Notes on maths for economists

Juan Alvaro Díaz Raimond Kedilhac

2024-12-09

Contents

Chapter 1

About

Welcome to this **maths for economists study guide**, a resource designed to support undergraduate students in their journey to understanding maths. These personal notes are based on lectures attended at **Universidad Panamericana**, alongside additional materials such as lecture notes, websites, books, and other sources.

While this guide is crafted to offer a quick and accessible approach to the subject, it is strongly recommended that students consult the original resources for a more comprehensive understanding.

1.1 Purpose

This site serves as a developing repository of knowledge, aimed at supplementing students' studies. Please note that the material is continually updated, so your patience is appreciated as we work on corrections and add new content.

1.2 Feedback

Your input is valuable! If you have comments, recommendations, or corrections, please reach out to us at **0251520@up.edu.mx**. We aim to address feedback and update the content promptly.

1.3 Support

If you find this project helpful and wish to support it, you can make a donation at buymeacoffee.com/jadrk040507. Your generosity is greatly appreciated!

Thank you for visiting, and happy learning!

Chapter 2

Linear Algebra

Based on Garrity and Pedersen [2007]

2.1 The Basic Vector Space $\mathbb{R}^n$

The quintessential vector space is $\mathbb{R}^n$, which is defined as the set of all n-tuples of real numbers. As we will see in the next section, what makes this set a vector space is the ability to add two n-tuples to obtain another n-tuple:

$$(v_1, \dots, v_n) + (w_1, \dots, w_n) = (v_1 + w_1, \dots, v_n + w_n)$$

and to multiply each n-tuple by a real number λ:

$$\lambda(v_1, \dots, v_n) = (\lambda v_1, \dots, \lambda v_n)$$

In this way, each n-tuple is commonly referred to as a vector, and the real numbers λ are known as scalars. When $n = 2$ or $n = 3$, this reduces to vectors in the plane and in space, which most of us learned in high school.

The natural relationship from $\mathbb{R}^n$ to $\mathbb{R}^m$ is established through matrix multiplication. We write a vector $x \in \mathbb{R}^n$ as a column vector:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Similarly, we can write a vector in $\mathbb{R}^m$ as a column vector with m entries. Let A be an $m \times n$ matrix:

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

Then, the product $A\mathbb{x}$ is the m-tuple:

$$A\mathbb{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

For any two vectors $\mathbb{x}$ and $\mathbb{y}$ in $\mathbb{R}^n$ and any two scalars λ and μ, the following property holds:

$$A(\lambda\mathbb{x} + \mu\mathbb{y}) = \lambda A\mathbb{x} + \mu A\mathbb{y}$$

In the next section, we will use the linearity of matrix multiplication to motivate the definition of a linear transformation between vector spaces. Now, let's relate all this to solving a system of linear equations. Suppose we are given numbers $b_1, \dots, b_m$ and numbers $a_{ij}, \dots, a_{mn}$. Our goal is to find n numbers $x_1, \dots, x_n$ that solve the following system of linear equations:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$
$$\vdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$$

Calculations in linear algebra often reduce to solving a system of linear equations. When there are only a few equations, we can find the solutions manually, but as the number of equations increases, the calculations quickly become less about pleasant algebraic manipulations and more about keeping track of many small individual details. In other words, it is an organizational problem.

We can write:

$$\mathbb{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

and our unknowns as:

$$\mathbb{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Then, we can rewrite our system of linear equations in the more visually appealing form:

$$A\mathbb{x} = \mathbb{b}$$

When $m > n$ (when there are more equations than unknowns), we generally do not expect solutions. For example, when $m = 3$ and $n = 2$, this corresponds geometrically to the fact that three lines in a plane generally do not have a common intersection point. When $m < n$ (when there are more unknowns than equations), we generally expect there to be many solutions. In the case where $m = 2$ and $n = 3$, this corresponds geometrically to the fact that two planes in space generally intersect in an entire line. Much of the machinery of linear algebra deals with the remaining case when $m = n$.

Therefore, we want to find the $n \times 1$ column vector $\mathbb{x}$ that solves $A\mathbb{x} = \mathbb{b}$, where A is a given $n \times n$ matrix and $\mathbb{b}$ is a given $n \times 1$ column vector.

