Vectors and Matrices

William F. Barnes

April 20, 2022

Contents

Chapter 1

Vectors and Matrices

1 Introduction to Vectors

1.1 Anatomy of a Vector

Definition

A *vector* is an ordered list of numbers or variables. One example of a vector is the sequence of numbers 2, 5, and 7, which can be written in *vector notation*:

$$\vec{V} = \langle 2, 5, 7 \rangle$$

Vector notation requires a *label* for the vector, $\vec{V}$ in our example, specially marked by an arrow ($\vec{\ }$). On the right, the the so-called *vector literal* is enclosed by left-and right-angle brackets $\langle\ \rangle$ as shown. Each number in the vector is separated by a comma.

Components

The individual elements in a vector are formally called *components*, and the total number of components is the *dimension* of the vector. The order in which the components of a vector are listed *does* matter. For example, the three-dimensional vector $\vec{V} = \langle 2, 5, 7 \rangle$ is completely different from its reversed version $\langle 7, 5, 2 \rangle$.

Subscripts

In a vector, any given component is represented using the vector's symbol without the arrow, but including an *index subscript*. For instance, we could represent $\vec{V}$ as

$$\vec{V} = \langle V_a, V_b, V_c \rangle$$

with $V_a = 2$, $V_b = 5$, $V_c = 7$, but the letters a, b, c could easily have been x, y, z, or perhaps 1, 2, 3. Vector component labels are, after the dust settles, purely for bookkeeping.

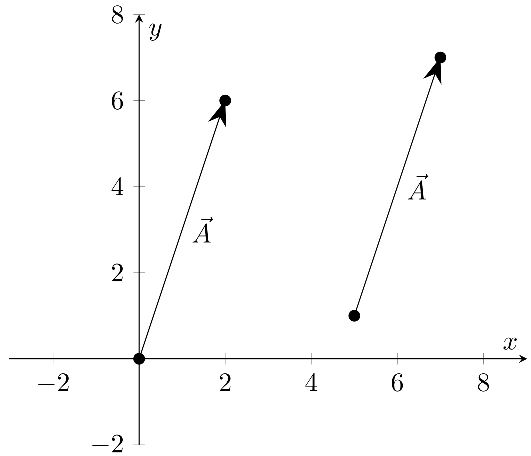

Figure 1.1: Vector $\vec{A} = \langle 2, 6 \rangle$ plotted from two different base points $(0, 0)$ and $(5, 1)$.

1.2 Visual Representation

Vectors of two dimensions are suited for visualization on the Cartesian plane. Given a vector $\vec{V} = \langle V_a, V_b \rangle$, we plot $\vec{V}$ with the following recipe:

- Choose any *base point* for the vector on the plane. From the base point:
- Measure V_a units horizontally, measure V_b units vertically.
- Plot the vector *tip point*. Connect base and tip with an arrow.

Plotted in Fig. 1.1 are two *equivalent* representations of the vector $\vec{A} = \langle 2, 6 \rangle$. Note that the vector doesn't 'care' about the choice of base point.

It should also follow that the above construction extends to dimensions beyond two. For instance, vectors of three dimensions can be visualized in a three-dimensional coordinate system, and so on.

1.3 Position Vector

A vector whose base point is the origin $(0, 0)$ is called a *position vector*, often denoted $\vec{R}$ or $\vec{X}$. A position vector $\vec{R} = \langle R_x, R_y \rangle$ is equivalent to the ordered pair (x, y), denoting a unique point in the Cartesian plane.

1.4 Magnitude and Direction

Given the 'arrow' representation of a vector, we notice two important features:

- Vectors have a *magnitude*, i.e. the total arrow length.

- Vectors have a *direction*, i.e. a notion of pointing somewhere.

The 'information' in a vector is completely represented by its magnitude and its direction. (This may grant some relief as to why we can be so loose about the choice of base point.)

Calculating the Magnitude

A vector $\vec{A}$ of dimension N has a magnitude given by

$$A = \left|\vec{A}\right| = \sqrt{A_1^2 + A_2^2 + A_3^2 + \cdots + A_N^2}\ . \tag{1.1}$$

Intuitively, the magnitude of a vector can be thought of the hypotenuse of an N-dimensional triangle. For the special case $N = 2$, the above reduces to the Pythagorean theorem.

Calculating the Direction

The direction of an N-dimensional vector $\vec{A}$ is always implied by the components A_j, but an explicit formula for the 'angle' of the vector is only trivial for small N. Working the $N = 2$ case, the direction in which a vector $\vec{A} = \langle A_x, A_y\rangle$ is pointing is given by

$$\phi = \arctan\left(\frac{A_y}{A_x}\right) \tag{1.2}$$

To justify (1.2), assume A_x and A_y are two sides of a right triangle such that

$$A_x = A\cos(\phi) \qquad\qquad A_y = A\sin(\phi)\ ,$$

and eliminate the magnitude A.

2 Vector Addition

2.1 Definition

Two vectors $\vec{A}$, $\vec{B}$ of equal dimension N can be added by combining like components, resulting in a vector $\vec{C}$ with N components:

$$\vec{C} = \vec{A} + \vec{B} \tag{1.3}$$

The jth component is given by

$$C_j = A_j + B_j \qquad\qquad j = 1, 2, 3, \ldots, N\ . \tag{1.4}$$

Commutativity of Addition

Following immediately from (1.3)-(1.4) is the *commutativity of addition*:

$$\vec{A} + \vec{B} = \vec{B} + \vec{A} \tag{1.5}$$

In particular, (1.5) tells us that the order in which two vectors are added does not affect the result.

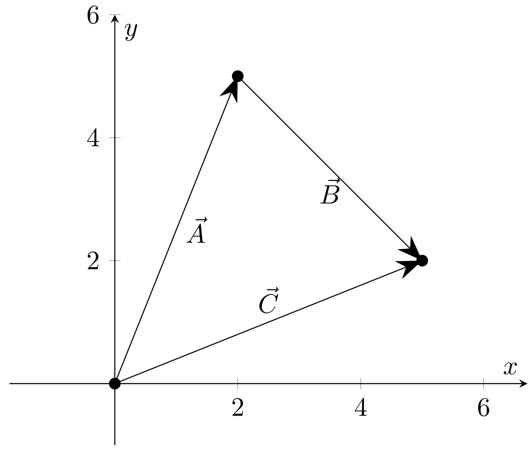

Figure 1.2: Vector addition $\vec{C} = \vec{A} + \vec{B}$.

