PATRICK ORSON

LINEAR ALGEBRA III

(THE END BIT)

Contents

Introduction

At this stage you will be familiar with the following concepts from Axler's *Linear Algebra Done Right* (4th edition).

- 6A. Inner product spaces
- 6B. Orthonormal bases
- 6C. Orthogonal complements
- 3F. Dual spaces
- 7A. Self-adjoint operators
- 7B. The Spectral Theorems
- 7C. Positive operators
- 7D. Isometries of inner product spaces
- 8A. Generalised eigenspaces
- 8B. Generalised eigenspace decomposition
- 8C. Jordan Normal Form

We will now take a departure from Axler.

Invariants. The aim of the game in much of mathematics is to classify objects of interest – one major idea is to define enough (computable!) invariants so that two objects are equivalent if and only if the invariants are all equal. The major theme for the rest of the course is the idea of an **invariant** of a matrix. This is a crude quantity, e.g. a number, that can be extracted from a matrix using some formula. Matrices can be used to represent linear maps, but for the matrix invariant to depend only on the linear map, it shouldn't change when you choose a different matrix representation (it should be invariant...). We will investigate operator invariants using matrix invariants called the **trace** and the **determinant**. We then use matrix invariants to study another important object in linear algebra: **bilinear forms**.

Notation

- $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.
- $M_{r\times s}(\mathbb{F})$ is the set of $r \times s$ matrices over $\mathbb{F}$.
- $\mathcal{P}(\mathbb{F})$ is the set of polynomials with coefficients in $\mathbb{F}$.
- $\mathcal{P}_n(\mathbb{F})$ is the set of polynomials with coefficients in $\mathbb{F}$ of degree at most n.
- $\mathbb{F}^S$ is the set of all functions $f\colon S \to \mathbb{F}$ from a set S to $\mathbb{F}$.
- $\mathcal{C}(X)$ is the set of continuous functions $f\colon X \to \mathbb{R}$ from X to $\mathbb{F}$.
- $\mathcal{L}(V)$ is the set of all linear maps $T\colon V \to V$.
- $V^\vee = \mathrm{Hom}(V, \mathbb{F})$ is the set of all linear maps $\varphi\colon V \to \mathbb{F}$.
- $V^\vee = \mathrm{Hom}(V, \mathbb{F})$ is the set of all linear maps $\varphi\colon V \to \mathbb{F}$.

1 Trace and determinant

For this chapter, assume that V is a finite dimensional vector space.

Trace of an operator

We now begin the hunt for matrix invariants that are secretly operator invariants.

Definition 1.1. The *trace* of a square matrix $A \in M_{n\times n}(\mathbb{F})$ is the sum of all the entries on its diagonal

$$\operatorname{tr}(A) := A_{11} + A_{22} + \cdots + A_{nn} \in \mathbb{F}.$$

Here is an important property of the trace.

Proposition 1.2. *If $A \in M_{r\times s}(\mathbb{F})$ and $B \in M_{s\times r}(\mathbb{F})$ then*

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

Proof. The formulae for the entries of AB and BA are given by

$$(AB)_{ij} = \sum_{k=1}^{s} A_{ik}B_{kj} \quad \text{and} \quad (BA)_{ij} = \sum_{k=1}^{r} B_{ik}A_{kj}.$$

So the diagonal entries are

$$(AB)_{ii} = \sum_{k=1}^{s} A_{ik}B_{ki} \quad \text{and} \quad (BA)_{ii} = \sum_{k=1}^{r} B_{ik}A_{ki}.$$

Hence

$$\operatorname{tr}(AB) = \sum_{i=1}^{r} \left(\sum_{k=1}^{s} A_{ik}B_{ki} \right)$$

and so

$$\operatorname{tr}(BA) = \sum_{i=1}^{s} \left(\sum_{k=1}^{r} B_{ik}A_{ki} \right) = \sum_{k=1}^{s} \left(\sum_{i=1}^{r} B_{ki}A_{ik} \right) = \operatorname{tr}(AB),$$

where the second equality is just switching the roles of the dummy indices i and k. $\square$

Corollary 1.3. *If $A, B \in M_{n\times n}(\mathbb{F})$ are two matrices representing the same operator $T \in \mathcal{L}(V)$ then* $\text{tr}(A) = \text{tr}(B)$.

