambda \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = 0.$$

Applying (III'), we see therefore that $\det(A'') = \det(A)$ where A'' is gotten by applying the row operation of type (III) to A.

We now prove (I'). Let A' be the matrix gotten from A by interchanging rows p and q. Note that by (III) (which we have now proved), we have

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1}+a_{q1} & a_{p2}+a_{q2} & \cdots & a_{pn}+a_{qn} \\ \vdots & \vdots & & \vdots \\ a_{q1} & a_{q2} & \cdots & a_{qn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}, \quad |A'| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1}+a_{q1} & a_{p2}+a_{q2} & \cdots & a_{pn}+a_{qn} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

These two matrices are identical expect in the qth row, and the sum of their qth rows is the same as the pth row, so by (III') and (I'), we have $\det(A) + \det(A') = 0$. Therefore $\det(A') = -\det(A)$.

Proof. (of Prop. 4.11) We proceed by induction on n.

For $n = 1$, note that there is nothing to prove for (I'), that (II') just says that $\lambda \det(a) = \lambda a = \det(\lambda a)$, and (III') says that $\det(a + a') = a + a' = \det(a) + \det(a')$.

Suppose now that $n > 1$, and that Prop. 4.11 (and therefore also Prop. 4.3) hold with n replaced by $n - 1$.

First we prove (II'). Let A' be the matrix gotten by mutliplying the ith row of A by λ. If $i = 1$, then the definition of the determinant gives

$$\det(A') = (\lambda a_{11})M_{11} - (\lambda a_{12})M_{12} + \cdots + (-1)^{n+1}(\lambda a_{1n})M_{1n} = \lambda \det(A).$$

If $2 \leq i \leq n$, then we get

$$\det(A') = (\lambda a_{11})M'_{11} - (\lambda a_{12})M'_{12} + \cdots + (-1)^{n+1}(\lambda a_{1n})M'_{1n},$$

where M'_{1j} is the $(1, j)$-minor of A'. However M'_{1j} is the determinant of the same $(n-1) \times (n-1)$ matrix as in the definition of M_{1j}, except that one of the rows has been multiplied by λ. So by the induction hypothesis $M'_{1j} = \lambda M_{1j}$ for $j = 1, \ldots, n$, and it follows that $\det(A') = \lambda \det(A)$.

Next we prove (III'). If $i = 1$, then

$$\begin{array}{rcl} \det(A'') & = & a''_{11}M_{11} - a''_{12}M_{12} + \cdots + (-1)^{n+1}a''_{1n}M_{1n} \\ & = & (\lambda a_{11} + a'_{11})M_{11} - (a_{12} + a'_{12})M_{12} + \cdots + (-1)^{n+1}(a_{1n} + a_{1n'})M_{1n} \\ & = & \det(A) + \det(A'). \end{array}$$

If $2 \leq i \leq n$, then

$$\det(A'') = a_{11}M''_{11} - a_{12}M''_{12} + \cdots + (-1)^{n+1}a_{1n}M''_{1n},$$

where M''_{1j} is the $(1, j)$-minor of A'', which is the determinant of an $(n-1) \times (n-1)$-matrix which is identical to the one in the definition of hte minors of A and A' except in one row, where it is their sum. So it follow from the induction hypothesis that $M''_{1j} = M_{1j} + M_{1j'}$ for each j, and therefore $\det(A'') = \det(A) + \det(A')$.

Finally we prove (I'). This is easy if $n = 2$, so assume $n > 2$. Suppose that the first two rows of A are identical, so $a_{1j} = a_{2j}$ for $j = 1, \ldots, n$. If $1 \leq j < k \leq n$, then let N_{jk} denote the denote the determinant of the

$(n-2) \times (n-2)$-matrix gotten by deleting the first two rows of A, and the jth and kth columns of A. Applying the definition of the determinant to the minors M_{1j} in the definition of $\det(A)$, we see that each M_{1j} is the alternating sum

$$a_{21}N_{1j}-a_{22}N_{2j}+\cdots+(-1)^j a_{2,j-1}N_{j-1,j}+(-1)^{j+1}a_{2,j+1}N_{j,j+1}+\cdots+(-1)^n a_{2n}N_{jn}.$$

In the resulting expansion of $\det(A)$, each N_{jk} (where $j < k$) will appear twice: once in the expansion of M_{1j} with coefficient $(-1)^{j+1}(-1)^k a_{1j}a_{2k}$, and once in the expansion of M_{1k} with coefficient $(-1)^{k+1}(-1)^{j+1}a_{1k}a_{2j}$. Since we are assuming the first two rows are identical, we have $a_{1j} = a_{2j}$ and $a_{1k} = a_{2k}$. Since the signs are opposite, these two terms cancel. Thus all terms cancel, and it follows that $\det(A) = 0$.

Now suppose instead that row 1 and row q are identical for some $q > 2$. Let A' be the matrix gotten from A by interchanging rows 2 and q. Then

$$\det(A') = a_{11}M'_{11} - a_{12}M'_{12} + \cdots + (-1)^{n+1}a_{1n}M'_{1n},$$

where each $M_{1j'}$ is the determinant of the same matrix as in the definition of M_{1j}, but with two rows interchanged. By Prop. 4.3 (I) for $n-1$, we therefore have $M'_{1j} = -M_{1j}$, and it follows that $\det(A) = -\det(A')$. But we have already shown that $\det(A') = 0$, so $\det(A) = 0$ as well.

Finally suppose that row p and row q are identical, and p and q are both greater than 1. Then the minors M_{1j} are determinants of $(n-1) \times (n-1)$-matrices, each of which has two identical rows. So by the induction hypothesis, each $M_{1j} = 0$, and therefore $\det(A) = 0$.

This finishes the proof of Prop. 4.11, and thus the proof of all the stated properties of the determinant. $\square$

Finally, we remark that there is an alternative definition of the determinant using permutations. This is also provided for your interest and will not be covered on the examination.

The second definition is given purely for your general interest and will not be covered in the course. It is somewhat more sophisticated but, once mastered, the proofs of the properties are more transparent. The definition requires some knowledge about permutations. Recall that a bijective mapping of $\{1, \ldots, n\}$ onto itself is called a *permutation*. The set of all such permutations with composition as the operation of 'product' forms a group denoted by S_n. (i.e. $\sigma\sigma' = \sigma \circ \sigma'$). Here are the main properties of permutations.

- Every permutation $\sigma \in S_n$ can be written as a product $\sigma = \epsilon_1\epsilon_2 \ldots \epsilon_s$ where each ϵ_i is a *transposition* (i.e. a permutation that exchanges two elements of $\{1, 2, \ldots, n\}$ and leaves the rest fixed).
- Although there are many such expressions for σ and the number s of transpositions varies, nevertheless, for given σ, s is either always even (in which case we say that σ is an *even* permutation) or always odd (when we say σ is *odd*). (See the exercises for a proof of this

fact using the definition and properties of the determinant we already know about.)

- We define the sign $\mathrm{sgn}(\sigma)$ of σ to be 1 if σ is even and -1 if σ is odd. (i.e. $\mathrm{sgn}(\sigma) = (-1)^s$).
- The map $\mathrm{sgn} : \sigma \mapsto \mathrm{sgn}(\sigma)$ is a *homomorphism* from S_n to the multiplicative group $\{1, -1\}$ i.e. $\mathrm{sgn}(\sigma\sigma') = \mathrm{sgn}(\sigma)\,\mathrm{sgn}(\sigma')\ \forall \sigma, \sigma' \in S_n$.

The second definition of determinant is based on the idea that each term in the evaluation of a determinant is a product of n entries with exactly one from each row and from each column. Thus from the first row a term might have a_{1n_1} where n_1 is any integer between 1 and n, the from the second a_{2n_2}, but now $n_2 \neq n_1$, and so on. It is clear that the mapping $i \mapsto n_i$ is a permutation of the integers 1 to n and that there one such term to each such permutation σ. If the permutation is even we attach a plus sign to the term $a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ and if σ is odd, we attach a minus sign. The sum of these signed terms is the determinant. Here is the formal definition:

The *determinant* of an $n \times n$ matrix $A = (a_{ij})$ is defined by

$$\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)}$$

where the sum is take over the group S_n of all permutations of the integers $\{1, 2, 3, \ldots, n\}$.

Here is a sketch of a proof that the two definitions agree: Let us provisionally call $\det'(A)$ the determinant of A as defined using permutations, and continue to use $\det(A)$ for the determinant as we originally defined it inductively. We first show that $\det'(A)$ has some of the same properties as $\det(A)$, and then explain how to deduce that $\det(A) = \det'(A)$. It is straightforward to check that $\det'(A)$ satisfies (II') and (III') in Prop. 4.11. In fact (I') is not so difficult either: let τ denote the transposition interchanging p and q. If $a_{pj} = a_{qj}$ for all j, then one sees that for any permutation σ, the terms in the permutation definition indexed by σ and $\sigma\tau$ are identical, except that they have opposite signs and therefore cancel.

Having proved $\det'(A)$ satifies these properties, one finds that $\det'(A)$ has all the same properties we proved about det. In particular, if E is elementary, we find that $\det(E) = \det'(E)$, and therefore $\det'(EA) = \det(E)\det'(A)$. If A is invertible, then we can write it as a product of elementary matrices to deduce in this case that $\det(A) = \det'(A)$. If A is not invertible, then there is an invertible matrix P so that PA has a zero row. It is easy to see that in this case $\det'(PA) = 0$, so that $\det'(A) = 0$, and so $\det'(A) = \det(A)$ in all cases.

5. Similarity and diagonalization

We now consider linear maps $f : V \to V$ from a finite-dimensional vector space V to itself, also called *linear operators* on V. We consider matrices of such maps with respect to the same basis of V as both the domain and co-domain. We then have that $A_{f \circ g} = A_f A_g$, where all matrices are with respect to the same basis for V. (Note that all the matrices are square.)

Changing the basis for V from S to S' (in both domain and co-domain) replaces the matrix A_f by $A'_f = QA_fQ^{-1}$ where P is the transition matrix from S to S'.

Definition 5.1. If A and B are $n \times n$-matrices, then we say that A is **similar** to B if there exists an $n \times n$ invertible matrix Q such that $B = QAQ^{-1}$. If A is similar to B, then we write $A \simeq B$.

So matrices representing the same linear operator, but with respect to different bases, are similar. Also, since every invertible matrix is a transition matrix, any matrix similar to A_f is the matrix of f with respect to some basis. One can check that similarity is an equivalence relation (exercise).

Example 5.2. The matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is similar to $B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ since $B = QAQ^{-1}$ with $Q = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Definition 5.3. We say that a linear map $f : V \to V$ is **diagonalizable** if the matrix of f with respect to some basis is diagonal. An $n \times n$-matrix A is **diagonalizable** if A is similar to a diagonal matrix.

A linear map f is diagonalizable if and only if A_f is diagonalizable and a matrix A is diagonalizable if and only if the corresponding map on $\mathbb{R}^n$ (defined by sending $\mathbf{v}$ to $A\mathbf{v}$) is diagonalizable (exercise).

Recall the notion of an eigenvector:

Definition 5.4. A non-zero vector $\mathbf{v} \in \mathbb{R}^n$ is an **eigenvector** for A with **eigenvalue** $\lambda \in \mathbb{R}$ if $A\mathbf{v} = \lambda\mathbf{v}$. Similarly if $f : V \to V$ is a linear map, then we say a non-zero vector $v \in V$ is an eigenvector for f with eigenvalue λ if $f(v) = \lambda v$.

Theorem 5.5. *A linear map $f : V \to V$ is diagonalizable if and only V has a basis consisting of eigenvectors for f. A matrix A is diagonalizable if and only if $\mathbb{R}^n$ has a basis consisting of eigenvectors for A.*

Proof. Suppose that f is diagonalizable. Then V has a basis $\{v_1, \ldots, v_n\}$ with respect to which

$$A_f = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

This means that $f(v_1) = \lambda_1 v_1$, $f(v_2) = \lambda_2 v_2$, $\ldots f(v_n) = \lambda_n v_n$; i.e., that $v_1, v_2, \ldots, v_n$ are eigenvectors. Conversely if V has a basis consisting of eigenvectors, then we see in the same way that the matrix of f with respect to this basis is diagonal.