Associativity of Addition

The sum of three vectors $\vec{A}$, $\vec{B}$, $\vec{C}$ involves two addition operations. Also following from (1.3)-(1.4) is the *associativity of addition*, telling us that the order of the two addition operations does not effect the result:

$$\vec{A} + \left(\vec{B} + \vec{C}\right) = \left(\vec{A} + \vec{B}\right) + \vec{C} \tag{1.6}$$

2.2 Arrow Trick

The 'arrow' representation of a vector avails a beautiful shortcut for vector addition. Given a pair of two-dimensional vectors $\vec{A}$, $\vec{B}$, recall that each vector can be drawn *anywhere* in the Cartesian plane. *By arranging the two vectors in tip-to-tail fashion, the vector sum goes from the tail of the first to the tip of the second.*

Fig. 1.2 demonstrates the 'arrow trick' on two example vectors $\vec{A} = \langle 2, 5 \rangle$, $\vec{B} = \langle 3, -3 \rangle$, whose sum easily comes out to $\vec{C} = \langle 5, 2 \rangle$. By plotting $\vec{A}$, $\vec{B}$ as suggested, the sum $\vec{C}$ is visually represented by an arrow beginning at the base of $\vec{A}$ and ending at the tip of $\vec{B}$.

2.3 Additive Inverse

Given any vector $\vec{A} = \langle A_1, A_2, A_3, \ldots, A_N \rangle$, the *additive inverse* is another vector that reverses the sign on all components in $\vec{A}$, denoted $-\vec{A}$, where

$$-\vec{A} = \langle -A_1, -A_2, -A_3, \ldots, -A_N \rangle \ . \tag{1.7}$$

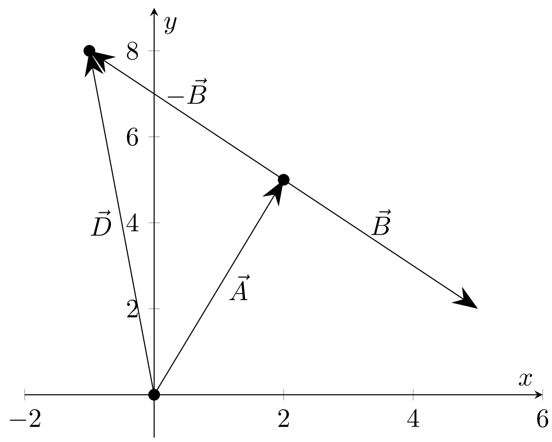

Figure 1.3: Vector subtraction $\vec{D} = \vec{A} - \vec{B}$.

2.4 Zero Vector

The so-called *zero vector* is the vector that contains only zeros:

$$\vec{0} = \langle 0, 0, 0, \ldots, 0 \rangle \tag{1.8}$$

For hopefully obvious reasons, turns out that the sum of any vector and its additive inverse always yields the zero vector:

$$\vec{A} + \left(-\vec{A}\right) = \vec{0}$$

In practice, the zero vector is simply written 0, omitting the arrow.

An interesting corollary to the rules of vector addition is that any *closed* sequence of vectors sums to zero. For example, drawing a triangle without lifting the pen from the surface is represented by $\vec{A} + \vec{B} + \vec{C} = 0$.

2.5 Vector Subtraction

With the additive inverse established, the notion of *vector subtraction* can be framed in terms of vector addition. Given two vectors $\vec{A}$, $\vec{B}$, the difference $\vec{D} = \vec{A} - \vec{B}$ can be visualized with the same 'arrow trick', so long as we reverse the direction on $\vec{B}$ as shown in Fig. 1.3.

3 Scalar Multiplication

A vector $\vec{A}$ can be 'scaled' by a number α called a *scalar*, which has the effect of multiplying the scalar into each component, yielding a new vector $\vec{B}$:

$$\vec{B} = \alpha\vec{A} = \langle \alpha A_1, \alpha A_2, \alpha A_3, \ldots, \alpha A_N \rangle \tag{1.9}$$

3.1 Parallel Vectors

Two vectors whose components are identical up to a scale factor α are said to be *parallel*. Somewhat like parallel lines, two parallel vectors can have different magnitudes, but point in the same direction. The vectors $\vec{A}$, $\vec{B}$ in (1.9) are necessarily parallel.

3.2 Straight Lines

Straight lines in the Cartesian plane are easily represented with vector addition and scalar multiplication. Consider the slope-intercept form of a line, namely $y = mx + b$, where m is the slope and b is the y-intercept at $(0, b)$. As a vector, the y-intercept can be written

$$\vec{b} = \langle 0, b \rangle \ .$$

Required next is a vector $\vec{m}$ that represents the slope of the line, which we capture by writing

$$\vec{m} = \langle m_x, m_y \rangle \qquad\qquad \frac{m_y}{m_x} = m \ .$$

Multiplying $\vec{m}$ by any scalar value α will lengthen, shorten, or reverse its effective placement.

Putting the two ingredients together, it follows that any point on the line $y = mx + b$ is equivalently represented as

$$\vec{r} = \alpha\, \vec{m} + \vec{b} \,, \tag{1.10}$$

where $\vec{r} = \langle x, y \rangle$ is the resulting position vector, as shown in Fig. 1.4. In case (1.10) isn't convincing, one may resolve $\vec{r}$ back into components

$$x = \alpha\, m_x \qquad\qquad y = \alpha\, m_y + b \,,$$

where eliminating α recovers the familiar $y = mx + b$.

Perpendicular Lines

Two lines in the Cartesian plane are perpendicular one line's slope is m, and the slope is $m_\perp = -1/m$. In terms of the components m_x, m_y, this means

$$m = \frac{m_y}{m_x} \qquad\qquad m_\perp = \frac{-m_x}{m_y} \ .$$

From this, the 'perpendicular slope vector' $\vec{m}_\perp$ is evidently $\vec{m}_\perp = \langle -m_y, m_x \rangle$.

3.3 Algebraic Properties

Associativity with Scalars

If a vector is modified by two scalars, the order in which they're applied does not matter:

$$\alpha \left(\beta \vec{A} \right) = (\alpha\beta)\, \vec{A} = (\beta\alpha)\, \vec{A} = \beta \left(\alpha \vec{A} \right) \tag{1.11}$$

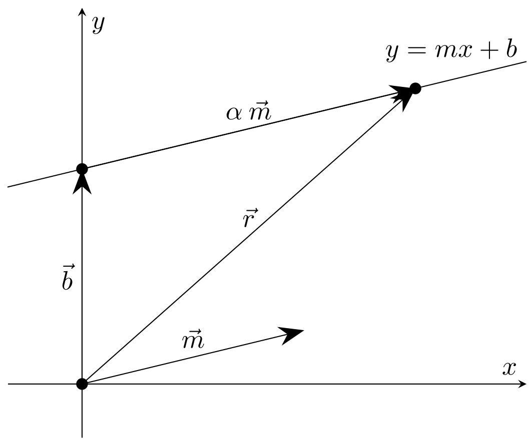

Figure 1.4: Vector construction of a straight line $y = mx + b$.