Proof. If A and B represent the same operator, then they are related by the change of basis formula $A = P^{-1}BP$, where $P \in M_{n\times n}(\mathbb{F})$ is a change of basis matrix. But then

$$\text{tr}(A) = \text{tr}(P^{-1}BP) = \text{tr}(P^{-1}PB) = \text{tr}(IB) = \text{tr}(B).$$

□

The previous corollary means the next definition makes sense.

Definition 1.4. The *trace* of any operator $T \in \mathcal{L}(V)$ is the trace of any matrix representing T.

Proposition 1.5. *The trace defines a linear map*

$$\text{tr}\colon \mathcal{L}(V) \to \mathbb{F}; \qquad T \mapsto \text{tr}(T)$$

such that $\text{tr}(ST) = \text{tr}(TS)$ *for all* $S, T \in \mathcal{L}(V)$.

Proof. As trace is independent of matrix representative choice, it is sufficient to prove the statement with matrices.
For additivity, we want to show $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$ for matrices A and B. But this is

$$\text{tr}(A+B) = \sum_{i=1}^{n}(A_{ii} + B_{ii}) = \left(\sum_{i=1}^{n} A_{ii}\right) + \left(\sum_{i=1}^{n} B_{ii}\right) = \text{tr}(A) + \text{tr}(B).$$

For homogeneity, let $\lambda \in \mathbb{F}$ and consider

$$\text{tr}(\lambda A) = \sum_{i=1}^{n}(\lambda A_{ii}) = \lambda \sum_{i=1}^{n} A_{ii}.$$

Finally, consider

$$\text{tr}(AB) = \text{tr}(BA),$$

as shown in Proposition 1.2. □

Finally, we can discuss the relationship between trace and eigenvalues.

Theorem 1.6. *If V is a complex vector space and $T \in \mathcal{L}(V)$, then*

$$\text{tr}(T) = d_1\lambda_1 + \ldots d_m\lambda_m,$$

where $\lambda_1, \ldots, \lambda_m$ is a list of all distinct eigenvalues of T and d_i is the multiplicity of λ_i.

Recall that $d_i = \dim G(T, \lambda_i)$, the dimension of the generalised eigenspace corresponding to λ_i.

Proof. Choose a basis of generalised eigenvalues. Then the eigenvalue λ_i appears on the diagonal of the corresponding matrix exactly d_i times. □

Corollary 1.7. *If V is a real vector space and $T \in \mathcal{L}(V)$, then* $\operatorname{tr}(T)$ *is given by taking a representative matrix A for T, considering it as a complex matrix, computing the generalised eigenspaces over the complex numbers, and computing*

$$\operatorname{tr}(T) = d_1\lambda_1 + \dots d_m\lambda_m,$$

where $\lambda_1, \dots, \lambda_m$ is a list of all distinct complex *eigenvalues of A and d_i is the multiplicity of λ_i.*

Matrix similarity and operator invariants

The idea of defining operator invariants by defining a matrix invariant $f(A)$ and then proving $f(AB) = f(BA)$ will now be studied in general. We recall some language and results about equivalence relations.

Definition 1.8. A relation $\sim$ on a set X is called an *equivalence relation* if it has the following properties.

(i) REFLEXIVE: $\forall x \in X, \quad x \sim x,$

(ii) SYMMETRIC: $\forall x, y \in X, \quad x \sim y \implies y \sim x,$

(iii) TRANSITIVE: $\forall x, y, z \in X, \quad (x \sim y \text{ and } y \sim z) \implies x \sim z.$

Given an equivalence relation $\sim$ on X, we have:

- An *equivalence class* $[x] = \{y \in X \mid x \sim y\} \subseteq X$, for every $x \in X$.
- The *equivalence set* $X/{\sim} = \{\, [x] \mid x \in X\}$.
- The *quotient map* $q\colon X \to X/{\sim};\ x \mapsto [x]$.