Suppose the square matrix A has an inverse matrix A^{-1} (which means that A^{-1} is also $n \times n$ and, more importantly, that $AA^{-1} = I$, where I is the identity matrix). Then our solution will be:

$$\mathbb{x} = A^{-1}\mathbb{b}$$

because:

$$A\mathbb{x} = A(A^{-1}\mathbb{b}) = I\mathbb{b} = \mathbb{b}$$

Thus, solving our system of linear equations reduces to understanding when the $n \times n$ matrix A has an inverse. (If an inverse matrix exists, then algorithms exist for its calculation). The key theorem of linear algebra, which is stated in section six, is essentially a list of many equivalencies for when an $n \times n$ matrix has an inverse, and it is therefore crucial to understanding when a system of linear equations can be solved.

2.2 Vector Spaces and Linear Transformations

The abstract approach to studying systems of linear equations begins with the notion of a vector space.

Definition 2.1. A set V is a vector space over the real numbers $\mathbb{R}$ if there are two operations:

1. Scalar multiplication: For every $a \in \mathbb{R}$ and $\mathbb{v} \in V$, there is an element $a \cdot \mathbb{v} \in V$, denoted by $a\mathbb{v}$, which satisfies the properties of scalar multiplication.

2. Vector addition: For every $\mathbb{v}, \mathbb{w} \in V$, there is an element $\mathbb{v} + \mathbb{w} \in V$, denoted by $\mathbb{v} + \mathbb{w}$, which satisfies the properties of vector addition.

These operations must satisfy the following properties:

a) **Additive identity**: There exists an element $\mathbb{0} \in V$ (the zero vector) such that $\mathbb{0} + \mathbb{v} = \mathbb{v}$ for all $\mathbb{v} \in V$.

b) **Additive inverse**: For each $\mathbb{v} \in V$, there exists an element $-\mathbb{v} \in V$ such that $\mathbb{v} + (-\mathbb{v}) = \mathbb{0}$.

c) **Commutativity**: For all $\mathbb{v}, \mathbb{w} \in V$, it holds that $\mathbb{v} + \mathbb{w} = \mathbb{w} + \mathbb{v}$.

d) **Distributivity of scalar multiplication over vector addition**: For all $a \in \mathbb{R}$ and for all $\mathbb{v}, \mathbb{w} \in V$, we have $a(\mathbb{v} + \mathbb{w}) = a\mathbb{v} + a\mathbb{w}$.

e) **Distributivity of scalar multiplication over scalar addition**: For all $a, b \in \mathbb{R}$ and for all $\mathbb{v} \in V$, it holds that $(a + b)\mathbb{v} = a\mathbb{v} + b\mathbb{v}$.

f) **Compatibility of scalar multiplication with field multiplication**: For all $a, b \in \mathbb{R}$ and all $\mathbb{v} \in V$, it holds that $a(b\mathbb{v}) = (ab)\mathbb{v}$.

g) **Multiplicative identity**: For all $\mathbb{v} \in V$, it holds that $1 \cdot \mathbb{v} = \mathbb{v}$, where 1 is the multiplicative identity in $\mathbb{R}$.

Real numbers can be replaced by complex numbers and, indeed, by any field.

As a matter of notation, and to be consistent with common usage, the elements of a vector space are called vectors, and the elements of $\mathbb{R}$ (or any field being used) are called scalars. It is worth noting that the space $\mathbb{R}^n$ mentioned in the previous section satisfies these conditions.

The natural map between vector spaces is that of a linear transformation.

:::{.definition} A linear transformation $T : V \to W$ is a function from a vector space V to a vector space W such that for any real numbers a_1 and a_2, and for any vectors v_1 and v_2 in V, it holds that

$$T(a_1 v_1 + a_2 v_2) = a_1 T(v_1) + a_2 T(v_2).$$

Matrix multiplication from $\mathbb{R}^n$ to $\mathbb{R}^m$ provides an example of a linear transformation. :::

:::{.definition} A subset U of a vector space V is a subspace of V if U itself is a vector space.