Example 5.6. The matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is diagonalizable, and $\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\}$ is a basis of eigenvectors. The matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable. The matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is not diagonalizable over $\mathbb{R}$, but it is diagonalizable over $\mathbb{C}$, a basis of eigenvectors being given by $\begin{pmatrix} 1 \\ i \end{pmatrix}$ (with eigenvalue i) and $\begin{pmatrix} 1 \\ -i \end{pmatrix}$ (with eigenvalue $-i$). For the reason ilustrated by the last example, it is often more convenient to work over $\mathbb{C}$ than over $\mathbb{R}$. We'll return to this point later.

Lemma 5.7. *If $f : V \to V$ is a linear map, and $v_1, v_2, \ldots, v_k \in V$ are eigenvectors for f with distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$, then $\{v_1, \ldots, v_k\}$ is linearly independent.*

Proof. We prove the lemma by induction on k. If $k = 1$, then $\{v_1\}$ is linearly independent since $v_1 \neq 0$.

Now suppose that $k > 1$ and the lemma is true with k replaced by $k-1$. In particular $\{v_1, \ldots, v_{k-1}\}$ is linearly independent. Suppose that $\sum_{i=1}^k \alpha_i v_i = 0$. Then also $f\left(\sum_{i=1}^k \alpha_i v_i\right) = 0$, but since f is linear,

$$f\left(\sum_{i=1}^{k} \alpha_i v_i\right) = \sum_{i=1}^{k} \alpha_i f(v_i).$$

Since $f(v_i) = \lambda_i v_i$ for $i = 1, \ldots, k$, we get

$$\alpha_1 \lambda_1 v_1 + \cdots + \alpha_{k-1} \lambda_{k-1} v_{k-1} + \alpha_k \lambda_k v_k = 0.$$

On the other hand, multiplying $\sum_{i=1}^k \alpha_i v_i = 0$ by λ_k gives

$$\lambda_k \alpha_1 v_1 + \cdots + \lambda_k \alpha_{k-1} v_{k-1} + \lambda_k \alpha_k v_k = 0.$$

Subtracting one equation from the other gives

$$\alpha_1(\lambda_1 - \lambda_k) v_1 + \cdots + \alpha_{k-1}(\lambda_{k-1} - \lambda_k) v_{k-1} = 0.$$

But since $\{v_1, \ldots, v_{k-1}\}$ is linearly independent, we get

$$\alpha_1(\lambda_1 - \lambda_k) = \alpha_2(\lambda_2 - \lambda_k) = \cdots \alpha_{k-1}(\lambda_{k-1} - \lambda_k) = 0.$$

Since the λ_i are distinct, we know that $\lambda_1 - \lambda_k$, $\lambda_2 - \lambda_k$, $\ldots$, $\lambda_{k-1} - \lambda_k$ are all non-zero. Therefore $\alpha_1 = \alpha_2 = \cdots = \alpha_{k-1} = 0$. Finally, since this gives $\alpha_k v_k = 0$, and $v_k \neq 0$, we conclude that $\alpha_k = 0$ as well. $\square$

Note that there's an equivalent form of the lemma with f replaced by an $n \times n$-matrix and $v_1, \ldots, v_k$ replaced vectors in $\mathbb{R}^n$ (or $\mathbb{C}^n$).

Since a set of n linearly independent vectors in an n-dimensional space is necessarily a basis, it follows from Theorem 5.5 and the lemma that:

Corollary 5.8. *If f (or A) has n distinct eigenvalues, then it is diagonalizable.*

A key tool for studying eigenvalues, eigenvectors and diagonalization of a matrix is its *characteristic polynomial*. Recall the definition:

Definition 5.9. The **characteristic polynomial** of an $n \times n$-matrix A is the polynomial $p_A(x) = \det(xI - A)$ (in the variable x). The **characteristic equation** of A is the equation $\det(xI - A) = 0$.

It is easy to see from the permutation definition of the determinant that $p_A(x)$ is a polynomial of degree n. See the exercises for a proof using the inductive definition.

The connection with eigenvalue is given by:

Lemma 5.10. *λ is an eigenvalue for A if and and only if $p_A(\lambda) = 0$.*

Proof. λ is an eigenvalue for A

$$\begin{aligned}
&\Leftrightarrow \quad A\mathbf{v} = \lambda\mathbf{v} \text{ for some } \mathbf{v} \neq \mathbf{0} \\
&\Leftrightarrow \quad (\lambda I - A)\mathbf{v} = \ \text{ for some } \mathbf{v} \neq \mathbf{0} \\
&\Leftrightarrow \quad \text{nullity}(\lambda I - A) > 0 \\
&\Leftrightarrow \quad \text{rank}(\lambda I - A) < n \\
&\Leftrightarrow \quad \det(\lambda I - A) = 0 \\
&\Leftrightarrow \quad p_A(\lambda) = 0.
\end{aligned}$$

□

Example 5.11. The characteristic polynomial of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is

$$\begin{vmatrix} x-1 & -1 \\ 0 & x-1 \end{vmatrix} = (x-1)^2$$

and its only eigenvalue is 1.

The characteristic polynomial of $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is

$$\begin{vmatrix} x & -1 \\ -1 & x \end{vmatrix} = x^2 - 1 = (x+1)(x-1),$$

so its eigenvalues are ± 1.

The characteristic polynomial of $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is

$$\begin{vmatrix} x & -1 \\ 1 & x \end{vmatrix} = x^2 + 1 = (x-i)(x+i),$$

so its eigenvalues are the complex numbers $\pm i$ (and it has no real eigenvalues).

Everything we've said so far works in exactly the same way for complex vector spaces and matrices as for real vector spaces and matrices. However the last example shows that some real matrices are only diagonalizable when viewed as complex matrices. Roughly speaking, the process of diagonalizing matrices works better over $\mathbb{C}$ than over $\mathbb{R}$ because factorization of polynomials works better, The reason for this is the following important property of the complex numbers, called the *Fundamental Theorem of Algebra.* We state it without proof (the proof being outside the scope of this course).

Theorem 5.12. *If $f(x) = x^n + a_{n-1}x^{n-1} + \cdots a_0$ with $a_n, \ldots, a_1, a_0 \in \mathbb{C}$, then*

$$f(x) = (x - \alpha_1)(x - \alpha_2)\cdots(x - \alpha_n)$$

for some $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}$.

In other words, any polynomial over $\mathbb{C}$ of degree n factors completely into linear factors, i.e., it has n roots, counting multiplicity.

For the rest of the section, we will be working with complex vector spaces and matrices unless otherwise stated.

Definition 5.13. The **algebraic multiplicity** of an eigenvalue λ of A is the multiplicity of λ as a root of the characteristic equation, i.e., if

$$p_A(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues for A, then the algebraic multiplicity of each λ_i is m_i. The **geometric multiplicity** of λ is the nullity of $(\lambda I - A)$.

Example 5.14. The algebraic multiplicity of the eigenvalue 1 of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is 2 (since the characteristic polynomial is $(x-1)^2$). The geometric multiplicity is the nullity of $\begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}$, which is 1.

For the 2×2 identity matrix, the characteristic polynomial is also $(x-1)^2$, so the only eigenvalue is 1, with algebraic multiplicity 2, which is also its geometric multiplicity.

For the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, the algebraic multiplicity of each eigenvalue is 1, as is its geometric multiplicity.

Proposition 5.15. *Suppose that B is similar to A. Then*

1. $p_A(x) = p_B(x)$ *(in particular, A and B have the same eigenvalues);*
2. *the algebraic multiplicity of an eigenvalue λ of A is the same as its algebraic multiplicity as an eigenvalue of B;*
3. *the geometric multiplicity of an eigenvalue λ of A is the same as its geometric multiplicity as an eigenvalue of B.*

Proof. Since $B = QAQ^{-1}$ for some invertible matrix Q, we have

$$Q(xI - A)Q^{-1} = (xQ - QA)Q^{-1} = (xQQ^{-1} - QAQ^{-1}) = xI - B.$$

Therefore,

(1) $p_B(x) = \det(xI - B) = \det(Q(xI - A)Q^{-1}) = \det(Q)\det(xI - A)\det(Q^{-1}) = \det(Q)\det(Q^{-1})p_A(x) = p_A(x)$.

(2) This is immediate from (1) and the definition of algebraic multiplicity.

(3) From above, we see that $\lambda I - B = Q(\lambda I - A)Q^{-1}$. Since Q is invertible, it follows that $\text{nullity}(\lambda I - B) = \text{nullity}(\lambda I - A)$. $\square$

When working with a fixed matrix A, we will write m_λ for the algebraic multiplicity of λ and n_λ for the geometric multiplicity of λ.

Theorem 5.16. *A is diagonalizable if and only $m_\lambda = n_\lambda$ for all eigenvalues λ.*

Proof. $\Rightarrow$: If A is diagonal, then $A \simeq D$ for some diagonal matrix

$$D = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{pmatrix}.$$

By Prop. 5.15, we just have to show that $m_\lambda = n_\lambda$ for the matrix D. Since $p_D(x) = (x - \alpha_1)(x - \alpha_2)\cdots(x - \alpha_n)$, we see that m_λ is the number times that $\lambda = \alpha_i$. On the other hand n_λ is the nullity of the matrix

$$D = \begin{pmatrix} \lambda - \alpha_1 & 0 & \cdots & 0 \\ 0 & \lambda - \alpha_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda - \alpha_n \end{pmatrix},$$

which is also the number of times that $\lambda = \alpha_i$.

$\Leftarrow$: Conversely, suppose that $m_\lambda = n_\lambda$ for all eigenvalues λ. We will show that $\mathbb{C}^n$ has a basis consisting of eigenvectors for A. Denote the distinct eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_k$, write m_i for the algebraic multiplicity of λ_i, and n_i for the geometric multiplicity. So for each $i = 1, \ldots, k$, the kernel of $\lambda_i I - A$ has dimension n_i. Let $S_i = \{\mathbf{v}_{i1}, \ldots, \mathbf{v}_{in_i}\}$ be a basis for $\ker(\lambda_i I - A)$. Note that each $\mathbf{v}_{ij}$ is an eigenvector for A with eigenvector λ_i. We will prove that

$$S = S_1 \cup S_2 \cup \cdots \cup S_k = \{\mathbf{v}_{11}, \ldots, \mathbf{v}_{1n_1}, \mathbf{v}_{21}, \ldots, \mathbf{v}_{2n_2}, \ldots, \mathbf{v}_{k1}, \ldots, \mathbf{v}_{kn_k}\}$$

is a basis for $\mathbb{C}^n$.

We first prove that S is linearly independent. Suppose that $\alpha_{11}, \ldots, \alpha_{1n_1}$, $\ldots$, $\alpha_{k1}, \ldots, \alpha_{kn_k} \in \mathbb{C}$ are such that

$$\alpha_{11}\mathbf{v}_{11} + \cdots + \alpha_{1n_1}\mathbf{v}_{1n_1}\alpha_{21}\mathbf{v}_{21} + \cdots + \alpha_{2n_2}\mathbf{v}_{2n_2} + \cdots + \alpha_{k1}\mathbf{v}_{k1} + \cdots + \alpha_{kn_k}\mathbf{v}_{kn_k} = \mathbf{0}.$$

Let $\mathbf{w}_1 = \alpha_{11}\mathbf{v}_{11} + \cdots + \alpha_{1n_1}\mathbf{v}_{1n_1}$, $\ldots$, $\mathbf{w}_k = \alpha_{k1}\mathbf{v}_{k1} + \cdots + \alpha_{kn_k}\mathbf{v}_{kn_k}$. Each $\mathbf{w}_i$ that is non-zero is an eigenvector for A with eigenvalue λ_i. Since the λ_i are distinct and $\mathbf{w}_1 + \cdots + \mathbf{w}_k = 0$, we see by Lemma 5.7 that $\mathbf{w}_1 = \mathbf{w}_2 = \cdots = \mathbf{w}_k = \mathbf{0}$. Since each S_i is linearly independent it follows that $\alpha_{ij} = 0$ for all $j = 1, \ldots, n_i$.