Distributive Properties

Readily provable from the properties of vector addition and scalar multiplication are the distributive properties involving the sum of two scalars:

$$(\alpha + \beta)\,\vec{A} = \alpha\vec{A} + \beta\vec{B} \tag{1.12}$$

$$\alpha\left(\vec{A} + \vec{B}\right) = \alpha\vec{A} + \alpha\vec{B} \tag{1.13}$$

4 Vector Products

4.1 Dot Product

Two vectors of equal dimension can be 'multiplied' to form a scalar, called the *dot product*, or the *scalar product*. The dot product is an operation that tells us how much of one vector's 'shadow' falls upon another vector, resulting in a scalar called a *projection*. For two vectors $\vec{A}$, $\vec{B}$ of dimension N, the dot product reads

$$\vec{A} \cdot \vec{B} = A_1 B_1 + A_2 B_2 + A_3 B_3 + \cdots + A_N B_N \ ,$$

or, in summation notation:

$$\vec{A} \cdot \vec{B} = \sum_{j=1}^{N} A_j B_j \tag{1.14}$$

Commutativity Relation

Implicit in the definition (1.14) is the commutativity of the dot product (in any number of dimensions):

$$\vec{A} \cdot \vec{B} = \vec{B} \cdot \vec{A} \tag{1.15}$$

Geometric Interpretation

The definition (1.14) becomes more intuitive by studying the $N = 2$ case. Consider two arbitrary vectors given by

$$\vec{A} = \langle A\cos(\phi_A), A\sin(\phi_A)\rangle$$
$$\vec{B} = \langle B\cos(\phi_B), B\sin(\phi_B)\rangle \ ,$$

where A and B are the respecctive magnitudes. Calculating $\vec{A} \cdot \vec{B}$ using the formula provided results in

$$\vec{A} \cdot \vec{B} = AB\left(\cos(\phi_A)\cos(\phi_B) + \sin(\phi_A)\sin(\phi_B)\right)$$
$$\vec{A} \cdot \vec{B} = AB\cos(\phi_B - \phi_A) \ ,$$

telling us that *the dot product is equal to the product of the magnitudes and the cosine of the angle between the vectors.* In general, this result reads

$$\cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{AB} \ , \tag{1.16}$$

where θ is the angle between the vectors in any number of dimensions. The special case $N = 2$ corresponds to $\theta = \phi_B - \phi_A$.

Orthogonality

From the two-dimensional dot product, note that the case $\phi_A - \phi_B = \pm\pi/2$ returns $\cos(\pm\pi/2) = 0$ on the left, telling us that *the dot product between perpendicular vectors is zero.* The formal term for 'perpendicular' is *orthogonal*, and this notion generalizes to N dimensions:

$$\vec{A} \cdot \vec{B} = 0 \qquad\qquad \vec{A} \text{ 'orthogonal to' } \vec{B} \tag{1.17}$$

In the Cartesian plane, recall that the slope of a line and another perpendicular line are represented by the vectors

$$\vec{m} = \langle m_x, m_y\rangle \qquad\qquad \vec{m}_\perp = \langle -m_y, m_x\rangle \ ,$$

respectively. We verify these vectors to be orthogonal by calculating

$$\vec{m} \cdot \vec{m}_\perp = -m_x m_y + m_y m_x = 0 \, .$$

Magnitude

The dot product is responsible for the formula (1.1) for calculating the magnitude of a vector. Indeed, for an N-dimensional vector $\vec{A}$, we find the dot product with itself to be

$$\vec{A} \cdot \vec{A} = A_1^2 + A_2^2 + A_3^2 + \cdots + A_N^2 \ ,$$

which is the square of the magnitude of A. More concisely:

$$A = \left|\vec{A}\right| = \sqrt{\vec{A} \cdot \vec{A}} \tag{1.18}$$

Distributive Property

For three vectors $\vec{A}$, $\vec{B}$, $\vec{C}$ of equal dimension, the dot product obeys the distributive property as one may expect:

$$\vec{A} \cdot \left(\vec{B} + \vec{C}\right) = \vec{A} \cdot \vec{B} + \vec{A} \cdot \vec{C} \tag{1.19}$$

Law of Cosines

An important relation from trigonometry called the *law of cosines* is derived using dot products. Consider the vector sum

$$\vec{A} - \vec{B} = \vec{C} \; ,$$

and then square both sides:

$$\begin{aligned} \left(\vec{A} - \vec{B}\right) \cdot \left(\vec{A} - \vec{B}\right) &= \vec{C} \cdot \vec{C} \\ \vec{A} \cdot \vec{A} + \vec{B} \cdot \vec{B} - 2\vec{A} \cdot \vec{B} &= \vec{C} \cdot \vec{C} \end{aligned}$$

Labeling θ as the angle between vectors $\vec{A}$, $\vec{B}$, the above simplifies to the law of cosines:

$$A^2 + B^2 - 2AB\cos(\theta) = C^2 \tag{1.20}$$

Note that all right triangles have $\theta = \pi/2$, in which case (1.20) reduces to the Pythagorean theorem.

4.2 Cross Product

Two vectors of equal dimension can be 'multiplied' to form a new vector, called the *cross product*, or the *vector product*. The cross product is, for most purposes, a strictly three-dimensional operation. Consider the pair of vectors with $N = 3$:

$$\vec{A} = \langle A_x, A_y, A_z \rangle \qquad\qquad \vec{B} = \langle B_x, B_y, B_z \rangle$$

The cross product $\vec{A} \times \vec{B}$ is given by

$$\vec{A} \times \vec{B} = \langle A_y B_z - A_z B_y, A_z B_x - A_x B_z, A_x B_y - A_y B_x \rangle \; , \tag{1.21}$$

and is *orthogonal* to both $\vec{A}$ and $\vec{B}$.

Determinant Notation

The cross product formula (1.21) is tricky to memorize, and can be more transparently represented as a 'block of numbers' (*not* a matrix), sometimes called *determinant notation*:

$$\vec{A} \times \vec{B} = \begin{vmatrix} (x) & (-y) & (z) \\ A_x & A_y & A_z \\ B_z & B_y & B_z \end{vmatrix}$$

Without sweating the details of determinant notation, you can play a matching game between the determinant representation of $\vec{A} \times \vec{B}$ and the formula (1.21) to remember how it goes.