In practice, it is usually easy to determine if a subset of a vector space is, in fact, a subspace, using the following proposition, whose proof is left to the reader. :::

:::{.proposition} A subset U of a vector space V is a subspace of V if U is closed under addition and scalar multiplication.

Given a linear transformation $T : V \to W$, there are naturally occurring subspaces both in V and W. :::

:::{.definition} If $T : V \to W$ is a linear transformation, then the kernel of T is:

$$\ker(T) = \{v \in V : T(v) = 0\}$$

and the image of T is

$$\operatorname{Im}(T) = \{w \in W : \text{there exists } v \in V \text{ such that } T(v) = w\}.$$

The kernel is a subspace of V, since if v_1 and v_2 are two vectors in the kernel and a and b are any two real numbers, then

$$T(av_1 + bv_2) = aT(v_1) + bT(v_2) = a \cdot 0 + b \cdot 0 = 0.$$

:::

Similarly, we can prove that the image of T is a subspace of W.

If the only vector spaces that existed were column vectors in $\mathbb{R}$, then even this level of abstraction would be trivial. However, this is not the case.

Here we consider just one example. Let $C^*[0,1]$ be the set of all real-valued functions with domain on the unit interval $[0,1]$:

$$f : [0,1] \to \mathbb{R}$$

such that the k-th derivative of f exists and is continuous. Since the sum of any two such functions and a multiple of any of these functions by a scalar will remain in $C^*[0,1]$, we have a vector space. Although we will define the dimension officially in the next section, $C^*[0,1]$ will have infinite dimension (and thus will definitely not be some $\mathbb{R}^n$). We can see the derivative as a linear transformation from $C^*[0,1]$ to those functions with one derivative less, $C^{k-1}[0,1]$:

$$\frac{d}{dx} : C^*[0,1] \to C^{k-1}[0,1].$$

The kernel of this transformation consists of those functions whose k-th derivative is zero, that is, constant functions.

Now consider the differential equation

$$f'' + 3f' + 2f = 0.$$

Let T be the linear transformation:

$$T : C^*[0,1] \to C^*[0,1]$$

defined by

$$T(f) = f'' + 3f' + 2f.$$

The problem of finding a solution $f(x)$ to the original differential equation can now be translated into finding an element in the kernel of T. This suggests the possibility (which is indeed true) that the language of linear algebra can be used to understand solutions to (linear) differential equations.

2.3   Bases, Dimension, and Linear Transformations as Matrices

Our next goal is to define the dimension of a vector space.

Definition 2.2. A set of vectors $(v_1, \dots, v_n)$ forms a basis for the vector space V if, for any vector $v \in V$, there exist unique scalars $a_1, \dots, a_n \in \mathbb{R}$ such that

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Definition 2.3. The dimension of a vector space V, denoted as $\dim(V)$, is the number of elements in a basis.

It is not obvious that the number of elements in a basis will always be the same, regardless of the chosen basis. To ensure that the definition of the dimension of a vector space is well-defined, we need the following theorem (which we will not prove):

Definition 2.4. All bases of a vector space V have the same number of elements.

For $\mathbb{R}^n$, the usual basis is $\{(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)\}$. Therefore, $\mathbb{R}^n$ has dimension n. If this were not the case, the previous definition of dimension would be incorrect and we would need another. This is an example of the principle mentioned in the introduction. We have an intuitive understanding of what the dimension should mean for specific examples: a line should be one-dimensional, a plane two-dimensional, and three-dimensional space. We then

formulate a precise definition. If this definition gives the "correct answer" for our three already understood examples, we are somewhat confident that the definition has captured what dimension means in this case. We can then apply the definition to examples where our intuitions fail.

Linked to the idea of a basis is:

Definition 2.5. The vectors $(v_1, \ldots, v_n)$ in a vector space V are linearly independent if, whenever $a_1 v_1 + \cdots + a_n v_n = 0$, it must be the case that the scalars $a_1, \ldots, a_n$ are all zero. Intuitively, a set of vectors is linearly independent if they all point in different directions. A basis, therefore, consists of a set of linearly independent vectors that span the vector space, where "span" means:

Definition 2.6. A set of vectors $(v_1, \ldots, v_n)$ spans the vector space V if, for any vector $v \in V$, there exist scalars $a_1, \ldots, a_n \in \mathbb{R}$ such that

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Our next goal is to show how all linear transformations $T : V \to W$ between finite-dimensional spaces can be represented as matrix multiplication, provided that we fix bases for the vector spaces V and W.