We have now shown that S is linearly independent, so to prove that S is a basis it suffices to prove that S has n elements. By construction, S has $n_1 + n_2 + \cdots + n_k$ elements, we assumed that $m_1 = n_1, \ldots, m_k = n_k$, and

$$m_1 + m_2 + \cdots + m_k = n$$

(by the Fundamental Theorem of Algebra). Since $\mathbb{C}^n$ has a basis of eigenvectors, A is diagonalizable by Theorem 5.5. $\square$

Recall that we are assuming A is a complex matrix. The results apply to real matrices, provided they are viewed as complex matrices. In particular, *diagonalizability* means as a *complex* matrix, and the equality $m_\lambda = n_\lambda$ must hold for all eigenvalues λ, including those that are complex but not real.

Example 5.17. For $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, we had $m_\lambda = 2$, $n_\lambda = 1$ for the (only) eigenvalue $\lambda = 1$.

While it is not the case that every matrix is similar to a diagonal matrix, the following is true:

Theorem 5.18. *Every square matrix is similar to an upper-triangular matrix (over $\mathbb{C}$).*

Proof. We prove this by induction on n. For $n = 1$ there is nothing to prove since every 1×1-matrix is upper-triangular.

Suppose that $n > 1$ and the theorem is true for $(n-1) \times (n-1)$-matrices. Let $\mathbf{v}$ be an eigenvector for A with eigenvalue $\lambda \in \mathbb{C}$. (By the Fundamental Theorem of Algebra, $p_A(x)$ has a root in $\mathbb{C}$, so A has an eigenvalue in $\mathbb{C}$.) Let S be any basis for $\mathbb{C}^n$ with v as its first vector, and let Q be the transition matrix from the standard basis to S, so the matrix of $f(\mathbf{x}) = A\mathbf{x}$ with respect to S is $B = QAQ^{-1}$. Since $A\mathbf{v} = \lambda\mathbf{v}$, the first column of B is $(\lambda, 0, \ldots, 0)^t$. Therefore

$$B = QAQ^{-1} = \begin{pmatrix} \lambda & \mathbf{r} \\ \mathbf{0} & A_1 \end{pmatrix}$$

for some row vector $\mathbf{r}$ of length $n-1$ and some $(n-1) \times (n-1)$-matrix A_1. By the induction hypothesis, there is an invertible $(n-1) \times (n-1)$-matrix Q_1 so that $T_1 = Q_1 A_1 Q_1^{-1}$ is upper-triangular. Let $R = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1 \end{pmatrix}$. Then R is invertible, and therefore so is RQ, and

$$\begin{aligned} (RQ)A(RQ)^{-1} &= RBR^{-1} \\ &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1 \end{pmatrix} \begin{pmatrix} \lambda & \mathbf{r} \\ \mathbf{0} & A_1 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1^{-1} \end{pmatrix} = \begin{pmatrix} \lambda & \mathbf{r}Q_1^{-1} \\ \mathbf{0} & T_1 \end{pmatrix} \end{aligned}$$

is upper-triangular. $\square$

Note that if A is similar to the upper-triangular matrix

$$T = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2n} \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & t_{nn} \end{pmatrix}$$

then $p_A(x) = p_T(x) = (x - t_{11})(x - t_{22}) \cdots (x - t_{nn})$.

Corollary 5.19. *If λ is an eigenvalue for A, then $n_\lambda \le m_\lambda$.*

Proof. By the theorem, we can assume A is the upper-triangular matrix T. Then m_λ is the number of times λ appears on the diagonal of T, and n_λ is the nullity of $\lambda I - T$. This matrix is upper-triangular and has $n - m_\lambda$ non-zero diagonal entries. The rows with non-zero diagonal entries are linearly independent, so $\text{rank}(\lambda I - T) \ge n - m_\lambda$, and $n_\lambda = n - \text{rank}(\lambda I - T) \le m_\lambda$. □

Note that if $f(x) = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0$ is a polynomial, say with coefficients in $\mathbb{C}$, and A is a square matrix, then we can "evaluate" the polynomial at $x = A$ by defining

$$f(A) = a_m A^m + a_{m-1}A^{m-1} + \cdots + a_1 A + a_0 I.$$

The following theorem states that when we substitute a matrix into its characteristic polynomial, we get the 0 matrix.

Theorem 5.20. [Cayley-Hamilton] *Every square matrix A satisfies its characteristic equation; i.e., $p_A(A) = 0$.*

Proof. Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be the eigenvalues of A repeated according to algebraic multiplicity. From Theorem 5.18 there is an invertible matrix R such that $A = RTR^{-1}$ where T is an upper-triangular matrix with $\lambda_1, \lambda_2, \cdots, \lambda_n$ as its diagonal entries. The characteristic polynomial p of A and of T is $(x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$. Writing this in the form $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, we see that

$$\begin{aligned} p(A) &= p(RTR^{-1}) \\ &= (RTR^{-1})^n + a_{n-1}(RTR^{-1})^{n-1} + \cdots + a_1 RTR^{-1} + a_0 I \\ &= RT^nR^{-1} + a_{n-1}RT^{n-1}R^{-1} + \cdots + a_1 RTR^{-1} + a_0 RIR^{-1} \\ &= R(T^n + a_{n-1}T^{n-1} + \cdots + a_1 T + a_0 I)R^{-1} \\ &= Rp(T)R^{-1}, \end{aligned}$$

so it is sufficient to show that $p(T) = (T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_n I) = 0$. Note that the i^{th} diagonal entry of $T - \lambda_i I$ is 0.

We complete the proof by showing that if $T_1, T_2, \ldots, T_n$ are upper-triangular matrices such that the i^{th} diagonal entry of T_i is 0 then $T_1 T_2 \cdots T_n = 0$. The proof is by induction on the size of the matrix. The result is obvious for $n = 1$. Suppose the result is true for $(n-1) \times (n-1)$ matrices. Write the

T_i as partitioned matrices

$$T_i = \begin{pmatrix} T_i' & \mathbf{v}_i \\ \mathbf{0} & t_i \end{pmatrix}, \text{ for } i = 1, \ldots, n-1, \text{ and } T_n = \begin{pmatrix} T_n' & \mathbf{v}_n \\ \mathbf{0} & 0 \end{pmatrix},$$

where each T_i' is an $(n-1) \times (n-1)$-upper triangular matrix with 0 in the i^{th} diagonal place. Therefore, by the induction hypothesis, $T_1'T_2' \cdots T_{n-1}' = 0$ and so

$$T_1T_2 \cdots T_{n-1} = \begin{pmatrix} T_1'T_2' \cdots T_{n-1}' & \mathbf{v} \\ \mathbf{0} & t \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{v} \\ \mathbf{0} & t \end{pmatrix}$$

where $\mathbf{v}$ and t are some vector and scalar (whose values are not important). Then

$$T_1T_2T_3 \cdot \cdots T_n = \begin{pmatrix} \mathbf{0} & \mathbf{v} \\ \mathbf{0} & t \end{pmatrix} \begin{pmatrix} T_n' & \mathbf{v}_n \\ \mathbf{0} & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} = 0$$

as required. $\square$

Example 5.21. The matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has characteristic polynomial $p_A(x) = x^2 + 1$ and satisfies $p_A(A) = 0$ since $A^2 + I = -I + I = 0$.

Lemma 5.22. *There is a unique monic polynomial $m_A(x)$ of lowest degree such that $m_A(A) = 0$. The polynomial $m_A(x)$ has the following properties:*

(1) *If $q(x)$ is a polynomial such that $q(A) = 0$, then $m_A(x)$ divides $q(x)$.*
(2) *The roots of $m_A(x) = 0$ are precisely the eigenvalues λ of A.*
(3) *If A and B are similar, then $m_A(x) = m_B(x)$.*

Proof. The Cayley-Hamilton Theorem shows that there is *some* monic polynomial $q(x)$ such that $q(A) = 0$, namely $q(x) = p_A(x)$. (Recall that a *monic* polynomial is one whose leading coefficient is 1.) Therefore there must be a monic polynomial $q(x)$ of lowest degree such that $q(A) = 0$. We denote it $m_A(x)$. The uniqueness will follow from part (1).

(1) Suppose that $q(x)$ is a polynomial such that $q(A) = 0$. By the division algorithm for polynomials, we have $q(x) = m_A(x)f(x) + r(x)$ for some polynomials $f(x)$, $r(x)$ where $\deg(r(x)) < \deg(m_A(x))$ (or $r(x) = 0$). Substituting $x = A$ gives

$$0 = q(A) = m_A(A)f(A) + r(A) = r(A)$$

since $m_A(A) = 0$. If $r(x) \neq 0$, then $r(x) = b_dx^d + \cdots + b_0$ for some $d < \deg(m_A(x))$ and $b_d \neq 0$, in which case $b_d^{-1}r(x)$ is a monic polynomial of degree less than $\deg(m_A(x))$ such that $b_d^{-1}r(A) = 0$, contradicting our definition of $m_A(x)$. Therefore $r(x) = 0$, and $q(x) = m_A(x)f(x)$ is divisible by $m_A(x)$.

The uniqueness of $m_A(x)$ follows since if $q(x)$ is a monic polynomial of the same degree as $m_A(x)$ and $q(x) = 0$, then by (1), $q(x) = m_A(x)f(x)$ where $f(x)$ is a scalar c (since $q(x)$ and $m_A(x)$ have the same degree). Since $m_A(x)$ and $q(x)$ are both monic, we must have $c = 1$, and $q(x) = m_A(x)$.

(2) By (1), we know that if $m_A(x)$ divides $p_A(x)$. Therefore if $m_A(\lambda) = 0$, then $p_A(\lambda) = 0$, so λ is an eigenvalue of A. Conversely, suppose that λ is an eigenvalue of A. We must show that $m_A(\lambda) = 0$. Let $\mathbf{v}$ be an eigenvector with eigenvalue λ. Note that since $A\mathbf{v} = \lambda\mathbf{v}$, we have $A^2\mathbf{v} = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda(A\mathbf{v}) = \lambda^2\mathbf{v}$, and similarly $A^i\mathbf{v} = \lambda^i\mathbf{v}$ for all $i \geq 0$. So writing $m_A(x) = x^m + b_{m-1}x^{m-1} + \cdots + b_0$, we see that

$$\begin{aligned}
m_A(A)\mathbf{v} &= (A^m + b_{m-1}A^{m-1} + \cdots + b_0 I)\mathbf{v} \\
&= A^m\mathbf{v} + b_{m-1}A^{m-1}\mathbf{v} + \cdots + b_0\mathbf{v} \\
&= \lambda^m\mathbf{v} + b_{m-1}\lambda^{m-1}\mathbf{v} + \cdots + b_0\mathbf{v} \\
&= (\lambda^m + b_{m-1}\lambda^{m-1} + \cdots + b_0)\mathbf{v} \\
&= m_A(\lambda)\mathbf{v}.
\end{aligned}$$

Since $m_A(A) = 0$, we therefore have $0 = m_A(A)\mathbf{v} = m_A(\lambda)\mathbf{v}$, and since $\mathbf{v} \neq \mathbf{0}$, it follows that the scalar $m_A(\lambda)$ is 0.

(3) is left an exercise. □

Definition 5.23. The polynomial $m_A(x)$ defined by the preceding lemma is called the **minimal polynomial** of A.

Example 5.24. Note that the lemma implies that $m_A(x)$ has the same roots as $p_A(x)$, but not necessarily with the same multiplicity; the multiplicity of the eigenvalue as a root of $m_A(x)$ may be less than as a root of $p_A(x)$. In particular, if A has n distinct eigenvalues (i.e., $p_A(x)$ has no repeated roots), then $m_A(x) = p_A(x)$, but this may fail if $p_A(X)$ has repeated roots. For example, if $n = 2$ and $A = I$, then $p_A(x) = (x-1)^2$, but $m_A(x) = x - 1$. On the other hand if $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, then $p_A(x) = (x-1)^2$, but $m_A(x) \neq x - 1$ since $A - I \neq 0$; therefore $m_A(x) = (x-1)^2$.

We give one more criterion for diagonalizability, now in terms of the minimal polynomial of A.

Theorem 5.25. *A square matrix A is diagonalizable if and only if its minimal polynomial $m_A(x)$ has no repeated roots.*

Proof. Suppose that A is diagonalizable. We must show that its minimal polynomial is $q(x) = (x - \lambda_1)\cdots(x - \lambda_k)$ where $\lambda_1, \ldots, \lambda_k$ are the *distinct* eigenvalues of A (i.e., listed without any repetition). Since each eigenvalue is a root of $m_A(x)$, we know that $m_A(x)$ is divisible by $q(x)$, so it suffices to show that $q(A) = 0$ as this implies (by Lemma 5.22) that $q(x)$ is also divisible by $m_A(x)$.