Orthogonality Check

To ensure that $\vec{A} \times \vec{B}$ is mutually orthogonal to $\vec{A}$, $\vec{B}$, calculate

$$\vec{A} \cdot \left(\vec{A} \times \vec{B}\right) \qquad\qquad \vec{B} \cdot \left(\vec{A} \times \vec{B}\right)$$

to see what comes out. In detail, the former case proceeds as

$$\begin{aligned}
&\vec{A} \cdot \left(\vec{A} \times \vec{B}\right) \\
&\quad = A_x A_y B_z - A_x A_z B_y + A_y A_z B_x - A_y A_x B_z + A_z A_x B_y - A_z A_y B_x \\
&\quad = B_z \left(A_x A_y - A_y A_x\right) - B_y \left(A_x A_z - A_z A_x\right) + B_z \left(A_x A_y - A_y A_x\right) \\
&\quad = 0\,,
\end{aligned}$$

and similar for the latter case. Note too that we could have kept going with the algebra:

$$\vec{A} \cdot \left(\vec{A} \times \vec{B}\right) = \vec{B} \cdot \left(\vec{A} \times \vec{A}\right) = 0$$

Null Case

The cross product of a vector with itself is identically zero:

$$\vec{A} \times \vec{A} = 0 \tag{1.22}$$

Anti-Commutativity Relation

Given the definition (1.21) of the cross product, one sees that swapping $\vec{A}$, $\vec{B}$ puts a minus sign on the result. This is known as the *anti-commutativity* of the cross product:

$$\vec{A} \times \vec{B} = -\vec{B} \times \vec{A} \tag{1.23}$$

Right Hand Rule

There is a trick that allows one to know the direction of $\vec{A} \times \vec{B}$ known as the (oft-dreaded) *right hand rule*. To know the direction of the vector $\vec{A} \times \vec{B}$, the steps are as follows:

1. On your right hand: point your thumb, index finger, and middle finger out in perpendicular directions.
2. Let your index finger be vector $\vec{A}$, let your middle finger be vector $\vec{B}$.
3. Your thumb points along vector $\vec{A} \times \vec{B}$.

Geometric Interpretation

The definition (1.21) becomes more intuitive by studying a special case. Consider the pair of three-dimensional vectors confined to the xy-plane given by

$$\vec{A} = \langle A\cos(\phi_A), A\sin(\phi_A), 0\rangle$$
$$\vec{B} = \langle B\cos(\phi_B), B\sin(\phi_B), 0\rangle \ ,$$

where A and B are the respecctive magnitudes. Calculating $\vec{A}\times\vec{B}$ using the formula provided results in

$$\vec{A}\times\vec{B} = \langle 0, 0, AB(\cos(\phi_A)\sin(\phi_B) - \cos(\phi_B)\sin(\phi_A))\rangle$$
$$\vec{A}\times\vec{B} = \langle 0, 0, AB\sin(\phi_B - \phi_A)\rangle \ ,$$

telling us that *the cross product is equal to the product of the magnitudes and the sine of the angle between the vectors.* In general, this result also tells us

$$\sin(\theta) = \frac{\left|\vec{A}\times\vec{B}\right|}{AB} \ , \tag{1.24}$$

where θ is the angle between the vectors at any relative orientation.

Area of a Parallelogram

The quantity $AB\sin(\theta)$ can be interpreted as the area of a parallelogram having base B and height $h = A\sin\phi$. For the right-angle case $\phi = \pi/2$, the parallelogram becomes a rectangle of area AB. In the language of vectors, the product $\left|\vec{A}\times\vec{B}\right|$ is the area of the parallelogram with sides A, B.

4.3 Identities

Consider three vectors $\vec{A}$, $\vec{B}$, $\vec{C}$, each of three dimensions.

Triple Product

The quantity

$$V = \vec{A}\cdot\left(\vec{B}\times\vec{C}\right) \tag{1.25}$$

is a scalar called the *triple product*. Intuitively, the triple product describes the volume of the parallel-piped with sides A, B, C. One can show by brute force that (1.25) obeys the cyclic relations:

$$\vec{A}\cdot\left(\vec{B}\times\vec{C}\right) = \vec{B}\cdot\left(\vec{C}\times\vec{A}\right) = \vec{C}\cdot\left(\vec{A}\times\vec{B}\right)$$

BAC-CAB Formula

A useful equation known as the *BAC-CAB* identity, reads

$$\vec{A} \times \left(\vec{B} \times \vec{C}\right) = \vec{B}\left(\vec{A} \cdot \vec{C}\right) - \vec{C}\left(\vec{A} \cdot \vec{B}\right) \ . \tag{1.26}$$

The proof of (1.26) is slightly long but straightforward, using (optional) determinant notation to contain the cross product:

$$\vec{A} \times \left(\vec{B} \times \vec{C}\right) = \begin{vmatrix} (x) & (-y) & (z) \\ A_x & A_y & A_z \\ B_y C_z - B_z C_y & B_z C_x - B_x C_z & B_x C_y - B_y C_x \end{vmatrix}$$

$$\begin{aligned} \vec{A} \times \left(\vec{B} \times \vec{C}\right) &= \langle A_y B_x C_y - A_y B_y C_x - A_z B_z C_x + A_z B_x C_z, 0, 0\rangle + \\ &\quad \langle 0, A_z B_y C_z - A_z B_z C_y - A_x B_x C_y + A_x B_y C_x, 0\rangle + \\ &\quad \langle 0, 0, A_x B_z C_x - A_x B_x C_z - A_y B_y C_z + A_y B_z C_y\rangle \\ &= B_x \langle A_y C_y + A_z C_z, 0, 0\rangle - A_x B_x \langle 0, C_y, C_z\rangle + \\ &\quad B_y \langle 0, A_z C_z + A_x C_x, 0\rangle - A_y B_y \langle C_x, 0, C_z\rangle + \\ &\quad B_z \langle 0, 0, A_x C_x + A_y C_y\rangle - A_z B_z \langle C_x, C_y, 0\rangle \end{aligned}$$

$$\begin{aligned} \vec{A} \times \left(\vec{B} \times \vec{C}\right) &= B_x \left\langle \vec{A} \cdot \vec{C}, 0, 0\right\rangle - A_x B_x \langle C_x, C_y, C_z\rangle + \\ &\quad B_y \left\langle 0, \vec{A} \cdot \vec{C}, 0\right\rangle - A_y B_y \langle C_x, C_y, C_z\rangle + \\ &\quad B_z \left\langle 0, 0, \vec{A} \cdot \vec{C}\right\rangle - A_z B_z \langle C_x, C_y, C_z\rangle \end{aligned}$$

$$\vec{A} \times \left(\vec{B} \times \vec{C}\right) = \vec{B}\left(\vec{A} \cdot \vec{C}\right) - \vec{C}\left(\vec{A} \cdot \vec{B}\right)$$

5 Polar Representation

In the Cartesian plane, consider a position vector

$$\vec{r} = \langle r_x, r_y\rangle \ .$$

The 'magnitude-and-direction' interpretation of $\vec{r}$ assigns the magnitude r to the hypotenuse of a right triangle, where the adjacent and opposite sides are respectively given by

$$r_x = r\cos(\phi) \tag{1.27}$$

$$r_y = r\sin(\phi) \ , \tag{1.28}$$

congruent with equations (1.1)-(1.2). The angle parameter ϕ is also known as the *phase* of the vector, a dimensionless argument unique on the interval $[0 : 2\pi)$.