Recall, the equivalence set is a partition of X.

Example 1.9. The rational numbers $\mathbb{Q}$ is the equivalence set consisting of all equivalence classes of symbols $\frac{a}{b}$ where $a, b \in \mathbb{Z}$ and $b \neq 0$, with the relation $\frac{a}{b} \sim \frac{c}{d}$ whenever $ad = bc$.

Let us formalise an equivalence relation on matrices we have been talking about for a while.

Definition 1.10. Consider the set of matrices called the *general linear group*

$$GL_n(\mathbb{F}) := \{A \in M_{n\times n}(\mathbb{F}) \mid A \text{ is invertible}\} \subseteq M_{n\times n}(\mathbb{F}).$$

Exercise 1.1.

(a) Prove that $GL_n(\mathbb{F})$ is not a vector subspace of $M_{n\times n}(\mathbb{F}))$ by showing it is neither closed under addition nor scalar multiplication.

(b) Prove that $GL_n(\mathbb{F})$ is closed under matrix multiplication.

Subsets $G \subseteq GL_n(\mathbb{F})$ that contain the identity, are closed under matrix multiplication, and are such that $A \in G$ implies $A^{-1} \in G$ are called *matrix groups*.

Definition 1.11. We say $A, B \in M_{n\times n}(\mathbb{F})$ are *similar* is there exists $P \in \mathrm{GL}_n(\mathbb{F})$ such that $A = P^{-1}BP$.

We know very well that the instruction "choose a matrix representative for the linear operator $T \in \mathcal{L}(V)$" does not even produce a well-defined function $\mathcal{L}(V) \to M_{n\times n}(\mathbb{F})$. Each operator has MANY such matrix representations. However, the function IS well defined up to basis change matrices.

Exercise 1.2. Prove the following.

- Prove that the relation $A \sim B$ if and only if A and B are similar is an equivalence relation.
- Prove that $A \sim B$ and $C \sim D$ then $AC \sim BD$.

  In other words that $[A] \cdot [B] := [AB]$ is a well-defined multiplication on the equivalence set.
- Find a counterexample to the statement: if $A \sim B$ and $C \sim D$ then $(A + B) \sim (C + D)$.

  In other words, show that $[A] + [B] := [A + B]$ does not make sense as an addition. So $M_{n\times n}(\mathbb{F})/\sim$ is not a vector space in the obvious way.
- Prove that the assignment

  $$\varphi\colon \mathcal{L}(V) \to M_{n\times n}(\mathbb{F})/\sim; \qquad T \mapsto [A],$$

  where A is any choice of matrix representing T, is a well-defined bijection.
- Prove that $\varphi(ST) = \varphi(S)\varphi(T)$.

In other words, you are proving that φ is an **isomorphism of multiplicative groups**. But φ is not a vector space isomorphism as the codomain isn't even naturally a vector space.

Trying to define functions on an equivalence set using a formula made up of a chosen representative of an equivalence class is one of the most common ways to produce badly defined functions. For example $f\colon \mathbb{Q} \to \mathbb{Z}$, defined by $f(a/b) = a$ does not make sense because $f(1/2) = 1 \neq f(2/4)$.

The next definition describes when this method does make sense.

Definition 1.12. Suppose X has an equivalence relation $\sim$. We say a function $f\colon X \to Y$ *descends to the quotient* if the function

$$X/\sim \to Y; \qquad [x] \mapsto f(x)$$

is well-defined. Or equivalently, if $x \sim y$ implies $f(x) \sim f(y)$.

- I wish to associate a *numerical invariant* to a linear map T. In other words, define a function $\mathcal{L}(V) \to \mathbb{F}$ with nice properties that tells me something interesting about my map.
- Matrices are computationally very helpful. So I might try and define my invariant using a matrix representative for T. The catch is this will initially only produce a map $f\colon M_{n\times n}(\mathbb{F}) \to \mathbb{F}$.