First, we fix a basis $\{v_1, \ldots, v_n\}$ for V and a basis $\{w_1, \ldots, w_m\}$ for W. Before examining the linear transformation T, we need to show how each element of the n-dimensional space V can be represented as a column vector in $\mathbb{R}^n$ and how each element of the m-dimensional space W can be represented as a column vector in $\mathbb{R}^m$. Given any vector $v \in V$, by the definition of a basis, there exist unique real numbers $a_1, \ldots, a_n$ such that:

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Thus, we represent the vector v with the column vector:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Similarly, for any vector $w \in W$, there exist unique real numbers $b_1, \ldots, b_m$ such that:

$$w = b_1 w_1 + \cdots + b_m w_m.$$

Here, we represent w as the column vector:

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

It is important to note that we have established a correspondence between the vectors in V and W and the column vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. More technically, we can prove that V is isomorphic to $\mathbb{R}^n$ (which means there is a one-to-one and onto linear transformation from V to $\mathbb{R}^n$) and that W is isomorphic to $\mathbb{R}^m$, although it must be emphasized that the real correspondence only exists after a basis has been chosen (which means that, while the isomorphism exists, it is not canonical; this is an important aspect because, in practice, we are often not provided with a basis).

Now, we want to represent a linear transformation $T : V \to W$ as a matrix A of size $m \times n$. For each basis vector v_i in the vector space V, $T(v_i)$ will be a vector in W. Therefore, there will be real numbers $a_{ij}, \dots, a_{im}$ such that:

$$T(v_i) = a_{ij}w_1 + \cdots + a_{im}w_m.$$

We want to see that the linear transformation T corresponds to the matrix A:

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

Given any vector $v \in V$, with $v = a_1 v_1 + \cdots + a_n v_n$, we have:

$$T(v) = a_1 T(v_1) + \cdots + a_n T(v_n) = a_1(a_{11}w_1 + \cdots + a_{1m}w_m) + \cdots + a_n(a_{n1}w_1 + \cdots + a_{nm}w_m).$$

Under the correspondences of the vector spaces with the respective column spaces, this can be seen as the matrix multiplication of A by the column vector corresponding to the vector v:

$$A \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

It is important to note that if $T : V \to V$ is a linear transformation of a vector space onto itself, then the corresponding matrix will be $n \times n$, i.e., a square matrix.

Given different bases for the vector spaces V and W, the matrix associated with the linear transformation T will change. A natural problem is to determine when two matrices actually represent the same linear transformation, but under different bases. This will be the subject of section seven.

2.4 The Determinant

Our next task is to define the determinant of a matrix. In fact, we will present three alternative descriptions of the determinant. These descriptions are equivalent, and each has its own advantages.

The first method involves defining the determinant of a 1×1 matrix and then recursively defining the determinant of an $n \times n$ matrix. Since 1×1 matrices are simply numbers, the following should not be surprising:

Definition 2.7. The determinant of a 1×1 matrix a is the real-valued function:

$$\det(a) = a.$$

This should not yet seem significant.

Before defining the determinant for a general $n \times n$ matrix, we need some notation. For an $n \times n$ matrix:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$

we denote by A_{ij} the $(n-1) \times (n-1)$ matrix obtained from A by removing the i-th row and j-th column. For example, if

$$A = \begin{pmatrix} 2 & 6 & 1 \\ 3 & 4 & 8 \\ 5 & 9 & 7 \end{pmatrix},$$

then

$$A_{23} = \begin{pmatrix} 2 & 6 \\ 5 & 9 \end{pmatrix}.$$

$$A_{23} = \begin{pmatrix} 2 & 1 \\ 3 & 8 \end{pmatrix}.$$

Similarly, if

$$A = \begin{pmatrix} 6 & 7 \\ 4 & 1 \\ 9 & 8 \end{pmatrix},$$

then

$$A_{12} = \begin{pmatrix} 7 & 8 \end{pmatrix}.$$

Since we have a definition for the determinant of 1×1 matrices, we will assume by induction that we know the determinant of any $(n-1) \times (n-1)$ matrix and use this to find the determinant of an $n \times n$ matrix.