5.1 Polar Coordinate System

Equations (1.27)-(1.28) represent a mapping from system of Cartesian coordinates to the system of *polar coordinates*. Any point in the plane that can be represented by the ordered pair (x, y) has an equivalent representation as the ordered pair (r, ϕ). In particular, we take the position vecor in polar coordinates to be

$$\vec{r} = \langle r\cos(\theta), r\sin(\theta)\rangle = r\langle\cos(\theta), \sin(\theta)\rangle \tag{1.29}$$

5.2 Rotated Vectors

Starting with a vector $\vec{r}$ in two dimensions, particularly

$$\vec{r} = \langle r\cos(\phi), r\sin(\phi)\rangle \ ,$$

we may inquire what happens when we modify the phase such that $\phi \to \phi + \theta$, effectively rotating the vector in the plane.

Carrying this out, one writes

$$(\vec{r})' = r\langle\cos(\phi+\theta), \sin(\phi+\theta)\rangle \ ,$$

or, expanding the trigonometry terms,

$$\begin{aligned} r'_x &= r\left(\cos(\phi)\cos(\theta) - \sin(\phi)\sin(\theta)\right) \\ r'_y &= r\left(\sin(\phi)\cos(\theta) + \cos(\phi)\sin(\theta)\right) \ . \end{aligned}$$

Things get interesting when we keep simplifying:

$$r'_x = r_x\cos(\theta) - r_y\sin(\theta) \tag{1.30}$$

$$r'_y = r_x\sin(\theta) + r_y\cos(\theta) \tag{1.31}$$

Written this way, we see that the 'new' components r'_j are a mixture of the 'old' components r_j scaled by trigonometry terms that depend only on θ.

5.3 Rotation Matrix

Equations (1.30)-(1.31) can be packed into a single statement using *matrix notation*:

$$\begin{bmatrix} r'_1 \\ r'_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{1.32}$$

Explicitly, we have made the associations

$$(\vec{r})' = \begin{bmatrix} r'_x \\ r'_y \end{bmatrix} = \begin{bmatrix} r'_1 \\ r'_2 \end{bmatrix} \qquad \vec{r} = \begin{bmatrix} r_x \\ r_y \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \ ,$$

and the 'block of numbers' containing the trigonometry terms is called the *rotation* matrix, or *rotation operator*, denoted $\tilde{R}$:

$$\tilde{R} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} \\ \tilde{R}_{21} & \tilde{R}_{22} \end{bmatrix} \tag{1.33}$$

Whenever a matrix such as $\tilde{R}$ occurs to the left of a vector such as $\vec{r}$ as in (1.32), there is an implied operation that 'applies' the matrix onto the vector by cross-multiplying certain components. This procedure is captured by writing (1.32) in component form:

$$r'_j = \sum_{k=1}^{2} \tilde{R}_{jk} \, r_k \tag{1.34}$$

6 Basis Vectors

6.1 Unit Vectors

Consider a vector $\vec{V}$ of dimension N, having magnitude V. A special vector $\hat{V}$, called a *unit vector*, is defined as $\vec{V}$ divided by its own magnitude:

$$\hat{V} = \frac{1}{V} \vec{V} \tag{1.35}$$

That is, a unit vector always has magnitude one, and points along the original vector. A vector of the form (1.35) is said to be *normalized*.

A more intuitive way to understand unit vectors is to rearrange (1.35) to write

$$\vec{V} = V \, \hat{V} \, ,$$

which tell is that a full vector $\vec{V}$ is the product of the magnitude V and the 'direction' unit vector $\hat{V}$.

6.2 Introduction to Basis Vectors

Consider an arbitrary vector $\vec{V}$ of dimension N. A curious way to express

$$\vec{V} = \langle V_1, V_2, V_3, \dots, V_N \rangle$$

is to fully pull apart each component so that $\vec{V}$ is the sum of N pure sub-vectors:

$$\vec{V} = \langle V_1, 0, 0, \dots \rangle + \langle 0, V_2, 0, \cdots \rangle + \langle \cdots, 0, V_3, 0, \cdots \rangle + \cdots + \langle \cdots, 0, V_N \rangle$$

Each sub-vector contains just one component V_j, which can be factored out of the sub-vector as a scalar. The sub-vectors that remain are called *basis vectors*, denoted $\hat{e}_j$.

$$\begin{aligned} \hat{e}_1 &= \langle 1, 0, 0, \cdots, 0 \rangle \\ \hat{e}_2 &= \langle 0, 1, 0, \cdots, 0 \rangle \\ \hat{e}_3 &= \langle 0, 0, 1, \cdots, 0 \rangle \\ &\cdots \\ \hat{e}_N &= \langle 0, 0, 0, \cdots, 1 \rangle \end{aligned} \tag{1.36}$$

There is one basis vector $\hat{e}_j$ for each of the N dimensions in which the vector is situated.

Cartesian Coordinates

In the Cartesian xy-plane, a vector is typically represented as $\vec{V} = \langle V_x, V_y \rangle$, suggesting basis vectors

$$\hat{e}_x = \langle 1, 0 \rangle \qquad\qquad \hat{e}_y = \langle 0, 1 \rangle \ .$$

Note that the same notation extrapolates to three dimensions, in which case

$$\hat{e}_x = \langle 1, 0, 0 \rangle \qquad\qquad \hat{e}_y = \langle 0, 1, 0 \rangle \qquad\qquad \hat{e}_z = \langle 0, 0, 1 \rangle$$

are the basis vectors.