The next theorem summarises the discussion above about construction operator invariants.

Theorem 1.13 (When a matrix invariant is an operator invariant)**.** *Let $f\colon M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ be a function. Then the function*

$$\widehat{f}\colon \mathcal{L}(V) \to \mathbb{F}; \qquad T \mapsto f(A),$$

where A is any matrix representation for T, is a well-defined function if and only if $A \sim B$ implies $f(A) \sim f(B)$.

A consequence Theorem 1.13 is that whenever $f(AB) = f(BA)$ we have that $\widehat{f}$ is a well-defined operator invariant. We saw this already with the trace. In detail: if $A \sim B$ then $A = P^{-1}BP$, so $f(A) = f(P^{-1}BP) = f(P^{-1}PB) = f(B)$.

Proof. You showed in Exercise 1.2 that the assignment

$$\varphi\colon \mathcal{L}(V) \to M_{n\times n}(\mathbb{F})/\sim; \qquad T \mapsto [A],$$

where A is any choice of matrix representing T, is a well-defined bijection. Furethermore, the map f descends to the quotient if and only if $A \sim B$ implies $f(A) \sim f(B)$. So the map $\widehat{f}$ is given by the (now well-defined) composition

$$\begin{array}{ccccc} \mathcal{L}(V) & \xrightarrow{\varphi} & M_{n\times n}(\mathbb{F})/\sim & \to & \mathbb{F} \\ T & \mapsto & [A] & \mapsto & f(A). \end{array}$$

$\square$

Determinant of a matrix

We begin with some examples.

Example 1.14. The determinant of a 2×2 matrix A is

$$\begin{aligned} \det A &= \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \\ &= A_{11}A_{22} - A_{12}A_{21} \\ &= A_{11}A_{22} - A_{1\sigma(1)}A_{2\sigma(2)} \end{aligned}$$

where $\sigma\colon \{1,2\} \to \{1,2\}$ is the permutation that switches 1 with 2.

The word *permutation* just means *bijection* in this context.

Example 1.15. The determinant of a 3×3 matrix A is

$$\begin{aligned} \det A &= \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} \\ &= A_{11}A_{22}A_{33} - A_{12}A_{21}A_{33} - A_{13}A_{22}A_{31} \\ &\qquad - A_{11}A_{23}A_{32} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} \\ &= A_{11}A_{22}A_{33} - A_{12}A_{21}A_{33} - A_{13}A_{22}A_{31} \\ &\qquad - A_{11}A_{23}A_{32} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32}. \end{aligned}$$

The pattern is that each term in the determinant is of the form

$$\pm A_{1\sigma(1)} A_{2\sigma(2)} A_{3\sigma(3)},$$

where $\sigma\colon \{1,2,3\} \to \{1,2,3\}$ is a permutation. Including the identity permutation, all six possible permutations of $\{1,2,3\}$ are being used.

To generalise this to $n \times n$ matrices, we need a little more language regarding permutations, to decide whether to use $+1$ of -1 as the sign.

Definition 1.16. The *symmetric group on n letters* is the set S_n of permutations

$$\sigma\colon \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}.$$

A *transposition* is a permutation σ that transposes two elements and fixes the remaining $n-2$ elements.

Example 1.17. Observe that the "3-cycle" $\sigma \in S_3$ given by

$$\sigma(1) = 2, \quad \sigma(2) = 3, \quad \sigma(3) = 1$$

can be decomposed into transpositions applied one after the other: "first transpose 1 and 2, then transpose 1 and 3". In fact this is true of any permutation as you will show in the next exercise.

Hint for Exercise 1.3: A cycle *is a closed loop, for example in S_4 the map $\sigma(1) = 2$, $\sigma(2) = 3$, $\sigma(3) = 1$, $\sigma(4) = 4$ has a 3-cycle $1 \to 2 \to 3 \to 1$ and a 1-cycle $4 \to 4$.*

Argue that every permutation is a composition of disjoint cycles. Next work out how to break each cycle into a composition of transpositions.