If $\mathbf{v}$ is an eigenvector for A, then $A\mathbf{v} = \lambda\mathbf{v}$ for some root λ of $q(x)$. So as in the proof of part (2) of Lemma 5.22, we see that $q(A)\mathbf{v} = q(\lambda)\mathbf{v} = \mathbf{0}$. Since A is diagonalizable, $\mathbb{C}^n$ has a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ consisting of eigenvectors for A, and we have seen that for each $i = 1, \ldots, n$, $q(A)\mathbf{v}_i = \mathbf{0}$, i.e., $\mathbf{v}_i$ is in the kernel of $q(A)$. Therefore $\text{nullity}(q(A)) = n$, so $q(A) = 0$.

Conversely, suppose that $m_A(x) = (x - \lambda_1) \cdots (x - \lambda_k)$ where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues. This means that

$$(A - \lambda_1 I) \cdots (A - \lambda_k I) = 0.$$

Recall that if B and C are $n \times n$-matrices, then $\text{nullity}(BC) \leq \text{nullity}(B) + \text{nullity}(C)$ (see Sheet 3, Exercise 9f, which in fact shows this is true for any matrices B, C such that BC is defined). Therefore

$$\begin{aligned} n &= \text{nullity}\, 0 = \text{nullity}((A - \lambda_1 I) \cdots (A - \lambda_k I)) \\ &\leq n_1 + n_2 + \cdots + n_k, \end{aligned}$$

where n_i is the geometric multiplicity of the eigenvalue λ_i. On the other hand we know that $n = m_1 + m_2 + \cdots + m_k$ where m_i is the algebraic multiplicity of λ_i, and that $n_i \leq m_i$ for each i (Cor. 5.19). Putting these inequalities together, we see that the only possibility is that $n_i = m_i$ for each i, and hence by Thm. 5.16 A is diagonalizable. $\square$

We finish the section with a discussion of *Jordan canonical form,* which is **not covered on the exam**. This gives a standard form for a matrix similar to A, which is as close as possible to being diagonal.

Theorem 5.26. *Every square matrix A is similar (over $\mathbb{C}$) to a partitioned matrix of the form:*

$$T = \begin{pmatrix} T_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & T_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & T_k \end{pmatrix},$$

where each T_i is a square matrix of the form

$$\begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix}$$

for some eigenvalue λ_i of A. Moreover the matrix T as above is unique, up to permuting the order of $T_1, \ldots, T_k$.

The form of the matrix in the theorem is called *Jordan canonical form.* We remark that there can be repetition among the eigenvalues $\lambda_1, \ldots, \lambda_k$ appearing in the expression, and the sizes of the $T_1, T_2, \ldots, T_k$ may be different from each other; also some of the T_i may be 1×1, in which case we have $T_i = (\lambda_i)$. We will sketch the proof, but first we illustrate the meaning of the theorem by listing the possible forms for T in the case $n = 3$.

- If $p_A(x)$ has distinct roots $\lambda_1, \lambda_2, \lambda_3$, then A is diagonalizable and
$$T = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$
So in this case $T_1 = (\lambda_1)$, $T_2 = (\lambda_2)$, $T_3 = (\lambda_3)$.
- If $p_A(x) = (x - \lambda_1)(x - \lambda_2)^2$ with $\lambda_1 \neq \lambda_2$, then
$$T = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{pmatrix}.$$
In the first case $k = 3$, $T_1 = (\lambda_1)$ and $T_2 = T_3 = (\lambda_2)$; in the second case $k = 2$, $T_1 = (\lambda_1)$ and $T_2 = \begin{pmatrix} \lambda_2 & 1 \\ 0 & \lambda_2 \end{pmatrix}$.
- If $p_A(x) = (x - \lambda)^3$, then
$$T = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}.$$
In the first case $k = 3$, $T_1 = T_2 = T_3 = (\lambda)$; in the second case $k = 2$, $T_1 = (\lambda)$ and $T_2 = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$; in the last case $k = 1$ and $T = T_1$.

We now sketch the proof of the theorem. By Theorem 5.18, we can assume A is upper-triangular. We will first reduce to the case where A has only one eigenvalue. Write $p_A(x) = (x - \lambda_1)^{n_1} \cdots (x - \lambda_r)^{n_r}$ where $\lambda_1, \ldots, \lambda_r$ are distinct, and consider the matrices $A_i = (A - \lambda_i)^{n_i}$ for $i = 1, \ldots, r$. Then considering the diagonal entries, we see that $\operatorname{rank} A_i \geq n - n_i$, so $\operatorname{nullity}(A_i) \leq n_i$. On the other hand, by the Cayley-Hamilton Theorem $A_1 A_2 \cdots A_r = 0$, so an argument like the one in the proof of Thm. 5.25 shows that in fact $\operatorname{nullity}(A_i) = n_i$. Let V_i denote the null space of A_i and suppose that $\mathbf{v}_i \in V_i$ for $i = 1, \ldots, r$ are such that $\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r = \mathbf{0}$. Suppose that $\mathbf{v}_i \neq \mathbf{0}$ for some i. If $j \neq i$, then $\mathbf{v}_i$ cannot be an eigenvector with eigenvalue λ_j (else we would have $A_i \mathbf{v}_i = (A - \lambda_i)^{n_i} \mathbf{v}_i = (\lambda_j - \lambda_i)^{n_i} \mathbf{v}_i \neq \mathbf{0}$). So $(A - \lambda_j)\mathbf{v}_i \neq \mathbf{0}$, but $(A - \lambda_j)$ commutes with A_i, so $(A - \lambda_j)\mathbf{v}_i$ is a non-zero vector in V_i. Inductively, we find that $A_j \mathbf{v}_i$ is a non-zero vector in V_i, and in fact, letting $A' = A_1 A_2 \cdots A_{i-1} A_{i+1} \cdots A_r$, we get that $A'\mathbf{v}_i$ is non-zero. On the other hand $A_j \mathbf{v}_j = \mathbf{0}$, so that $A'\mathbf{v}_j = \mathbf{0}$ for $j \neq i$, so applying A' to $\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r = \mathbf{0}$ gives a contradiction. It follows then as in the proof of Thm. 5.16 that if S_i is a basis for V_i for $i = 1, \ldots, r$, then $S = S_1 \cup \cdots \cup S_r$ is a basis for $\mathbb{C}^n$. Since $\mathbf{v}_i \in V_i$ implies $A\mathbf{v}_i \in V_i$, it follows that the matrix for A with respect to such a basis will be "block diagonal" with the i^{th} diagonal block B_i being a matrix for the linear map defined by A on V_i with respect to S_i. Moreover B_i satisfies $(B_i - \lambda_i)^{n_i} = 0$, so λ_i is the only eigenvalue of B_i.

It suffices to prove that each B_i has the required form, so we are reduced to the case where A has only one eigenvalue, so we will just write λ instead of λ_i. Replacing A by $A - \lambda I$, we can even assume $\lambda = 0$ from now on. (If $A - \lambda I$ is similar to T, then A is similar to $T + \lambda I$.) Suppose then that $m_A(x) = x^d$, so $A^d = 0$, but $A^{d-1} \neq 0$. Let $\mathbf{v}_1^{(1)}, \mathbf{v}_2^{(1)}, \ldots, \mathbf{v}_{s_1}^{(1)}$ be a basis for the image of A^{d-1} (viewed as the linear map $\mathbf{v} \mapsto A\mathbf{v}$), and for $i = 1, \ldots, s_1$, choose $\mathbf{w}_i^{(1)}$ so that $A^{d-1}\mathbf{w}_i^{(1)} = \mathbf{v}_i^{(1)}$. Consider the linear map f_1 from the image of A^{d-2} to the image of A^{d-1} defined by multiplication by A. Then $\mathbf{v}_1^{(1)}, \mathbf{v}_2^{(1)}, \ldots, \mathbf{v}_{s_1}^{(1)}$ are linearly independent vectors in the kernel, and so can be extended to a basis of the kernel of f_1:

$$\mathbf{v}_1^{(1)}, \mathbf{v}_2^{(1)}, \ldots, \mathbf{v}_{s_1}^{(1)}, \mathbf{v}_1^{(2)}, \mathbf{v}_2^{(2)}, \ldots, \mathbf{v}_{s_2}^{(2)}.$$

By lemma 2.10, these $s_1 + s_2$ vectors, together with

$$A^{d-2}\mathbf{w}_1^{(1)}, A^{d-2}\mathbf{w}_2^{(1)}, \ldots, A^{d-2}\mathbf{w}_{s_1}^{(1)},$$

form a basis for the image of A^{d-2}. Now choose $\mathbf{w}_i^{(2)}$ so $A^{d-2}\mathbf{w}_i^{(2)} = \mathbf{v}_i^{(2)}$ for $i = 1, \ldots, s_2$. Iterating the process for $j = 3, \ldots, d$ for the map $f_{j-1}(\mathbf{v}) = A\mathbf{v}$ from the image of A^{d-j} to the image of A^{d-j+1}, we get vectors

$$\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \ldots, \mathbf{w}_{s_1}^{(1)}, \mathbf{w}_1^{(2)}, \mathbf{w}_2^{(2)}, \ldots, \mathbf{w}_{s_2}^{(2)}, \ldots, \mathbf{w}_1^{(j)}, \mathbf{w}_2^{(j)}, \ldots, \mathbf{w}_{s_j}^{(j)}$$

with the following properties:

- The vectors $\mathbf{v}_i^{(t)} = A^{d-t}\mathbf{w}_i^{(t)}$, for $1 \leq t \leq j$, $1 \leq i \leq s_t$, form a basis for $\ker f_{j-1}$.
- The vectors $A^u\mathbf{w}_i^{(t)}$ for $1 \leq t \leq j$, $d-t \leq u \leq d-j$, $1 \leq i \leq s_t$, form a basis for the image of A^{d-j}.

In particular, for $j = d$, this gives a basis for $\mathbb{C}^n$. Now divide the basis into its $k = s_1 + s_2 + \cdots + s_d$ subsets of the form:

$$\mathbf{v}_i^{(t)} = A^{d-t}\mathbf{w}_i^{(t)}, A^{d-t-1}\mathbf{w}_i^{(t)}, \ldots, A\mathbf{w}_i^{(t)}, \mathbf{w}_i^{(t)}$$

(where $t = 1, \ldots, d$, and $i = 1, \ldots, s_t$ for each t). Note that the effect of A on each such subset is precisely that of a matrix T_i as in the statement of the theorem, so the matrix of A with respect to this basis has the required form (with $\lambda = 0$).

The uniqueness follows from the fact that the sizes of the blocks for each eigenvalue λ are determined by the nullities of the maps $(A - \lambda)^i$ for $1 \leq i \leq n_\lambda$.

6. Inner products

Definition 6.1. The **inner product** of two vectors $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{C}^n$ is defined by

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} u_i \bar{v}_i = u_1\bar{v}_1 + u_2\bar{v}_2 + \cdots + u_n\bar{v}_n.$$

Note that $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^t \bar{\mathbf{v}} = \mathbf{u} \cdot \bar{\mathbf{v}}$ where $\bar{\mathbf{v}} = (\bar{v}_1, \ldots, \bar{v}_n)$ is the complex conjugate of $\mathbf{v}$ (taken coordinatewise) and $\cdot$ dnotes the dot (or scalar) product of the two vectors. Thus $\langle \mathbf{u}, \mathbf{v} \rangle$ is a scalar in $\mathbb{C}$.

Example 6.2. If $\mathbf{u} = (3, 1+i, -1)$ and $\mathbf{v} = (1-2i, -i, 2)$, then

$$\langle \mathbf{u}, \mathbf{v} \rangle = 3(1+2i) + (1+i)i + (-1)2 = 7i.$$

If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then the same definition applies, simply giving the real number $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^t \mathbf{v} = \mathbf{u} \cdot \mathbf{v}$ since $\mathbf{v} = \bar{\mathbf{v}}$.