Orthogonality

Basis vectors are all mutually orthogonal by necessity. For two different basis vectors $\hat{e}_j$, $\hat{e}_k$, the *orthogonality relation* is

$$\hat{e}_j \cdot \hat{e}_k = 0 \, . \tag{1.37}$$

On the other hand, two of the same basis vector $\hat{e}_k$ obeys

$$\hat{e}_k \cdot \hat{e}_k = 1 \, . \tag{1.38}$$

In the general case, any set of basis vectors $\{\hat{e}_j\}$ that obeys (1.37), (1.38) is said to be *orthonormal*.

6.3 Linear Combinations

Having established the notion of basis vectors, we are free to an express arbitrary vector $\vec{V}$ as a *linear combination* of each $\hat{e}_j$, namely

$$\vec{V} = V_1 \, \hat{e}_1 + V_2 \, \hat{e}_2 + V_3 \, \hat{e}_3 + \cdots + V_N \, \hat{e}_N \ , \tag{1.39}$$

or in summation notation:

$$\vec{V} = \sum_{j=1}^{N} V_j \, \hat{e}_j \tag{1.40}$$

In the above, $\vec{V}$ can potentially point to any 'place' in the N-dimensional space in which it lives. Such a place is formally called a *vector space*.

Isolating Components

One may 'solve' for the V_jth component in a vector $\vec{V}$ by exploiting the orthogonality relations (1.37), (1.38). Start with (1.39), and multiply any particular $\hat{e}_k$ into both sides:

$$\hat{e}_k \cdot \vec{V} = V_1 \, \hat{e}_k \cdot \hat{e}_1 + V_2 \, \hat{e}_k \cdot \hat{e}_2 + V_3 \, \hat{e}_k \cdot \hat{e}_3 + \cdots + V_N \, \hat{e}_k \cdot \hat{e}_N$$

Next, observe that *all except one* of the dot products on the right will cancel due to (1.38). The whole sum collapses to the term with $j = k$, namely $V_k\, \hat{e}_k \cdot \hat{e}_k$, simplifying to V_k. Formally, we have uncovered the obvious yet satisfying statement:

$$V_k = \vec{V} \cdot \hat{e}_k \tag{1.41}$$

With an explicit formula for the V_jth component of a vector, it's curious to see happens by replacing V_j in (1.40). Carrying this out, we can write a component-free way to reference a vector and its contents:

$$\vec{V} = \sum_{j=1}^{N} \left(\vec{V} \cdot \hat{e}_j \right) \hat{e}_j \tag{1.42}$$

Spanning the Vector Space

It's important to notice that a linear combination $\vec{V}$, with appropriate values of V_j, could represent *any* point in the N-dimensional *vector space* in which the vector is embedded. This is possible because the set of basis vectors $\{\hat{e}_j\}$ are said to *span* the vector space.

7 Change of Basis

Consider a two-dimensional vector $\vec{V} = \langle V_x, V_y \rangle$, naturally expressed as a linear combination in the Cartesian basis

$$\hat{e}_1 = \hat{x} = \langle 1, 0 \rangle \qquad\qquad \hat{e}_2 = \hat{y} = \langle 0, 1 \rangle \ .$$

By convention, the Cartesian coordinate system is usually aligned with the edges of a rectangular sheet of paper or computer screen. The orientation of the coordinate system is of course arbitrary, and we must be free to *rotate* the basis vectors without 'physical' consequences.

Figure 1.5 shows a two-dimensional example with two sets of basis vectors $\{\hat{e}_j\}$, $\{\hat{u}_j\}$ embedded on the Cartesian plane. In particular, the basis vector $\hat{u}$ is rotated up from $\hat{x}$ by some arbitrary angle, and similarly $\hat{v}$ corresponds to $\hat{y}$ by the same angle. Any given linear combination $\vec{r}$ has a different representation in each basis.

Generalizing this idea to N diensions, we can say that linear combinations of the form (1.40) can be re-expressed in terms of a different set of orthonormal basis vectors $\{\hat{u}_j\}$:

$$(\vec{V})' = \sum_{j=1}^{N} V'_j\, \hat{u}_j \tag{1.43}$$

Analogous to (1.41), the primed components relate to the vector by:

$$V'_k = (\vec{V})' \cdot \hat{u}_k \tag{1.44}$$

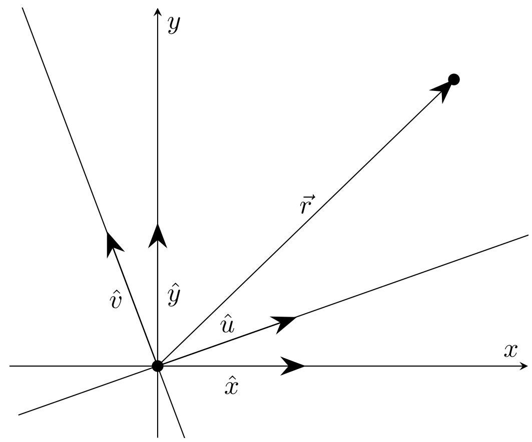

Figure 1.5: Vector $\vec{r}$ as a linear combination in two different bases.

7.1 Two Dimensions

Rotated Cartesian Coordinates

Suppose a different set basis vectors $\hat{u}$, $\hat{v}$ is given in terms of the original $\{\hat{e}_j\}$ basis such that

$$\hat{u}_1 = \hat{u} = \left\langle \frac{\sqrt{3}}{2}, \frac{1}{2} \right\rangle \qquad \hat{u}_2 = \hat{v} = \left\langle \frac{-1}{2}, \frac{\sqrt{3}}{2} \right\rangle$$

or equivalently,

$$\hat{u}_1 = \frac{\sqrt{3}}{2}\,\hat{e}_1 + \frac{1}{2}\,\hat{e}_2 \qquad \hat{u}_2 = -\frac{1}{2}\,\hat{e}_1 + \frac{\sqrt{3}}{2}\,\hat{e}_2 \;.$$

Note that each $\hat{u}_j$ is a linear combination of each $\hat{e}_j$. The coefficients $\sqrt{3}/2$, $1/2$, etc. are carefully chosen to assure orthonormality between $\hat{u}_{1,2}$.