Exercise 1.3. Every $\sigma \in S_n$ is the composition of some number of transpositions.

The number of transpositions required to decompose a general permutation is not well-defined. However, the *parity* of number of transpositions required is well-defined. We will not prove this, but it is used to make the following definition.

For example, the identity map in S_2 can be written with 0 transpositions, but is also equal to σ^2 where σ transposes 1 and 2.

Definition 1.18. Given $\sigma \in S_n$, let k be the number of transpositions in some decomposition of σ into transpositions. The *sign* of σ is

$$\varepsilon(\sigma) := \begin{cases} +1 & \text{if } k \text{ is even} \\ -1 & \text{if } k \text{ is odd} \end{cases}$$

We now have the ingredients to define the determinant.

Definition 1.19. The determinant of $A \in M_{n\times n}(\mathbb{F})$ is given by

$$\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \ldots A_{n\sigma(n)}.$$

Exercise 1.4. Check that this agrees with the 2×2 and 3×3 cases. In particular, check you understand the signs in those cases.

Let's prove some useful lemmas!

Lemma 1.20. $\det A = \det A^T$

Proof. Given $\sigma \in S_n$, observe that

$$A_{\sigma(1)1}A_{\sigma(2)2}\dots A_{\sigma(n)n}$$

includes the numbers $1, 2, \dots, n$ jumbled up somewhere in the row-entry indices. So this can be shuffled back to the standard form

$$A_{1\tau(1)}A_{2\tau(2)}\dots A_{n\tau(n)}.$$

The τ is just notation for the result of shuffling the factors back to the standard form. Let's work out what τ is. We have, for each $i \in \{1, \dots, n\}$, that $(i, \tau(i)) = (\sigma(j), j)$ for some j. So $i = \sigma(j)$ and $j = \tau(i)$, implying $i = \sigma(\tau(i))$. As this is true for all i, we have $\sigma \circ \tau = \mathrm{Id}$. So $\tau = \sigma^{-1}$ is the inverse to σ. We note as well that $\varepsilon(\sigma) = \varepsilon(\sigma^{-1})$.
Let's compute! Write $B = A^T$. Then

$$\begin{aligned}
\det A^T = \det B &= \sum_{\sigma \in S_n} \varepsilon(\sigma) B_{1\sigma(1)}B_{2\sigma(2)}\dots B_{n\sigma(n)} \\
&= \sum_{\sigma \in S_n} \varepsilon(\sigma) A_{\sigma(1)1}A_{\sigma(2)2}\dots A_{\sigma(n)n} \\
&= \sum_{\sigma \in S_n} \varepsilon(\sigma) A_{1\sigma^{-1}(1)}A_{2\sigma^{-1}(2)}\dots A_{n\sigma^{-1}(n)} \\
&= \sum_{\sigma \in S_n} \varepsilon(\sigma^{-1}) A_{1\sigma^{-1}(1)}A_{2\sigma^{-1}(2)}\dots A_{n\sigma^{-1}(n)}
\end{aligned}$$

But this computes $\det A$. To see this, consider that as each $\sigma \in S_n$ corresponds to a unique $\sigma^{-1} \in S_n$, the sum is the same as if we had written

$$\sum_{\sigma \in S_n} \varepsilon(\sigma) A_{1\sigma(1)}A_{2\sigma(2)}\dots A_{n\sigma(n)}.$$

□

Lemma 1.21. *If A is upper triangular then* $\det A = A_{11}A_{22}\dots A_{nn}$.