Proposition 6.3. *If* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^n$ *and* $\lambda, \mu \in \mathbb{C}$, *then*

(1) $\langle \lambda \mathbf{u} + \mu \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{u}, \mathbf{w} \rangle + \mu \langle \mathbf{v}, \mathbf{w} \rangle$;
(2) $\langle \mathbf{v}, \mathbf{u} \rangle = \overline{\langle \mathbf{u}, \mathbf{v} \rangle}$;
(3) $\langle \mathbf{u}, \mathbf{u} \rangle \in \mathbb{R}$, *and* $\langle \mathbf{u}, \mathbf{u} \rangle > 0$, *unless* $\mathbf{u} = \mathbf{0}$, *in which case* $\langle \mathbf{u}, \mathbf{u} \rangle = 0$.

The proof is straighforward and is left as an exercise. Note that according to (2), the order of $\mathbf{u}$ and $\mathbf{v}$ matters; interchanging the vectors replaces the inner product by its complex conjugate, so for example if $\mathbf{u} = (3, 1+i, -1)$ and $\mathbf{v} = (1-2i, -i, 2)$, then $\langle \mathbf{u}, \mathbf{v} \rangle = 7i$, but $\langle \mathbf{v}, \mathbf{u} \rangle = -7i$

Definition 6.4. If $\mathbf{v} \in \mathbb{C}^n$, then the **norm** (or **length**) of $\mathbf{v}$ is defined as $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. We say $\mathbf{u}$ is **orthogonal** to $\mathbf{v}$ if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Note that $||\mathbf{v}||$ is a non-negative real number. For the vectors from Example 6.2, with $\mathbf{u} = (3, 1+i, -1)$ and $\mathbf{v} = (1-2i, -i, 2)$, we have $||\mathbf{u}|| = \sqrt{12} = 2\sqrt{3}$ and $||\mathbf{v}|| = \sqrt{10}$.

More generally, suppose V is any finite-dimensional complex vector space equipped with a map $V \times V \to \mathbb{C}$ denoted $\langle , \rangle$; i.e., a rule associating a complex number $\langle u, v \rangle$ to any elements $u, v \in \mathbb{C}$. Then V is an **inner product space** if it satisies (1), (2) and (3) of the preceding proposition. In particular, there is the notion of the norm of a vector in V defined by the formula $||v|| = \sqrt{\langle v, v \rangle}$ (a non-negative real number by property (3)); there is also a notion of orthogonality as for $\mathbb{C}^n$. Note that $\mathbb{C}^n$, with the inner product defined above, is an inner product space. There is a similar notion of a real inner product space, which is a real vector space V equipped with an inner product $\langle v, w \rangle$ satisfying analogous axioms. Everything we do for (complex) inner product spaces carries over without change, except that complex conjugation plays no role.

Proposition 6.5. *If* V *is an inner product space, then for* $u, v, w \in V$ *and* $\lambda, \mu \in \mathbb{C}$, *we have*

(1) $\langle u, \lambda v + \mu w \rangle = \overline{\lambda} \langle u, v \rangle + \overline{\mu} \langle u, w \rangle$;
(2) $||\lambda v|| = |\lambda| \, ||v||$;
(3) $\langle 0, v \rangle = \langle v, 0 \rangle = 0$;
(4) [Pythagorean Theorem] *if* $\langle u, v \rangle = 0$, *then* $||u+v||^2 = ||u||^2 + ||v||^2$.

The proofs of these are also straightforward and left as exercises. Note that the proposition applies to to $V = \mathbb{C}^n$ with its usual inner product. The case of (4) for $u, v \in \mathbb{R}^2$ is the usual Pythagorean Theorem for triangles in the plane.

Theorem 6.6. *Suppose that V is an inner product space and $u, v \in V$. Then*

(1) [**Cauchy-Schwartz Inequality**] $\quad |\langle u, v\rangle| \leq ||u||\,||v||$;
(2) [**Triangle Inequality**] $\quad ||u+v|| \leq ||u|| + ||v||$.

Proof. (1) If $v = 0$, then $\langle u, v\rangle = 0$, so the assertion is obvious. If $v \neq 0$, then $||v||^2 \neq 0$, and we let

$$w = u - \frac{\langle u, v\rangle}{||v||^2}\, v.$$

Then $\langle w, v\rangle = \langle u, v\rangle - \frac{\langle u,v\rangle}{||v||^2}\langle v, v\rangle = \langle u, v\rangle - \langle u, v\rangle = 0$. Therefore w is also orthogonal to $v' = \frac{\langle u,v\rangle}{||v||^2}\, v$, and since $u = w + v'$, the Pythagorean Theorem implies that

$$||u||^2 = ||w||^2 + ||v'||^2 = ||w||^2 + \frac{\langle u, v\rangle^2}{||v||^2}.$$

It follows that $\frac{\langle u,v\rangle^2}{||v||^2} \leq ||u||^2$,and (1) follows on taking square roots and multiplying through by $||v||$.

(2) We have

$$\begin{aligned} ||u+v||^2 &= \langle u+v, u+v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= ||u||^2 + \langle u, v\rangle + \overline{\langle u, v\rangle} + ||v||^2 \\ &= ||u||^2 + 2\mathrm{Re}(\langle u, v\rangle) + ||v||^2 \end{aligned}$$

where $\mathrm{Re}(z) = \frac{1}{2}(z + \bar{z})$ is the real part of z. Since $\mathrm{Re}(z) \leq |z|$ for any $z \in \mathbb{C}$, this gives (using part (1))

$$\begin{aligned} ||u+v||^2 &\leq ||u||^2 + 2|\langle u, v\rangle| + ||v||^2 \\ &\leq ||u||^2 + 2||u||\,||v|| + ||v||^2 \\ &= (||u|| + ||v||)^2, \end{aligned}$$

and (2) follows on taking square roots. □

Example 6.7. Again using the vectors $\mathbf{u}$ and $\mathbf{v}$ from Example 6.2, we have $|\langle \mathbf{u}, \mathbf{v}\rangle| = 7$, $||\mathbf{u}|| = 2\sqrt{3}$, $||\mathbf{v}|| = \sqrt{10}$ and $\mathbf{u} + \mathbf{v} = (4 - 2i, 1, 1)$, so $||\mathbf{u} + \mathbf{v}|| = \sqrt{22}$. The inequalities in Theorem 6.6 are $7 \leq 2\sqrt{30}$ and $2\sqrt{3} + \sqrt{10} \leq \sqrt{22}$.

Definition 6.8. A set of vectors $\{u_1, u_2, \ldots, u_k\}$ in an inner product space V is **orthonormal** if

(1) $\langle u_i, u_i\rangle = 1$ for $i = 1, \ldots, k$;
(2) $\langle u_i, u_j\rangle = 0$ if $i \neq j$ (with $i, j \in \{1, \ldots, k\}$).

Note that the first condition is equivalent to $||u_i|| = 1$ for $i = 1, \ldots, k$.

Example 6.9. Let $\mathbf{u}_1 = \frac{1}{2}(1, i, 1, 1)^t$, $\mathbf{u}_2 = \frac{1}{\sqrt{2}}(1, 0, -1, 0)^t$ and $\mathbf{u}_3 = \frac{1}{2}(i, 1, i, -i)^t$. Then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal subset of $\mathbb{C}^4$.

Lemma 6.10. *An orthonormal set is linearly independent.*

Proof. Suppose $\{u_1, \dots, u_k\}$ is orthonormal and $\alpha_1, \dots, \alpha_k \in \mathbb{C}$ are such that

$$\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_k u_k = 0.$$

We must show that $\alpha_i = 0$ for $i = 1, \dots, k$. Since $\langle 0, u_i \rangle = 0$ for each i, we know that

$$0 = \langle \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_k u_k\, ,\, u_i \rangle = \alpha_1 \langle u_1, u_i \rangle + \alpha_2 \langle u_2, u_i \rangle + \cdots + \alpha_k \langle u_k, u_i \rangle.$$

But $\langle u_j, u_i \rangle = 0$ unless $j = i$, so the only surviving term on the right is $\alpha_i \langle u_i, u_i \rangle = \alpha_i$, which must therefore be 0. □

If an orthonormal set spans V, then (being linearly independent) it is necessarily a basis, called an orthonormal basis.

Example 6.11. The standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$, where $\mathbf{e}_i$ is the vector $(0, 0, \dots, 0, 1, 0, \dots, 0)^t$ (with the 1 in the i^{th} place) is an orthonormal basis for $\mathbb{C}^n$. For another example of an orthonormal basis, take $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ where $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are as in Example 6.9 and $\mathbf{u}_4 = \frac{1}{\sqrt{2}}(0, i, 0, -1)^t$.

Lemma 6.12. *Let $(\alpha_1, \alpha_2, \dots, \alpha_n)^t$ and $(\beta_1, \beta_2, \dots, \beta_n)^t$ be the coordinates of v and w with respect to an orthonormal basis $\{u_1, u_2, \dots, u_n\}$ of an inner product space V. Then*

(1) $\alpha_i = \langle v, u_i \rangle$ *for* $i = 1, \dots, n$*;*
(2) $\langle v, w \rangle = \alpha_1 \overline{\beta}_1 + \cdots + \alpha_n \overline{\beta}_n$.

Proof. (1) For $(\alpha_1, \alpha_2, \dots, \alpha_n)^t$ to be the coordinates of v means that $v = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n$. Therefore

$$\langle v, u_i \rangle = \alpha_1 \langle u_1, u_i \rangle + \alpha_2 \langle u_2, u_i \rangle + \cdots + \alpha_n \langle u_n, u_i \rangle,$$

and since $\langle u_j, u_i \rangle = 0$ unless $i = j$, in which case it is 1, we get $\langle v, u_i \rangle = \alpha_i$.

(2) Since $w = \beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n$, we have

$$\langle v, w \rangle = \overline{\beta}_1 \langle v, u_1 \rangle + \overline{\beta}_2 \langle v, u_2 \rangle + \cdots + \overline{\beta}_n \langle v, u_n \rangle,$$

which by (1) is $\alpha_1 \beta_1 + \cdots + \alpha_n \beta_n$. □

Example 6.13. Let $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ be the orthonormal basis of $\mathbb{C}^4$ from Example 6.11. If $\mathbf{v} = (1 + i, -2, 0, -i)^t$, then the coordinates of $\mathbf{v}$ are given by $\alpha_1 = \langle \mathbf{v}, \mathbf{u}_1 \rangle = (1+2i)/2$, $\alpha_2 = \langle \mathbf{v}, \mathbf{u}_2 \rangle = (1+i)/\sqrt{2}$, $\alpha_3 = \langle \mathbf{v}, \mathbf{u}_3 \rangle = -i/2$ and $\alpha_4 = \langle \mathbf{v}, \mathbf{u}_4 \rangle = 3i/\sqrt{2}$.

Definition 6.14. For a complex $n \times n$ matrix $A = (a_{ij})$, the **adjoint** of A, denoted A^*, is the matrix whose (i, j)-entry is $\overline{a}_{ji}$; in other words $A^* = \overline{A}^t$. A complex $n \times n$ matrix U is **unitary** if $UU^* = I$; in other words, U is invertible and $U^{-1} = U^*$. A real $n \times n$ matrix P is **orthogonal** if $PP^t = I$; in other words P is orthogonal if P is invertible and $P^{-1} = P^t$. (Note that if A is real, then $A^* = A^t$, so a real matrix is orthogonal if and only if it is unitary.)

Example 6.15. The matrix $U = \frac{1}{5}\begin{pmatrix} 3 & 4i \\ -4 & 3i \end{pmatrix}$ is unitary since $U^* = \frac{1}{5}\begin{pmatrix} 3 & -4 \\ -4i & -3i \end{pmatrix} = U^{-1}$.

Proposition 6.16. *Let A and U be complex $n \times n$ matrices, and P a real $n \times n$ matrix. Then*

(1) $(A^*)^* = A$;
(2) *if U is unitary, then so are $U^* = U^{-1}$, $\overline{U}$ and $U^t = \overline{U}^{-1}$;*
(3) *if P is real orthogonal, then so is $P^t = P^{-1}$.*

The proof is left as an exercise.