Definition 2.8. Let A be an $n \times n$ matrix. The determinant of A is defined as:

$$\det(A) = \sum_{k=1}^{n} a_{1k} \det(A_{1k}).$$

Thus, for

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

we have:

$$\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}),$$

which aligns with the determinant formula most of us think of. The determinant of our earlier 3×3 matrix is:

$$\det \begin{pmatrix} 2 & 6 & 1 \\ 3 & 4 & 8 \\ 5 & 9 & 7 \end{pmatrix} = 2 \det(A_{11}) - 3 \det(A_{12}) + 5 \det(A_{13}).$$

Although this definition is an efficient way to describe the determinant, it obscures many of its uses and intuitions.

The second way to describe the determinant incorporates its key algebraic properties. It highlights the functional properties of the determinant. Denote the $n \times n$ matrix A as $A = (A_1, A_2, \ldots, A_n)$, where A_i denotes the i-th column:

Definition 2.9. The determinant of A is the unique real-valued function:

$$\det : \text{Matrices} \to \mathbb{R},$$

that satisfies:
1. $\det(A_1, \ldots, cA_k, \ldots, A_n) = c \det(A_1, \ldots, A_k, \ldots, A_n)$.
2. $\det(A_1, \ldots, A_k + A_{k'}, \ldots, A_n) = \det(A_1, \ldots, A_k, \ldots, A_n)$ for $k \neq k'$.
3. $\det(\text{Identity Matrix}) = 1$.

Thus, treating each column vector of a matrix as a vector in $\mathbb{R}^n$, the determinant can be viewed as a special function from $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ to the real numbers.

To use this definition, one would need to prove that such a function on the space of matrices, satisfying these conditions, exists and is unique. Existence can be shown by verifying that our first (inductive) definition satisfies these conditions, although this involves tedious computation. The proof of uniqueness can be found in most linear algebra texts.

The third definition of the determinant is the most geometric but also the vaguest. Think of an $n \times n$ matrix A as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. Then, A maps the unit cube in $\mathbb{R}^n$ to some other object (a parallelepiped). The unit cube has, by definition, a volume of one.

Definition 2.10. The determinant of the matrix A is the signed volume of the image of the unit cube.

This is not well-defined since the method of defining the volume of the image has not been described. In fact, most would define the signed volume of the image as the number given by the determinant using one of the two previous definitions. However, this can be made rigorous, albeit at the cost of losing much of the geometric intuition.

For example, the matrix

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

maps the unit cube to a region with doubled area, so we must have:

$$\det(A) = 2.$$

Signed volume means that if the orientations of the edges of the unit cube are reversed, we must have a negative sign for the volume. For example, consider the matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Here, the image is:

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

Note that the orientations of the sides are reversed. Since the area is still doubled, the definition forces:

$$\det(A) = -2.$$

Defining orientation rigorously is somewhat complicated (we will do this in Chapter Six), but its meaning is straightforward.

The determinant has many algebraic properties. For example:

Lemma 2.1. *If A and B are $n \times n$ matrices, then:*

$$\det(AB) = \det(A) \cdot \det(B).$$

This can be proven via lengthy calculation or by focusing on the definition of the determinant as the change in volume of the unit cube.

2.5   The Key Theorem of Linear Algebra

Below is the fundamental theorem of linear algebra. *(Note: We have not yet defined eigenvalues or eigenvectors, but we will do so in Section 8.)*

Theorem 2.1. *(Key Theorem 1.6.1) Let A be an $n \times n$ matrix. Then, the following statements are equivalent:*
1. A is invertible.
2. $\det(A) \neq 0$.
3. $\ker(A) = \{0\}$.
4. If $\mathbf{b}$ is a column vector in $\mathbb{R}^n$, there exists a unique column vector $\mathbf{x}$ in $\mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{b}$.
5. The columns of A are linearly independent $n \times 1$ column vectors.
6. The rows of A are linearly independent $1 \times n$ row vectors.
7. The transpose $A^\top$ of A is invertible. (Here, if $A = (a_{ij})$, then $A^\top = (a_{ji})$).
8. All eigenvalues of A are nonzero.