Proof. The claim is that the only term in the formula for the determinant that is non-zero is the one corresponding to the identity permutation. To see this is the case, let $\sigma \in S_n$ be a non-identity permutation. Then for some $i \in \{1, \dots, n\}$ we must have $\sigma(i) < i$ (think about why this is true..). But as the matrix is upper triangular, we then have $A_{i\sigma(i)} = 0$, so the summand of the determinant corresponding to σ must vanish. □

Determinant of an operator

We wish to show that the determinant of a matrix representative of an operator is independent of the representative chosen. In other words we need to show $\det(P^{-1}BP) = \det B$, where P is a change of basis matrix. We begin with a lemma.

Lemma 1.22. *Let $A, A' \in M_{n\times n}(\mathbb{F})$ be the same matrix, except that the i^{th} and j^{th} rows are switched, for $i \neq j$. Then $\det A = -\det A'$. The same statement holds if the i^{th} and j^{th} columns are switched.*

Proof. We start with the switched-rows version. The resulting condition on the matrices is that $A_{rs} = A'_{rs}$ in all cases except when r is i or j, in which case we have $A_{is} = A'_{js}$ and $A_{js} = A'_{is}$. Let $\tau \in S_n$ by the transposition of i and j. Then $S_n \to S_n$ given by $\sigma \mapsto \sigma\tau$ is a bijection. Based on this discussion, we compute

$$\begin{aligned}
\det A &= \sum_{\sigma\in S_n} \varepsilon(\sigma) A_{1\sigma(1)} \dots A_{i\sigma(i)} \dots A_{j\sigma(j)} \dots A_{n\sigma(n)} \\
&= \sum_{\sigma\tau\in S_n} \varepsilon(\sigma\tau) A_{1\sigma\tau(1)} \dots A_{i\sigma\tau(i)} \dots A_{j\sigma\tau(j)} \dots A_{n\sigma\tau(n)} \\
&= -\sum_{\sigma\tau\in S_n} \varepsilon(\sigma) A_{1\sigma\tau(1)} \dots A_{i\sigma\tau(i)} \dots A_{j\sigma\tau(j)} \dots A_{n\sigma\tau(n)} \\
&= -\sum_{\sigma\tau\in S_n} \varepsilon(\sigma) A_{1\sigma(1)} \dots A_{i\sigma(j)} \dots A_{j\sigma(i)} \dots A_{n\sigma(n)} \\
&= -\sum_{\sigma\in S_n} \varepsilon(\sigma) A'_{1\sigma(1)} \dots A'_{i\sigma(i)} \dots A'_{j\sigma(j)} \dots A'_{n\sigma(n)}
\end{aligned}$$

So $\det A = -\det A'$ as claimed.
To see that the switched-columns version also holds, just use Lemma 1.20. □

Corollary 1.23. *If two rows (or two columns) of A are the same then $\det A = 0$.*

Proof. Switching the rows (or columns) that are the same changes the sign on the determinant, but does not change the matrix. Hence $\det A = -\det A$, implying $\det A = 0$. □

Corollary 1.24. *If A' is the result of permuting the columns (or rows) of A using the permutation $\sigma \in S_n$ then*

$$\det(A') = \varepsilon(\sigma) \det A.$$

Proof. Decompose σ into k transpositions and apply Lemma 1.22 k times, to pick up a sign of $(-1)^k = \varepsilon(\sigma)$. □

Lemma 1.25. *Let $A \in M_{n\times n}(\mathbb{F})$, fix $i \in \{1, \dots, n\}$ and denote by A_x the effect of replacing the ith row of A by a column vector $x \in \mathbb{F}^n$. Then the map*

$$T\colon V \to \mathbb{F}; \qquad v \mapsto \det(A_x)$$

is linear.
The similar statement is true if we replace the ith column.