Proposition 6.17. *Suppose that $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $S' = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ are bases for $\mathbb{C}^n$ and that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is orthonormal. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is orthonormal if and only the transition matrix from S' to S is unitary.*

Proof. Let $U = (\alpha_{ij})$ be the transition matrix from S' to S. Then the j^{th} column $(\alpha_{1j}, \ldots, \alpha_{nj})^t$ of U is given by the coordinates of $\mathbf{v}_j$ with respect to S. Therefore the i^{th} *row* of U^* is $(\overline{\alpha}_{1i}, \ldots, \overline{\alpha}_{ni})$. It follows that the (i, j)-entry of U^*U is

$$\alpha_{1j}\overline{\alpha}_{1i} + \alpha_{2j}\overline{\alpha}_{2i} + \cdots + \alpha_{nj}\overline{\alpha}_{ni},$$

which by Lemma 6.12 is precisely $\langle \mathbf{v}_j, \mathbf{v}_i \rangle$. So $U^*U = I$ is equivalent to having $\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1$ for all i, and $\langle \mathbf{v}_j, \mathbf{v}_i \rangle = 0$ whenever $i \neq j$. $\square$

Corollary 6.18. *A complex $n \times n$ matrix is unitary if and only if its columns form an orthonormal basis for $\mathbb{C}^n$.*

Proof. The basis $S = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal, and the transition matrix from $S' = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ to S is the matrix whose columns are precisely the vectors in S' $\square$

Example 6.19. From Example6.11, we see that the matrix

$$\frac{1}{2}\begin{pmatrix} 1 & \sqrt{2} & i & 0 \\ i & 0 & 1 & i\sqrt{2} \\ 1 & -\sqrt{2} & i & 0 \\ 1 & 0 & -i & -\sqrt{2} \end{pmatrix}$$

is unitary.

Analagous statemenes are true for real matrices; in particular, a real $n \times n$ matrix is orthogonal if and only if its columns form an orthonormal basis for $\mathbb{R}^n$.

Definition 6.20. If A and B are complex square matrices, then we say that A is **unitarily similar** to B if $A = U^{-1}BU = U^*BU$ for some unitary matrix U, and A is **unitarily diagonalizable** if it is unitarily similar to a diagonal matrix. If A and B are real square matrices, then we say that A is **orthogonally similar** to B if $A = P^{-1}BP = P^tBP$ for some orthogonal matrix P, and A is **orthogonally diagonalizable** if it is orthogonally similar to a diagonal matrix.

It is easy to see that unitary (and orthogonal) similarity are equivalence relations using the fact that if U and U' are unitary, then so are U^{-1} and UU' (see the exercises).

Theorem 6.21. *A complex (or real) matrix A is unitarily (or orethogonally) diagonalizable if and only if there is an orthonormal basis consisting of eigenvectors for A.*

Proof. This is clear from the definitions and Corollary 6.18 since U^*AU is diagonal if and only the columns of U are eigenvectors for A. □

Example 6.22. The matrix $A = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$ is unitarily diagonalizable since $\frac{1}{\sqrt{2}}(1,1)^t$ and $\frac{1}{\sqrt{2}}(1,-1)^t$ form an orthonormal basis of eigenvectors with eigenvalues i and $-i$. More explicitly, we have $U^*AU = D$ where $U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $D = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$.

Definition 6.23. A square matrix A is

(1) **symmetric** if $A = A^t$;
(2) **self-adjoint** (or **Hermitian**) if $A^* = A$;
(3) **normal** if $AA^* = A^*A$.

Example 6.24. If A is self-adjoint, then A is normal. The matrix $A = \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}$ is not self-adjoint, but it is normal since $A^* = -A$, so $A^*A = AA^* = -A^2$.

Proposition 6.25.

(1) *If a real matrix is orthogonally diagonalizable, then it is symmetric.*
(2) *If a complex matrix is unitarily diagonalizable, then it is normal.*

Proof. (1) If A is orthogonally diagonalizable, then $A = PDP^t$ for some orthogonal P and diagonal D. Then $A^t = (PDP^t)^t = (P^t)^tD^tP^t = PDP^t = A$, so A is symmetric.

(2) If A is unitarily diagonalizable, then $A = UDU^*$ for some unitary U and diagonal D. Then $A^* = (UDU^*)^* = (U^*)^*D^*U^* = UD^*U^*$, so

$$AA^* = (UDU^*)(UD^*U^*) = U(DD^*)U^*$$

and similarly $A^*A = U(D^*D)U^*$. Since D is diagonal $D^* = \overline{D}$ and $DD^* = D^*D$, so A is normal. □

We will later prove the converse to both parts of the proposition. First we note the behavior of adjoint and unitary matrices with respect to the inner product.

Proposition 6.26. *Suppose that A is a complex $n \times n$-matrix.*

(1) *If $\langle A\mathbf{v}, \mathbf{w}\rangle = 0$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$, then $A = 0$.*
(2) *If $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$, then $\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, A^*\mathbf{w}\rangle$.*
(3) *If U is a unitary $n \times n$ matrix and $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$, then*

$$\langle U\mathbf{v}, U\mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{w}\rangle \quad \textit{and} \quad ||U\mathbf{v}|| = ||\mathbf{v}||.$$

Proof. (1) Let $\mathbf{e}_i$ be the i^{th} standard basis vector for $i = 1, \ldots, n$. Then $A\mathbf{e}_j$ is the j^{th} column of $A = (a_{ij})$, so $A\mathbf{e}_j = \sum_{k=1}^n a_{kj}\mathbf{e}_k$. Therefore

$$\langle A\mathbf{e}_j, \mathbf{e}_i\rangle = \left\langle \sum_{k=1}^n a_{kj}\mathbf{e}_k \,,\, \mathbf{e}_i \right\rangle = \sum_{k=1}^n a_{kj}\langle \mathbf{e}_k, \mathbf{e}_i\rangle.$$

The only term that survives on the right is for $i = k$, giving $\langle A\mathbf{e}_j, \mathbf{e}_i\rangle = a_{ij}$. Applying $\langle A\mathbf{v}, \mathbf{w}\rangle = 0$ with $\mathbf{v} = \mathbf{e}_i, \mathbf{w} = \mathbf{e}_j$ therefore gives $a_{ij} = 0$ for all i, j, so $A = 0$.

(2) Since $\overline{A^*} = A^t$, we have

$$\langle \mathbf{v}, A^*\mathbf{w}\rangle = \mathbf{v}^t\overline{A^*\mathbf{w}} = \mathbf{v}^t\overline{A^*}\,\overline{\mathbf{w}} = \mathbf{v}^t A^t\overline{\mathbf{w}} = (A\mathbf{v})^t\overline{\mathbf{w}} = \langle A\mathbf{v}, \mathbf{w}\rangle.$$

(3) Applying part (2) with $A = U$ and with $U\mathbf{w}$ in place of $\mathbf{w}$ gives

$$\langle U\mathbf{v}, U\mathbf{w}\rangle = \langle \mathbf{v}, U^*U\mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{w}\rangle$$

since $U^*U = I$. The assertion about $||U\mathbf{v}||$ follows from the case $\mathbf{v} = \mathbf{w}$ and taking square roots. $\square$

Analogous statements hold for real matrices. (See the exercises.)

Theorem 6.27. *If A self-adjoint, then*

(1) *the eigenvalues of A are real;*
(2) *eigenvectors with distinct eigenvalues are orthogonal.*

Proof. (1) Suppose $\lambda \in \mathbb{C}$ is an eigenvalue of A, so $A\mathbf{v} = \lambda\mathbf{v}$ for some $\mathbf{v} \neq \mathbf{0}$. Then $\langle A\mathbf{v}, \mathbf{v}\rangle = \langle \lambda\mathbf{v}, \mathbf{v}\rangle = \lambda\langle \mathbf{v}, \mathbf{v}\rangle$. Since $A = A^*$, part (1) of Prop. 6.26 implies that

$$\langle A\mathbf{v}, \mathbf{v}\rangle = \langle \mathbf{v}, A^*\mathbf{v}\rangle = \langle \mathbf{v}, A\mathbf{v}\rangle = \langle \mathbf{v}, \lambda\mathbf{v}\rangle = \overline{\lambda}\langle \mathbf{v}, \mathbf{v}\rangle.$$

Since $\lambda\langle \mathbf{v}, \mathbf{v}\rangle = \overline{\lambda}\langle \mathbf{v}, \mathbf{v}\rangle$ and $\langle \mathbf{v}, \mathbf{v}\rangle \neq 0$, it follows that $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$.

(2) Suppose that $A\mathbf{v} = \lambda\mathbf{v}$ and $A\mathbf{w} = \mu\mathbf{w}$ for some non-zero $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$, and some $\lambda \neq \mu \in \mathbb{C}$. Then $\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \lambda\mathbf{v}, \mathbf{w}\rangle = \lambda\langle \mathbf{v}, \mathbf{w}\rangle$, and

$$\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, A\mathbf{w}\rangle = \langle \mathbf{v}, \mu\mathbf{w}\rangle = \overline{\mu}\langle \mathbf{v}, \mathbf{w}\rangle = \mu\langle \mathbf{v}, \mathbf{w}\rangle$$

since $\mu \in \mathbb{R}$. Since $\lambda\langle \mathbf{v}, \mathbf{w}\rangle = \mu\langle \mathbf{v}, \mathbf{w}\rangle$ and $\lambda \neq \mu$, it follows that $\langle \mathbf{v}, \mathbf{w}\rangle = 0$. $\square$

Note that the theorem applies to real symmetric matrices.

The theorem has the following corollary (which we will improve upon later, removing the assumption that A has n eigenvalues).

Corollary 6.28.

(1) *If A is self-adjoint with n distinct eigenvalues, then A is unitarily diagonalizable.*
(2) *If A is real symmetric with n distinct eigenvalues, then A is orthogonally diagonalizable.*

Proof. (1) Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis consisting of egenvectors for the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $\mathbf{u}_i = \frac{\mathbf{v}_i}{||\mathbf{v}_i||}$, for $i = 1, \ldots, n$. Then each $\mathbf{u}_i$ is an eigenvector with eigenvalue λ_i. By Thm. 6.27 we have $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ for $i \neq j$. Since $||\mathbf{u}_i|| = 1$ for $i = 1, \ldots, n$, we see that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal basis of eigenvectros for A, so A is unitarily diagonalizable. The proof of (2) is the same. $\square$

Example 6.29. Let A be the self-adjoint matrix $\begin{pmatrix} 1 & 2i \\ -2i & 4 \end{pmatrix}$. Then $\det(xI - A) = x^2 - 5x = x(x-5)$, so the eigenvalues are the real number 0 and 5. For $\lambda_1 = 0$, we have the eigenvector $\mathbf{v}_1 = (-2i, 1)^t$, and for $\lambda_2 = 5$, we have the eigenvector $\mathbf{v}_2 = (1, -2i)^t$. Then $\mathbf{v}_1$ is orthogonal to $\mathbf{v}_2$ and an orthonormal basis is given by multiplying each by its norm $\sqrt{1}5$. Thus if we take $U = \frac{1}{\sqrt{5}} \begin{pmatrix} -2i & 1 \\ 1 & -2i \end{pmatrix}$, then U is unitary, and $U^*AU = \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix}$.

Before treating the general cases of normal and real symmetric matrices, we establish the following useful algorithm for constructing orthonormal sets.

Theorem 6.30 (Gram-Schmidt Process)**.** *Suppose that $\{v_1, v_2, \ldots, v_k\}$ is a linearly independent subset of an inner product space V. Let $u_1, u_2, \ldots, u_k$ be vectors defined inductively as follows:*

$$\begin{array}{rrcll} \textit{Let} & w_1 & = & v_1, & u_1 = \frac{w_1}{||w_1||}, \\ & w_2 & = & v_2 - \langle v_2, u_1 \rangle u_1, & u_2 = \frac{w_2}{||w_2||}, \\ & w_3 & = & v_3 - \langle v_3, u_1 \rangle u_1 - \langle v_3, u_2 \rangle u_2, & u_3 = \frac{w_3}{||w_3||}, \\ & & \vdots & & \quad\vdots \\ \textit{and} & w_k & = & v_k - \sum_{i=1}^{k-1} \langle v_k, u_i \rangle u_i, & u_k = \frac{w_k}{||w_k||}. \end{array}$$

Then $\{u_1, \ldots, u_k\}$ is orthonormal and has the same span as $\{v_1, \ldots, v_k\}$.

Proof. We prove the theorem by induction on k. For $k = 1$, we have $||u_1|| = 1$, so $\{u_1\}$ is orthonormal, and since u_1 is a non-zero multiple of v_1, they have the same span.

Now suppose $k > 1$ and the theorem holds with k replaced by $k - 1$. Then $\{u_1, \ldots, u_{k-1}\}$ is orthonormal and has the same span as $\{v_1, \ldots, v_{k-1}\}$.