We can restate this theorem in terms of linear transformations:

Theorem 2.2. *(Key Theorem 1.6.2) Let $T : V \to V$ be a linear transformation. Then, the following statements are equivalent:*
1. T is invertible.
2. $\det(T) \neq 0$, where the determinant is defined via a choice of basis in V.
3. $\ker(T) = \{0\}$.
4. If $\mathbf{b}$ is a vector in V, there exists a unique vector $\mathbf{v}$ in V such that $T(\mathbf{v}) = \mathbf{b}$.
5. For any basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V, the image vectors $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent.
6. For any basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V, if S denotes the transpose linear transformation of T, then the image vectors $S(\mathbf{v}_1), \ldots, S(\mathbf{v}_n)$ are linearly independent.
7. The transpose $T^\top$ is invertible. (Here, the transpose is defined via a choice of basis in V.)
8. All eigenvalues of T are nonzero.

To clarify the correspondence between the two theorems, we note that we currently have definitions for the determinant and the transpose only for matrices, not for linear transformations. However, both notions can be extended to linear transformations by fixing a basis (or equivalently, by choosing an inner product, which will be defined in Chapter 13 on Fourier series).

It is important to note that while the actual value of $\det(T)$ depends on the chosen basis, the condition $\det(T) \neq 0$ does not. Similar remarks apply to conditions (6) and (7).

An exercise (Exercise 7) encourages the reader to find any linear algebra textbook and complete the proof of this theorem. It is unlikely that the textbook will present the result in this exact form, making the act of translation itself part of the exercise's purpose.

Each of these equivalences is significant and can be studied in its own right. It is remarkable that they are all equivalent.

2.6 Similar Matrices

Recall that given a coordinate system for a vector space V of dimension n, we can represent a linear transformation $T : V \to V$ as an $n \times n$ matrix A. However, if we choose a different coordinate system V', the matrix representing the linear transformation T' will generally differ from the original matrix A. The goal of this section is to establish a clear criterion to determine when two matrices represent the same linear transformation under different choices of bases.

Definition 2.11. Two $n \times n$ matrices A and B are **similar** if there exists an invertible matrix C such that

$$A = C^{-1}BC.$$

We aim to show that two matrices are similar precisely when they represent the same linear transformation. Let us choose two bases for the vector space V, say $\{v_1, \dots, v_n\}$ (the v-basis) and $\{w_1, \dots, w_n\}$ (the w-basis). Let A be the matrix representing the linear transformation T with respect to the v-basis, and B the matrix representing T with respect to the w-basis. We seek to construct the matrix C such that

$$A = C^{-1}BC.$$

Recall that given the v-basis, any vector $\mathbf{z} \in V$ can be written as an $n \times 1$ column vector as follows: there exist unique scalars $a_1, \dots, a_n$ such that

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

We then represent $\mathbf{z}$ with respect to the v-basis as the column vector:

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Similarly, there exist unique scalars $b_1, \ldots, b_n$ such that

$$\mathbf{z} = b_1 \mathbf{w}_1 + \cdots + b_n \mathbf{w}_n,$$

which means that with respect to the w-basis, the vector $\mathbf{z}$ is represented as:

$$\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

The desired matrix C is the one that satisfies the relationship:

$$CA = BC.$$

Determining when two matrices are similar is a result that frequently arises in mathematics and physics. Often, we need to select a coordinate system (a basis) to express anything, but the underlying mathematics or physics of interest is independent of the initial choice. The key question then becomes: *What is preserved when the coordinate system changes?* Similar matrices provide a starting point to understand these issues.