Proof. We begin with the row-replacement version. We must first show that $\det A_{x+y} = \det A_x + \det A_y$. For this, we just compute

$$\begin{aligned}
&\det A_{x+y} \\
=&\sum_{\sigma\in S_n} \varepsilon(\sigma) A_{1\sigma(1)} \dots A_{(i-1)\sigma(i-1)} (x_{\sigma(i)} + y_{\sigma(i)}) A_{(i+1)\sigma(i+1)} \dots A_{n\sigma(n)} \\
=&\sum_{\sigma\in S_n} \varepsilon(\sigma) A_{1\sigma(1)} \dots A_{(i-1)\sigma(i-1)} x_{\sigma(i)} A_{(i+1)\sigma(i+1)} \dots A_{n\sigma(n)} \\
&\quad + \sum_{\sigma\in S_n} \varepsilon(\sigma) A_{1\sigma(1)} \dots A_{(i-1)\sigma(i-1)} y_{\sigma(i)} A_{(i+1)\sigma(i+1)} \dots A_{n\sigma(n)} \\
=&\det A_x + \det A_y.
\end{aligned}$$

Similarly, to show that $\det A_{\lambda x} = \lambda \det A_x$, just compute

$$\begin{aligned}
&\det A_{\lambda x} \\
=&\sum_{\sigma\in S_n} \varepsilon(\sigma) A_{1\sigma(1)} \dots A_{(i-1)\sigma(i-1)} (\lambda x_{\sigma(i)}) A_{(i+1)\sigma(i+1)} \dots A_{n\sigma(n)} \\
=&\lambda \sum_{\sigma\in S_n} \varepsilon(\sigma) A_{1\sigma(1)} \dots A_{(i-1)\sigma(i-1)} x_{\sigma(i)} A_{(i+1)\sigma(i+1)} \dots A_{n\sigma(n)} \\
=&\lambda \det A_x
\end{aligned}$$

For the column version, just use the row argument and then apply Lemma 1.20. □

We can finally prove the desired result.

Theorem 1.26. $\det(AB) = (\det A)(\det B)$.

Proof. Define $P(i, j)$ to be the effect of swapping the ith and jth columns of the identity matrix. Define $E(i, j)$ to be the matrix consisting of all 0's except for a 1 is the (i, j) entry. These matrices fully describe elementary row and column operations on a matrix B in the sense the effect of an elementary row or column operation on B is the matrix XB or BX where X is one of $X = P(i, j)$ or $X = I + \alpha E(i, j)$ for some $\alpha \in \mathbb{F}$
Any square matrix can be converted via elementary row and column operations to a diagonal matrix with 1's and 0's on the diagonal. Consequently there exist square matrices X_i such that $B = X_1 X_2 \dots X_m$ where X_i is one of $P(i, j)$, or $I + \alpha E(i, j)$ for some $\alpha \in \mathbb{F}$, or a diagonal matrix with only 1's and 0's on the diagonal.
We claim that if X is any of these three types then $\det(YX) = \det(Y)\det(X)$ for any square matrix Y. This will prove the theorem

by iteration, as

$$\begin{aligned}\det(AB) &= \det(AX_1 \dots X_m) \\ &= \det(AX_1 \dots X_{m-1}) \det(X_m) \\ &\vdots \\ &= \det(A) \det(X_1) \dots \det(X_m) \\ &= \det(A) \det(X_1 \dots X_m) \\ &= \det(A) \det(B).\end{aligned}$$

To prove the claim, start with $X = P(i, j)$. Then X is the effect of swapping the i and j columns of the identity matrix and YX is the effect of swapping the i and j columns of Y. By Lemma 1.22 we have $\det(YX) = -\det(Y)$ and $\det(X) = -\det(I) = -1$. Hence $\det(YX) = \det(Y) \det(X)$.
Now assume $X = I + \alpha E(i, j)$ for some $\alpha \in \mathbb{F}$. We have that YX is the effect of adding α times the ith column of Y to the jth column of Y. By Lemma 1.25, we thus have $\det(YX) = \det Y + \alpha \det Z$, where Z is the effect of taking Y, deleting the ith column and filling the jth column in its place. But now Z has two columns the same, so $\det Z = 0$. Finally, note that $\det X = 1$. Hence we see that $\det(YX) = \det(Y) \det(X)$ in this case.
Finally, if X is a diagonal matrix with 0's and 1's on its diagonal then $\det X = 0$ unless $X = I$, in which case $\det X = 1$. If there is a 0 on the diagonal of X, then YX has a column of all 0's, hence $\det(YX) = 0 = \det Y \det X$. If $X = I$, then clearly $\det(YX) = \det(Y) \det(X)$.
This completes the case analysis and hence the proof. □