We must first check that $w_k \neq 0$ in order for the formula defining u_k to make sense. Suppose then that $w_k = 0$. Then

$$v_k = \sum_{i=1}^{k-1} \langle v_k, u_i \rangle u_i \ \in \operatorname{span}\{u_1, \ldots, u_{k-1}\} = \operatorname{span}\{v_1, \ldots, v_{k-1}\}$$

contradicts the assumption that $\operatorname{span}\{v_1, \ldots, v_{k-1}, v_k\}$ is linearly independent (recall Lemma 1.19).

Next we check that $\{u_1, \ldots, u_k\}$ is orthonormal. Since $\{u_1, \ldots, u_{k-1}\}$ is orthonormal, we already know that $||u_i|| = 1$ for $i = 1, \ldots, k-1$ and $\langle u_i, u_j \rangle = 0$ if $i \neq j \in \{1, \ldots, k-1\}$. It is clear that $||u_k|| = 1$, so we only need to check that $\langle w_k, u_j \rangle = 0$ for $j = 1, \ldots, k-1$ (as this implies that $\langle u_k, u_j \rangle = 0$ and $\langle u_j, u_k \rangle = \overline{\langle u_k, u_j \rangle} = 0$ for $j = 1, \ldots, k-1$). But for each such j, we have

$$\begin{aligned} \langle w_k, u_j \rangle &= \left\langle v_k - \textstyle\sum_{i=1}^{k-1} \langle v_k, u_i \rangle u_i \ , \ u_j \right\rangle \\ &= \langle v_k, u_j \rangle - \textstyle\sum_{i=1}^{k-1} \langle v_k, u_i \rangle \langle u_i, u_j \rangle. \end{aligned}$$

By the induction hypothesis, $\langle u_i, u_j \rangle = 0$ unless $i = j$, in which case it is 1, so the right hand side becomes $\langle v_k, u_j \rangle - \langle v_k, u_j \rangle = 0$, as required.

Finally we must check that $\operatorname{span}\{u_1, \ldots, u_k\} = \operatorname{span}\{v_1, \ldots, v_k\}$. By the induction hypothesis, we know $\operatorname{span}\{u_1, \ldots, u_{k-1}\} = \operatorname{span}\{v_1, \ldots, v_{k-1}\}$, so in particular $u_i \in \operatorname{span}\{v_1, \ldots, v_{k-1}\} \subset \operatorname{span}\{v_1, \ldots, v_k\}$ for $i = 1, \ldots, k-1$. By construction, we have $w_k \in \operatorname{span}\{u_1, \ldots, u_{k-1}, v_k\} \subset \operatorname{span}\{v_1, \ldots, v_{k-1}, v_k\}$, and therefore so is u_k. It follows that $\operatorname{span}\{u_1, \ldots, u_k\} \subset \operatorname{span}\{v_1, \ldots, v_k\}$. Since we have shown that $\{u_1, \ldots, u_k\}$ is orthonormal, we know it is linearly independent, so $\operatorname{span}\{u_1, \ldots, u_k\}$ has dimension k. We know $\operatorname{span}\{v_1, \ldots, v_k\}$ also has dimension k, so in fact $\operatorname{span}\{u_1, \ldots, u_k\} = \operatorname{span}\{v_1, \ldots, v_k\}$. $\square$

The Gram-Schmidt Process works in exactly the same way for real inner product spaces. For $\mathbb{R}^n$, it has the folowing geometric interpretation: first rescale v_1 to get a vector u_1 of length 1. Then the orthogonal projection of v_2 onto the line spanned by u_1 is $\langle v_2, u_1 \rangle u_1$; subtract this from v_2 to get w_2, and rescale to get u_2. Then the orthogonal projection of v_3 to the plane spanned by u_1 and u_2 is $\langle v_3, u_1 \rangle u_1 + \langle v_3, u_2 \rangle u_2$; subtract this from v_3 and rescale to get $u_3 \ldots$

Example 6.31. Let us apply the Gram-Schmidt Process to the vectors $\mathbf{v}_1 = (1, 0, i)^t$, $\mathbf{v}_2 = (-1, i, 1)^t$ and $\mathbf{v}_3 = (0, -1, 1+i)^t$ in $\mathbb{C}^3$. Then $\mathbf{w}_1 = \mathbf{v}_1$, $\mathbf{u}_1 = \frac{1}{\sqrt{2}} \mathbf{v}_1 = \frac{1}{\sqrt{2}}(1, 0, i)^t$, and

$$\mathbf{w}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{u}_1 \rangle \mathbf{u}_1 == \mathbf{v}_2 - \frac{1}{2} \langle \mathbf{v}_2, \mathbf{w}_1 \rangle \mathbf{w}_1.$$

Since $\langle \mathbf{v}_2, \mathbf{w}_1 \rangle = -1 - i$, this gives $\mathbf{w}_2 = \left(\frac{-1+i}{2}, i, \frac{1+i}{2})^t \right)$, so

$$\mathbf{u}_2 = \frac{1}{\sqrt{2}} \left(\frac{-1+i}{2}, i, \frac{1+i}{2} \right))^t.$$

Now

$$\mathbf{w}_3 = \mathbf{v}_3 - \frac{1}{2}\langle\mathbf{v}_3, \mathbf{w}_1\rangle\mathbf{w}_1 - \frac{1}{2}\langle\mathbf{v}_3, \mathbf{w}_2\rangle\mathbf{w}_2 = (\frac{i}{2}, -\frac{1+i}{2}, \frac{1}{2})^t$$

happens to have $||\mathbf{w}_3|| = 1$, so let $\mathbf{u}_3 = \mathbf{w}_3$.

We record some consequences of the Gram-Schmidt process.

Corollary 6.32.

(1) *Every subspace of $\mathbb{C}^n$ has an orthonormal basis;*
(2) *if $\mathbf{v} \in \mathbb{C}^n$ with $||\mathbf{v}|| = 1$, then $\mathbf{v}$ is the first column of a unitary matrix;*
(3) *if Q is an invertible $n \times n$ complex matrix, then $Q = UT$ for some unitary matrix U and upper-triangular matrix T.*

Proof. (1) Apply the Gram-Schmidt process to any basis $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ for the subspace.

(2) Extend $\{\mathbf{v}\}$ to a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for $\mathbb{C}^n$ with $\mathbf{v}_1 = \mathbf{v}$ (by Thm. 1.26 and then apply the Gram-Schmidt process to obtain an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$. Note that since $||\mathbf{v}_1|| = 1$, we have $\mathbf{u}_1 = \mathbf{v}_1 = \mathbf{v}$, and since $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is orthonormal, the matrix whose columns are $\mathbf{u}_1, \dots, \mathbf{u}_n$ is unitary.

(3) Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ be the columns of Q. If A is invertible, then $\operatorname{rank} Q = n$, so the columns span $\mathbb{C}^n$, hence form a basis. Now apply the Gram-Schmidt process to $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ to get an orthonormal basis $S' = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$. Then the matrix whose columns are $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ is unitary. Then $T = U^{-1}Q$ is the transition matrix from S to S'. Recall that the j^{th} column of T is given by the coordinates of $\mathbf{v}_j$ with respect to S'. Since $\mathbf{v}_j$ is in the span of $\{\mathbf{u}_1, \dots, \mathbf{u}_j\}$, its i^{th} coordinate is 0 for $i > j$, which means that T is upper-triangular. □

The analagous statements hold for vectors and subspaces of $\mathbb{R}^n$ and real matrices, replacing *unitary* with *orthogonal* throughout; the proofs are the same as for $\mathbb{C}$.

We discuss some examples before proceeding with the proof of unitary (resp. orthogonal) diagonalizability of normal (resp. real symmetric) matrices.

Example 6.33. Let V be the null space in $\mathbb{C}^4$ of the matrix

$$A = \begin{pmatrix} 1 & i & 0 & i+1 \\ 1 & 0 & 2 & i \end{pmatrix}.$$

We can find an orthonormal basis for V as follows. First find a basis for V in the usual way, by applying row operations to A to get

$$\begin{pmatrix} 1 & 0 & 2 & i \\ 0 & 1 & 2i & -i \end{pmatrix}.$$

We see therefore that $\mathbf{v}_1 = (-2, -2i, 1, 0)^t$, $\mathbf{v}_2 = (-i, i, 0, 1)^t$. Applying Gram-Schmidt to get an orthonormal basis gives $\mathbf{u}_1 = \frac{1}{3}(-2, -2i, 1, 0)^t$ and

$$\mathbf{w}_2 = (-i, i, 0, 1)^t + \frac{1}{9}(2+2i)(-2, -2i, 1, 0)^t = \frac{1}{9}(-5i-4, 5i-4, 2-2i, 9)^t.$$

Normalizing then gives $\mathbf{u}_2 = \frac{1}{3\sqrt{19}}(-5i-4, 5i-4, 2-2i, 9)^t$.

Example 6.34. Let $\mathbf{v} = \frac{1}{3}(1, 2, 2)^t \in \mathbb{R}^3$. Then $||\mathbf{v}|| = 1$, and we can extend $\{\mathbf{v}\}$ to a basis using, for example, the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2$. Applying the Gram-Schmidt process to $\{\mathbf{v}, \mathbf{e}_1, \mathbf{e}_2\}$ gives

$$\mathbf{w}_2 = \mathbf{e}_1 - \langle \mathbf{e}_1, \mathbf{v} \rangle \mathbf{v} = (1, 0, 0)^t - \frac{1}{9}(1, 2, 2)^t = \frac{2}{9}(4, -1, -1)^t.$$

Then $||\mathbf{w}_2|| = \frac{2}{9} \cdot 3\sqrt{2}$, so we let

$$\mathbf{u}_2 = \frac{\mathbf{w}_2}{||\mathbf{w}_2||} = \frac{1}{3\sqrt{2}}(4, -1, -1)^t.$$

Now $\langle \mathbf{e}_2, \mathbf{v} \rangle = \frac{2}{3}$ and $\langle \mathbf{e}_2, \mathbf{u}_2 \rangle = -\frac{1}{3\sqrt{2}}$, so we let

$$\begin{aligned} \mathbf{w}_3 &= \mathbf{e}_2 - \langle \mathbf{e}_2, \mathbf{v} \rangle \mathbf{v} - \langle \mathbf{e}_2, \mathbf{u}_2 \rangle \mathbf{u}_2 \\ &= (0, 1, 0)^t - \tfrac{2}{9}(1, 2, 2)^t + \tfrac{1}{18}(4, -1, -1)^t = \tfrac{1}{2}(0, 1, -1)^t, \end{aligned}$$

giving $\mathbf{u}_3 = \frac{1}{\sqrt{2}}(0, 1, -1)^t$. We thus obtain an orthogonal matrix

$$\begin{pmatrix} \frac{1}{3} & \frac{2}{\sqrt{3}} & 0 \\ \frac{2}{3} & \frac{-1}{2\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{2}{3} & \frac{-1}{2\sqrt{3}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$$

with first column $\mathbf{v} = \frac{1}{3}(1, 2, 2)^t$.

Example 6.35. Let Q be the matrix whose columns are the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ from Example **??**, so

$$Q = \begin{pmatrix} 1 & -1 & 0 \\ 0 & i & -1 \\ i & 1 & 1+i \end{pmatrix}.$$

We then have the matrix

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1+i}{2\sqrt{2}} & \frac{i}{2} \\ 0 & \frac{i}{\sqrt{2}} & \frac{-1-i}{2} \\ \frac{i}{\sqrt{2}} & \frac{1+i}{2\sqrt{2}} & \frac{1}{2} \end{pmatrix},$$

whose columns are the vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ gotten from the Gram-Schmidt Process. Now $T = U^{-1}Q = U^*Q =$

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{-i}{\sqrt{2}} \\ \frac{-1-i}{2\sqrt{2}} & \frac{-i}{\sqrt{2}} & \frac{1-i}{2\sqrt{2}} \\ \frac{-i}{2} & \frac{-1+i}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 0 & i & -1 \\ i & 1 & 1+i \end{pmatrix} = \begin{pmatrix} \sqrt{2} & \frac{-1-i}{\sqrt{2}} & \frac{1-i}{\sqrt{2}} \\ 0 & \sqrt{2} & \frac{1+i}{\sqrt{2}} \\ 0 & 0 & 1 \end{pmatrix}.$$

So far the only examples of inner product spaces we've considered were $\mathbb{R}^n$ and $\mathbb{C}^n$. For more examples, note that a subspace of an inner product space is still an inner product space. For more interesting examples, one can consider spaces of functions.