2.7   Eigenvalues and Eigenvectors

In the previous section, we observed that two matrices represent the same linear transformation under different choices of bases precisely when they are similar. However, this does not tell us how to choose a basis for a vector space so that a linear transformation has a particularly simple matrix representation. For instance, the diagonal matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

is similar to the matrix

$$B = \begin{pmatrix} 1 & -5 & 15 \\ 1 & -1 & 84 \\ -4 & -15 & 15 \end{pmatrix},$$

but the simplicity of A is evident compared to B. (Incidentally, it is not obvious that A and B are similar; I started with A, chose a nonsingular matrix C, and then used Mathematica software to compute $C^{-1}AC$ to obtain B. This was not immediately apparent but intentionally set up.)

One of the purposes of the following definitions of eigenvalues and eigenvectors is to provide tools for selecting good bases. However, there are many other reasons to understand eigenvalues and eigenvectors.

Definition 2.12. Let $T : V \to V$ be a linear transformation. A nonzero vector $\mathbf{v} \in V$ is an eigenvector of T with eigenvalue λ, a scalar, if

$$T(\mathbf{v}) = \lambda\mathbf{v}.$$

For an $n \times n$ matrix A, a nonzero column vector $\mathbf{x} \in \mathbb{R}^n$ is an eigenvector with eigenvalue λ, a scalar, if

$$A\mathbf{x} = \lambda\mathbf{x}.$$

Geometrically, a vector $\mathbf{v}$ is an eigenvector of the linear transformation T with eigenvalue λ if T stretches $\mathbf{v}$ by a factor of λ.

For example, consider the matrix

$$\begin{pmatrix} -2 & 1 \\ 2 & 7 \end{pmatrix}.$$

Here, $\lambda = 2$ is an eigenvalue, and $\mathbf{v}$ is an eigenvector for the linear transformation represented by this 2×2 matrix.

Fortunately, there is a simple way to describe the eigenvalues of a square matrix, allowing us to see that the eigenvalues of a matrix are preserved under a similarity transformation.

Definition 2.13. A scalar λ is an eigenvalue of a square matrix A if and only if λ is a root of the polynomial

$$P(t) = \det(tI - A),$$

where $P(t)$ is called the characteristic polynomial of A.

Proof: Suppose λ is an eigenvalue of A, with eigenvector $\mathbf{v}$. Then,

$$A\mathbf{v} = \lambda\mathbf{v},$$

which implies

$$A\mathbf{v} - \lambda\mathbf{v} = 0.$$

Introducing the identity matrix I, we rewrite this as

$$0 = (\lambda I - A)\mathbf{v}.$$

Thus, the matrix $\lambda I - A$ has a nontrivial kernel, $\mathbf{v}$. By the key theorem of linear algebra, this occurs precisely when

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

which means λ is a root of the characteristic polynomial $P(t) = \det(tI - A)$. Since these implications are reversible, the theorem is established.

Theorem 2.3. *If A and B are similar matrices, then the characteristic polynomial of A is equal to the characteristic polynomial of B.*

Proof: For A and B to be similar, there must exist an invertible matrix C such that $A = C^{-1}BC$. Then,

$$\det(tI - A) = \det(tI - C^{-1}BC) = \det(C^{-1}(tI - B)C) = \det(C^{-1})\det(tI - B)\det(C) = \det(tI - B),$$

using the property $\det(C^{-1}C) = 1 = \det(C^{-1})\det(C)$.

Since similar matrices have the same characteristic polynomial, it follows that their eigenvalues must also be the same.

Corollary 2.1. *The eigenvalues of similar matrices are identical. Thus, to determine if two matrices are not similar, check whether their eigenvalues differ. If they do, the matrices are not similar.*

However, in general, identical eigenvalues do not imply that matrices are similar. For example, the matrices

$$A = \begin{pmatrix} 4 & 2 \\ 7 & 5 \end{pmatrix} \quad and \quad B = \begin{pmatrix} 2 & 3 \\ 5 & 4 \end{pmatrix}$$

have eigenvalues 1 and 2, but they are not similar. (This can be shown by assuming the existence of an invertible matrix C such that $C^{-1}AC = B$ and then demonstrating that $\det(C) = 0$, which contradicts the invertibility of C.)