Corollary 1.27. *If $A, B \in M_{n\times n}(\mathbb{F})$ are two matrices representing the same operator $T \in \mathcal{L}(V)$ then $\det(A) = \det(B)$.*

Proof. If A and B represent the same operator, then they are related by the change of basis formula $A = P^{-1}BP$, where $P \in M_{n\times n}(\mathbb{F})$ is a change of basis matrix. But then

$$\begin{aligned}\det(A) = \det(P^{-1}BP) &= \det(P^{-1}) \det(B) \det(P) \\ &= \det(P^{-1}PB) = \det(IB) = \det(B).\end{aligned}$$

□

The previous corollary means the next definition makes sense.

Definition 1.28. The *determinant* of any operator $T \in \mathcal{L}(V)$ is the determinant of any matrix representing T.

Proposition 1.29. *The determinant defines a linear map*

$$\det\colon \mathcal{L}(V) \to \mathbb{F}; \qquad T \mapsto \det(T)$$

such that moreover $\det(ST) = \det(S)\det(T)$ *for all* $S, T \in \mathcal{L}(V)$.

Finally, we can discuss the relationship between determinant and eigenvalues.

Theorem 1.30. *If V is a complex vector space and $T \in \mathcal{L}(V)$, then*

$$\det(T) = \lambda_1^{d_1} \dots \lambda_m^{d_m}$$

where $\lambda_1, \dots, \lambda_m$ is a list of all distinct eigenvalues of T and d_i is the multiplicity of λ_i.

Corollary 1.31. *If V is a real vector space and $T \in \mathcal{L}(V)$, then then*

$$\det(T) = \lambda_1^{d_1} \dots \lambda_m^{d_m}$$

where $\lambda_1, \dots, \lambda_m$ is a list of all complex *eigenvalues of T and d_i is the multiplicity of λ_i.*

Remark 1.32. Here we finally get an intuition for the geometric meaning of the determinant. In the complex case, we may choose a basis of generalised eigenvectors. We can roughly think of these vectors each being scaled by the corresponding eigenvalue, when the operator T is applied. (This is only actually the case if they are all actual eigenvectors, not just generalised.) Then the determinant is the product of all these scale factors at once. Think of it as the scale factor for the n-dimensional volume of an n-dimensional unit cube. If one of the eigenvalues is 0, one of the dimensions of the cube gets collapsed by the operator, so the volume scale factor is 0.

We finish with one of the most-used properties of the determinant.

Proposition 1.33. *An operator T on a finite dimensional vector space V is invertible if and only if* $\det T \neq 0$.
In this case

$$\det(T^{-1}) = (\det T)^{-1}.$$

Proof. Suppose $\det T \neq 0$. Assume, for a contradiction, that T is not invertible. Then T is not injective and there exists $v \in V \setminus \{0\}$ with $Tv = 0$. Let $v, e_2, \dots, e_n$ be a basis for V. Let A be the matrix representation of T in this basis. Then the first column of A is all 0's. So $\det T = \det A = 0$. This is a contradiction, implying T is invertible. Conversely, suppose T is invertible. Then

$$1 = \det(I) = \det(TT^{-1}) = (\det T)(\det(T^{-1})),$$

so in particular $\det T \neq 0$. □

The definition of determinant makes sense generally for square matrices $A \in M_{n\times n}(R)$ with entries in a commutative ring R, resulting in $\det A \in R$. For example $R = \mathbb{Z}$ is a very popular case. In general, the assumption that $\det A \neq 0$ is not strong enough to conclude invertibility. It must be replaced by the condition that $\det A \in R$ has a multiplicative inverse. For example, for integer matrices this means $\det A = \pm 1$ is the condition to check invertibility.

2 *Bilinear forms*

Coming soon!