Example 6.36. Let V be the set of real polynomials of degree at most n. Define an inner product on V by

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx.$$

This satisfies the conditions in the definition of a real inner product space, since

(1) $\int_0^1(\alpha f(x)+\beta g(x))h(x)\,dx = \alpha\int_0^1 f(x)h(x)\,dx+\beta\int_0^1 g(x)h(x)\,dx$ shows that $\alpha\langle f,h\rangle+\beta\langle g,h\rangle$;
(2) $\langle g,f\rangle = \int_0^1 f(x)g(x)\,dx = \langle g,f\rangle$;
(3) $\langle f,f\rangle = \int_0^1 f(x)^2\,dx > 0$ unless $f(x)=0$.

In the context of abstract inner product spaces, the Gram-Schmidt Process has the following corollary.

Corollary 6.37. *Every inner product space has an orthonormal basis.*

Proof. Let $\{v_1,\ldots,v_n\}$ be any basis for V, and apply the Gram-Schmidt Process to get an orthonormal basis. □

We already saw how this works if V is a subspace of $\mathbb{C}^n$ (see Example **??**). Consider an example where V is a space of polynomials, as in Example **??**.

Example 6.38. Let V be the space of real polynomials of degree at most 2. Let us take the basis $\{f_1,f_2,f_3\}$ where $f_1(x)=1$, $f_2(x)=x$ and $f_3(x)=x^2$, and apply the Gram-Schmidt Process to get an orthonormal basis. Since $||f_1(x)|| = \int_0^1 dx = 1$, we do not need to rescale, and we let $h_1=f_1$. Since $\langle f_2,h_1\rangle = \int_0^1 x\,dx = \frac{1}{2}$, we set $g_2(x) = x-\frac{1}{2}$. Now

$$||g_2||^2 = \int_0^1\left(x-\frac{1}{2}\right)^2 dx = \frac{1}{3}\left(x-\frac{1}{2}\right)^3\Big|_0^1 = \frac{1}{12},$$

so $||g_2|| = 1/2\sqrt{3}$. Our second basis vector is therefore

$$h_2(x) = \frac{g_2(x)}{||g_2||} = 2\sqrt{3}\left(x-\frac{1}{2}\right) = \sqrt{3}(2x-1).$$

Since $\langle f_3,h_1\rangle = \int_0^1 x^2\,dx = \frac{1}{3}$ and

$$\begin{aligned}\langle f_3,h_2\rangle = \sqrt{3}\textstyle\int_0^1 x^2(2x-1)\,dx &= \sqrt{3}\textstyle\int_0^1(2x^3-x^2)\,dx\\ &= \sqrt{3}\left(\tfrac{1}{2}x^4-\tfrac{1}{3}x^3\right)\Big|_0^1 = \sqrt{3}\tfrac{1}{6},\end{aligned}$$

we set

$$g_3(x) = x^2-\frac{1}{3}-\frac{1}{2}(2x-1) = x^2-x+\frac{1}{6}.$$

Then another integral calculation gives $||g_3||^2 = 1/180$, so for the third vector in the orthonormal basis, take

$$h_3(x) = \frac{g_3(x)}{||g_3||} = \sqrt{5}(6x^2 - 6x + 1).$$

Example 6.39. If we momentarily drop the assumption that V be finite-dimensional, we can consider more interesting examples. Let V be the space of continuous complex-valued functions on the real unit interval $[0,1]$ (so $f \in V$ means that $f(x) = s(x) + it(x)$ where s and t are continuous real-valued functions on $[0,1]$). Define

$$\langle f, g\rangle = \int_0^1 f(x)\overline{g(x)}\,dx$$

for $f, g \in V$. This satisfies the axioms to be an inner product on V. (Since $\langle f, f\rangle$ is the integral of the continuous non-negative real-valued function $|f(x)|^2$ on $[0,1]$, its value is positive unless $|f(x)|^2$, and hence $f(x)$, are identically 0.) Consider the functions

$$f_n(x) = e^{2\pi inx} = \cos(2\pi nx) + i\sin(2\pi nx)$$

for $n \in \mathbb{Z}$. Then $\overline{f_n(x)} = e^{-2\pi inx} = \cos(2\pi nx) - i\sin(2\pi nx)$, $|f_n(x)|^2 = 1$, $||f_n|| = \int_0^1 dx = 1$, and if $m \neq n$, then

$$\begin{array}{rcl}
\langle f_m, f_n\rangle & = & \int_0^1 e^{2\pi imx}e^{-2\pi inx}\,dx \\
& = & \int_0^1 e^{2\pi i(m-n)x}\,dx \\
& = & \int_0^1 (\cos(2\pi(m-n)x) + i\sin(2\pi(m-n)x))\,dx \\
& = & \int_0^1 \cos(2\pi(m-n)x)\,dx + i\int_0^1 \sin(2\pi(m-n)x)\,dx \\
& = & 0.
\end{array}$$

Therefore $\{\, f_n \mid n \in \mathbb{Z}\}$ is an example of *infinite* orthonormal set. The set is linearly independent, by the same argument as for finite orthonormal sets.

Now we return to the task of proving unitary diagonalizability. We first prove:

Theorem 6.40.

(1) *Every square matrix is unitarily similar to an upper-triangular matrix.*
(2) *Every square matrix with only real eigenvalues is orthogonally similar to an upper-triangular matrix.*

Proof. (1) Recall from Theorem 5.18 that if A is a square matrix, then A is similar to an upper-triangular matrix, so $Q^{-1}AQ = T$ is upper-triangular for some invertible Q and upper-triangular T. By Corollary **??**(3), $Q = UT'$ for some unitary U and upper-triangular T'. It follows that

$$T = Q^{-1}AQ = (UT')^{-1}A(UT') = (T')^{-1}U^{-1}AUT',$$

and therefore $U^*AU = U^{-1}AU = T'T(T')^{-1}$ is upper-triangular.

(2) If A is a real square matrix with only real eigenvalues, then the proof of Theorem 5.18 goes through in exactly the same way to show that A is similar (over $\mathbb{R}$) to an upper-triangular matrix, and the proof of Corollary **??** goes through to show that if Q is real invertible, then $Q = PT$ for some orthogonal P and upper-triangular T. The proof of (2) is then the same as (1). $\square$

Theorem 6.41.

(1) *A matrix is unitarily diagonalizable if and only if it is normal.*
(2) *A real matrix is orthogonally diagonalizabe if and only if it is symmetric.*

Proof. (1) We already saw in Prop. 6.25 that if A is unitarily diagonalizable, then A is normal. We must prove the converse.

Suppose then that A is normal. By Thm. **??**, we know that $U^*AU = T$ for some unitary matrix U and upper-triangular matrix T. Since $AA^* = A^*A$ and $T^* = (U^*AU)^* = U^*A^*(U^*)^* = U^*A^*U$, it follows that

$$\begin{aligned} TT^* = (U^*AU)(U^*A^*U) &= U^*A(UU^*)A^*U = U^*(AA^*)U \\ &= U^*(A^*A)U = U^*A^*UU^*AU = T^*T; \end{aligned}$$

i.e., T is normal.

We complete the proof by showing that a normal upper-triangular $n \times n$ matrix is diagonal. We prove this by induction on n. The case $n = 1$ is obvious. Suppose then that $n > 1$ and the statement is true for $(n-1) \times (n-1)$ matrices. Let T be a normal upper-triangular $n \times n$ matrix. Since T is upper-triangular, we can write

$$T = \begin{pmatrix} \lambda & \mathbf{v}^t \\ \mathbf{0} & T_1 \end{pmatrix},$$

where $\lambda \in \mathbb{C}$, $\mathbf{v} \in \mathbb{C}^{n-1}$ and T_1 is an $(n-1) \times (n-1)$ matrix. Then $T^* = \begin{pmatrix} \overline{\lambda} & \mathbf{0} \\ \overline{\mathbf{v}} & T_1^* \end{pmatrix}$, so

$$TT^* = \begin{pmatrix} |\lambda|^2 + \mathbf{v}^t\overline{\mathbf{v}} & \mathbf{v}^tT_1^* \\ T_1\overline{\mathbf{v}} & T_1T_1^* \end{pmatrix} \quad \text{and} \quad T^*T = \begin{pmatrix} |\lambda|^2 & \overline{\lambda}\mathbf{v}^t \\ \lambda\overline{\mathbf{v}} & \overline{\mathbf{v}}\mathbf{v}^t + T_1^*T_1 \end{pmatrix}.$$

Since T is normal, $TT^* = T^*T$, and equating the upper-left entries gives $|\lambda|^2 + ||\mathbf{v}||^2 = |\lambda|^2$, so $||\mathbf{v}||^2 = 0$, which implies that $\mathbf{v} = 0$. Therefore $T = \begin{pmatrix} \lambda & \mathbf{0} \\ \mathbf{0} & T_1 \end{pmatrix}$, and $T_1T_1^* = T_1^*T_1$. Now T_1 is a normal upper-triangular $(n-1) \times (n-1)$ matrix, so by the induction hypothesis, T_1 is diagonal, and it follows that so is T.

(2) Again, we've already seen one implication in Prop. 6.25, namely that if A is real and orthogonally diagonalizable, then A is symmetric. We must prove the converse, so suppose that A is symmetric. By Thm. 6.27, the eigenvalues of A are all real, so Thm. **??** implies that $P^tAP = T$ for some orthogonal P and upper-triangular T. Now $T^t = P^tA^tP = P^tAP = T$, so T

is symmetric, and a symmetric upper-triangular matrix is clearly diagonal. □

Finding the matrix U (or P) as in the theorem amounts to finding an orthonormal basis of eigenvectors. If the eigenvalues are distinct then the eigenvectors are automatically orthogonal (see Theorem 6.27 for the real symmetric/orthogonal case; the normal/unitary case is similar), so only need to be rescaled to obtain an orthonormal basis (see Example 6.29). If the characteristic has repeated roots, then for each repeated root λ, we apply the Gram-Schmidt Process to the linearly independent eigenvectors with that eigenvalue to obtain an orthonornmal basis for the kernel of $A - \lambda I$.

Example 6.42. Let A denote the matrix

$$\begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}.$$

Since A is symmetric, we know by the theorem that is orthogonally diagonalizable. The characteristic polynomial of A is

$$\begin{aligned} p_A(x) &= (x-3)^2(x-6) + 16 + 16 - 16(x-6) - 4(x-3) - 4(x-3) \\ &= x^3 - 12x^2 + 21x + 98. \end{aligned}$$

We see that $x = -2$ is a root and $p_A(x)$ factors as $(x+2)(x^2 - 14x + 49) = (x+2)(x-7)^2$. Applying row operations to $A + 2I$ gives

$$\begin{pmatrix} 5 & -2 & 4 \\ -2 & 8 & 2 \\ 4 & 2 & 5 \end{pmatrix} \to \cdots \to \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{pmatrix},$$

so that $(-2,-1,2)^t$ is an eigenvector with eigenvalue -2. Normalizing gives $\mathbf{u}_1 = \frac{1}{3}(-2,-1,2)^t$. Applying row operations to $A - 7I$ gives

$$\begin{pmatrix} -4 & -2 & 4 \\ -2 & -1 & 2 \\ 4 & 2 & -4 \end{pmatrix} \to \cdots \to \begin{pmatrix} 1 & 1/2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

so that $(1,0,1)^t$ and $(1,-2,0)^t$ are linearly independent eigenvectors with eigenvalue 7. Applying the Gram-Schmidt Process to these gives $\mathbf{u}_2 = \frac{1}{\sqrt{2}}(1,0,1)^t$ and $\mathbf{u}_3 = \frac{1}{3\sqrt{2}}(-1,4,1)$. Therefore setting

$$P = \begin{pmatrix} \frac{-2}{3} & \frac{1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} \\ \frac{-1}{3} & 0 & \frac{2\sqrt{2}}{3} \\ \frac{2}{3} & \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} \end{pmatrix}$$

gives an orthogonal matrix such that P^tAP is the diagonal matrix with entries $-2, 7, 7$ on the diagonal.