2.7.1 Eigenvalues and Determinants

Since the characteristic polynomial $P(t)$ does not change under a similarity transformation, the coefficients of $P(t)$ are also preserved. As these coefficients are polynomial functions (complex) of the entries of the matrix A, we obtain special polynomials of the entries of A that are invariant under similarity transformations. One such coefficient, already seen in another form, is the determinant of A, as highlighted in the following theorem. This theorem establishes a deeper connection between the eigenvalues of A and the determinant of A.

Theorem 2.4. *Let* $\lambda_1, \lambda_2, \dots, \lambda_n$ *be the eigenvalues (counted with multiplicity) of a matrix A. Then,*

$$\det(A) = \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_n.$$

Before proving this theorem, we need to clarify the concept of counting eigenvalues "with multiplicity." The difficulty arises because a polynomial can have a root that must be counted more than once. For example, the polynomial $(z-2)^2$ has a single root 2, but we count it twice. This situation is particularly relevant for the characteristic polynomial.

For instance, consider the matrix

$$A = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 4 \end{pmatrix},$$

which has the characteristic polynomial $P(t) = (t-5)(t-5)(t-4)$. Here, we list the eigenvalues as $4, 5$, and 5, counting the eigenvalue 5 twice.

Proof: Since the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ are the roots of the characteristic polynomial $\det(tI - A)$, we can write

$$\det(tI - A) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n).$$

Setting $t = 0$, we obtain

$$(-1)^n \lambda_1 \lambda_2 \cdots \lambda_n = \det(-A).$$

In the matrix $-A$, each column of A is multiplied by -1. Using the linearity of the determinant, we can factor out -1 from each column, yielding

$$\det(-A) = (-1)^n \det(A).$$

Substituting, we have

$$(-1)^n \lambda_1 \lambda_2 \cdots \lambda_n = (-1)^n \det(A),$$

and thus,

$$\det(A) = \lambda_1 \cdot \lambda_2 \cdots \lambda_n.$$

To find a "good" basis for representing a linear transformation, we measure "goodness" by how close the matrix is to being diagonal. Here, we focus on a special but common class of matrices: symmetric matrices. A matrix A is symmetric if $A = (a_{ij})$ satisfies $a_{ij} = a_{ji}$, meaning the element in the i-th row and j-th column equals the element in the j-th row and i-th column.

For example:

$$\begin{pmatrix} 5 & 3 & 4 \\ 3 & 5 & 2 \\ 4 & 2 & 4 \end{pmatrix}$$

is symmetric, but

$$\begin{pmatrix} 5 & 6 & 2 \\ 2 & 5 & 18 \\ 3 & 3 & 4 \end{pmatrix}$$

is not.

Theorem 2.5. *If A is a symmetric matrix, then there exists a matrix B similar to A that is diagonal. Moreover, the entries along the diagonal of B are precisely the eigenvalues of A.*

Proof: The proof relies on showing that the eigenvectors of A form a basis such that A becomes diagonal in this basis. We assume that A's eigenvalues are distinct since technical difficulties arise when eigenvalues have multiplicity.

Let $v_1, v_2, \ldots, v_n$ be the eigenvectors of A with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Form the matrix

$$C = \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix},$$

where the i-th column of C is the column vector v_i. We will show that the matrix $C^{-1}AC$ satisfies our theorem. Specifically, we aim to demonstrate

$$C^{-1}AC = B,$$

where B is the desired diagonal matrix.

Define B as

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

The diagonal matrix B is the only matrix satisfying $Be_i = \lambda_i e_i$ for all i, where e_i is the standard basis vector. Observe that $Ce_i = v_i$ for all i. Thus,

$$C^{-1}ACe_i = C^{-1}Av_i = C^{-1}(\lambda_i v_i) = \lambda_i C^{-1} v_i = \lambda_i e_i,$$

proving that

$$C^{-1}AC = B.$$

This result is not the end of the story. For nonsymmetric matrices, there are other canonical forms to find "good" similar matrices, such as the Jordan canonical form, upper triangular form, and rational canonical form.

Bibliography

Thomas A. Garrity and Lori Pedersen. *All the mathematics you missed: but need to know for graduate school*. Cambridge Univ. Press, Cambridge, 11. print edition, 2007. ISBN 978-0-521-79285-1 978-0-521-79707-8.