If a vector $\vec{r}$ is expressed in the $\{\hat{e}_j\}$ basis as the linear combination

$$\vec{r} = r_1\,\hat{e}_1 + r_2\,\hat{e}_2 \;,$$

the so-called 'change of basis' occurs if we algebraically replace all $\hat{e}_j$ with $\hat{u}_j$, which first requires inverting the above relations:

$$\hat{e}_1 = \frac{\sqrt{3}}{2}\,\hat{u}_1 - \frac{1}{2}\,\hat{u}_2 \qquad \hat{e}_2 = \frac{1}{2}\,\hat{u}_1 + \frac{\sqrt{3}}{2}\,\hat{u}_2$$

Then, the vector $\vec{r}$ can be written

$$(\vec{r})' = r_1\left(\frac{\sqrt{3}}{2}\,\hat{u}_1 - \frac{1}{2}\,\hat{u}_2\right) + r_2\,\left(\frac{1}{2}\,\hat{u}_1 + \frac{\sqrt{3}}{2}\,\hat{u}_2\right)$$

$$(\vec{r})' = \left(r_1\frac{\sqrt{3}}{2} + r_2\frac{1}{2}\right)\hat{u}_1 + \left(-r_1\frac{1}{2} + r_2\frac{\sqrt{3}}{2}\right)\hat{u}_2 \;,$$

where the components $r_{1,2}$ are finally readable as

$$r_1' = r_1 \frac{\sqrt{3}}{2} + r_2 \frac{1}{2} \qquad\qquad r_2' = -r_1 \frac{1}{2} + r_2 \frac{\sqrt{3}}{2} \ ,$$

and a form like (1.43) is attained:

$$(\vec{r})' = r_1' \, \hat{u}_1 + r_2' \, \hat{u}_2$$

General Coordinate Rotations

The above example can be easily generalized such that $\hat{u}$ points anywhere in the Cartesian plane, with $\hat{v}$ appropriately perpendicular to $\hat{u}$. To achieve this, we introduce an arbitrary parameter θ such that

$$\begin{aligned}\hat{u}_1 &= \cos(\theta) \ \hat{e}_1 + \sin(\theta) \ \hat{e}_2 \\ \hat{u}_2 &= -\sin(\theta) \ \hat{e}_1 + \cos(\theta) \ \hat{e}_2 \ .\end{aligned}$$

By straightforward algebra, we find the inverted version to be

$$\begin{aligned}\hat{e}_1 &= \cos(\theta) \ \hat{u}_1 - \sin(\theta) \ \hat{u}_2 \\ \hat{e}_2 &= \sin(\theta) \ \hat{u}_1 + \cos(\theta) \ \hat{u}_2 \ .\end{aligned}$$

The pairs of equations above are suggestive of a matrix formulation, particularly

$$\begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \end{bmatrix} \ , \tag{1.45}$$

and

$$\begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} \ , \tag{1.46}$$

respectively. Comparing the above to (1.33), we see (1.46) contains the same rotation matrix $\tilde{R}$ that rotates vectors in a fixed basis. Denoting the other matrix in (1.45) as R, the component version of the above reads:

$$\hat{u}_j = \sum_{k=1}^{2} R_{jk} \, \hat{e}_k \qquad\qquad \hat{e}_j = \sum_{k=1}^{2} \tilde{R}_{jk} \, \hat{u}_k \ .$$

With a convenient representation for the basis vectors $\{\hat{e}_j\}$, an arbitrary linear combination

$$\vec{r} = r_1 \, \hat{e}_1 + r_2 \, \hat{e}_2$$

becomes

$$(\vec{r})' = r_1 \sum_{k=1}^{2} \tilde{R}_{1k} \, \hat{u}_k + r_2 \sum_{k=1}^{2} \tilde{R}_{2k} \, \hat{u}_k = \sum_{k=1}^{2} \left(\sum_{j=1}^{2} \tilde{R}_{jk} \, r_j \right) \hat{u}_k \ ,$$

telling us that the kth component of the vector $(\vec{r})'$ is given by

$$r'_k = \sum_{j=1}^{2} \tilde{R}_{jk}\, r_j \ , \tag{1.47}$$

which is a cousin to equation (1.34). To do a fair comparison, let us swap the j-index and the k-index in (1.47) to write

$$r'_j = \sum_{k=1}^{2} \tilde{R}_{kj}\, r_k \ .$$

Looking carefully, the above differs from (1.34) by the order of the subscripts on the $\tilde{R}$-term, ultimately equivalent to reversing the sign on θ. Said another way, a 'positive' rotation in the basis vectors with $\vec{r}$ fixed is equivalent to a 'negative' rotation of $\vec{r}$ with the basis fixed.

7.2 N Dimensions

Change of Basis Vectors

At the center of the change-of-basis problem is the issue of relating the two orthonormal bases $\hat{e}_j$, $\hat{u}_j$ to one another. In N dimensions, the basis vectors are related by linear combinations

$$\hat{u}_j = \sum_{k=1}^{N} U_{jk}\, \hat{e}_k \tag{1.48}$$

$$\hat{e}_j = \sum_{k=1}^{N} \tilde{U}_{jk}\, \hat{u}_k \ . \tag{1.49}$$

Having two subscripts, the terms U_{jk} and $\tilde{U}_{jk}$ are not vector components, but instead *matrix* components. These typically end up being coefficients like $\sqrt{3}/2$, $1/2$, and so on.

To isolate the matrix components U_{jk} and $\tilde{U}_{jk}$, multiply (via dot product) the basis vectors $\hat{e}_m$, $\hat{u}_m$, respectively into (1.48), (1.49):

$$\hat{e}_m \cdot \hat{u}_j = \sum_{k=1}^{N} U_{jk}\, \hat{e}_m \cdot \hat{e}_k$$

$$\hat{u}_m \cdot \hat{e}_j = \sum_{k=1}^{N} \tilde{U}_{jk}\, \hat{u}_m \cdot \hat{u}_k \ .$$

Due to orthonormality, the right side of each equation resolves to zero except for the case with $m = k$, allowing the components to be isolated:

$$U_{jm} = \hat{e}_m \cdot \hat{u}_j \tag{1.50}$$

$$\tilde{U}_{jm} = \hat{u}_m \cdot \hat{e}_j \tag{1.51}$$

From the pair (1.50)-(1.51), we can further relate the matrix components:

$$\tilde{U}_{jm} = U_{mj} \tag{1.52}$$

Linear Combinations

An arbitrary linear combination

$$\vec{V} = \sum_{j=1}^{N} V_j \, \hat{e}_j$$

can be written with all $\hat{e}_j$ replaced according to (1.49):

$$(\vec{V})' = \sum_{j=1}^{N}\sum_{k=1}^{N} \tilde{U}_{jk} \, V_j \, \hat{u}_k = \sum_{k=1}^{N} \left(\sum_{j=1}^{N} \tilde{U}_{jk} \, V_j \right) \hat{u}_k$$

Comparing the above to (1.43) gives a formula for the kth component of the vector $(\vec{V})'$:

$$(V')_k = \sum_{j=1}^{N} \tilde{U}_{jk} \, V_j \tag{1.53}$$

Unity Condition

We can learn a bit more about the matrix components U_{jk}, $\tilde{U}_{jk}$ by eliminating $\hat{u}_k$ between (1.48)-(1.49):

$$\hat{e}_j = \sum_{k=1}^{N} \tilde{U}_{jk} \left(\sum_{m=1}^{N} U_{km} \, \hat{e}_m \right) = \sum_{k=1}^{N}\sum_{m=1}^{N} \left(\tilde{U}_{jk} U_{km} \right) \hat{e}_m$$

Next, use the symbol I to represent the quantity

$$I_{jm} = \sum_{k=1}^{N} \left(\tilde{U}_{jk} U_{km} \right) ,$$

and the above becomes

$$\hat{e}_j = \sum_{m=1}^{N} I_{jm} \, \hat{e}_m \, .$$

For the above to make sense, all terms in the right-hand sum must vanish except for that with $m = j$. Explicitly, this means I_{jm} obeys

$$I_{jm} = \begin{cases} 1 & m = j \\ 0 & m \neq j \end{cases} ,$$

reminiscent of (1.37)-(1.38).

8 Matrix Formalism

Formally, a *matrix* is a collection of numbers or variables arranged in a block with fixed rows M and columns N. Each element, i.e. *component* in the matrix requires two subscripts.

8.1 Matrix-Operator Equivalence

A primary use for a matrix is to 'operate' on a vector of dimension N, yielding a new vector of dimension M. (The term 'matrix' is often interchanged with the term 'operator'.) Symbolically, this is written

$$A\vec{x} = \vec{y} \, ,$$

and in full *block notation*, the same statement looks like

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1N} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{2N} \\ A_{31} & A_{32} & A_{33} & \cdots & A_{3N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{M1} & A_{M2} & A_{M3} & \cdots & A_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_M \end{bmatrix}$$

More transparently, we use *index notation* to express the same caluation:

$$\sum_{k=1}^{N} A_{jk} \, x_k = y_j \qquad\qquad j = 1, 2, 3, \ldots, M \tag{1.54}$$

8.2 Matrix Components

Consider two vectors $\vec{x}$, $\vec{y}$, each a linear combination in some N-dimensional basis such that

$$\vec{x} = \sum_{j=1}^{N} x_j \, \hat{e}_j \qquad\qquad \vec{y} = \sum_{j=1}^{M} y_j \, \hat{e}_j \, .$$

While $\vec{y}$ is perfectly happy being expressed as a linear combination in the basis $\{\hat{e}_j\}$, it's instructive to re-express $\vec{y}$ in terms of its brother, $\vec{x}$. To do so, we propose an operator A such that

$$\vec{y} = A\vec{x} \, .$$

To proceed, write the above as

$$\sum_{k=1}^{M} y_k \, \hat{e}_k = \sum_{k=1}^{N} x_k \, A\hat{e}_k \, ,$$

and multiply the basis vector $\hat{e}_j$ (via dot product) into both sides:

$$\sum_{k=1}^{M} y_k \, \hat{e}_j \cdot \hat{e}_k = \sum_{k=1}^{N} x_k \, \hat{e}_j \cdot A\hat{e}_k$$

On the left, every term in the sum vanishes except that with $j = k$, and the above becomes

$$y_j = \sum_{k=1}^{N} \left(\hat{e}_j \cdot A\hat{e}_k\right) x_k \qquad\qquad j = 1, 2, 3, \dots, M \; .$$

The parenthesized quantity is what we're after:

$$A_{jk} = \hat{e}_j \cdot A\hat{e}_k \tag{1.55}$$

The term A_{jk} is the component of the matrix A corresponding to the jth row, kth column.

8.3 Projector

Consider the curious quantity

$$P_x = \vec{x}\,\vec{x} \; , \tag{1.56}$$

called the the *projector* of $\vec{x}$. By itself, P_x does nothing - there is no operation between the two copies of $\vec{x}$. What the projector *does* is 'wait' to be multiplied into another vector, resulting in a scaled version of $\vec{x}$. For example, applying the projector to a different vector $\vec{y}$ (of the same dimension as $\vec{x}$) goes like

$$P_x\vec{y} = \vec{x}\left(\vec{x} \cdot \vec{y}\right) \; .$$

8.4 Identity Operator

Consider a vector $\vec{x}$ as a linear combination in some N-dimensional basis:

$$\vec{x} = \sum_{j=1}^{N} x_j \, \hat{e}_j$$

For any one of the basis vectors $\hat{e}_k$, write the projector

$$P_{e_k} = \hat{e}_k \, \hat{e}_k \; ,$$

and then multiply $\vec{x}$ onto the right side to get

$$P_{e_k}\vec{x} = \hat{e}_k \, \hat{e}_k \cdot \vec{x} = x_k \, \hat{e}_k$$

By summing over the index k, the right side is identically $\vec{x}$:

$$\left(\sum_{k=1}^{N} P_{e_k}\right)\vec{x} = \sum_{k=1}^{N} x_k \, \hat{e}_k = \vec{x}$$

For the left side to also equal $\vec{x}$, the parenthesized quantity must be equivalent to 'multiplying by one', which we call the *identity* operator:

$$I = \sum_{k=1}^{N} P_{e_k} \tag{1.57}$$

The identity operator leaves a vector unchanged:

$$I\vec{x} = \vec{x}$$

The matrix-equivalence of I is square, has no mixed components, and has ones along the diagonal:

$$I = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{1.58}$$

8.5 Unified Notation

Recall that a matrix A relates to its components A_{jk} in a way given by (1.55), namely

$$A_{jk} = \hat{e}_j \cdot A\hat{e}_k \; .$$

To establish this directly using projectors, start with $A = IAI$ and watch what happens:

$$A = IAI = \sum_{j=1}^{M} \sum_{k=1}^{N} P_{e_j} A P_{e_k} = \sum_{j=1}^{M} \sum_{k=1}^{N} \hat{e}_j \left(\hat{e}_j \cdot A\hat{e}_k \right) \hat{e}_k$$

The parenthesized quantity is precisely A_{jk}. Evidently, the symbolic notation unifies with the index notation in the equation

$$A = \sum_{j=1}^{M} \sum_{k=1}^{N} \hat{e}_j \left(A_{jk} \right) \hat{e}_k \tag{1.59}$$

The presence of $\hat{e}_j \; \hat{e}_k$ is like a projector - it couples the component to the operator.

9 Matrix Operations

9.1 Matrix Addition

Two matrices A and B of identical dimensions, meaning M rows, N columns, can be combined to form a new matrix C such that

$$A + B = C \; ,$$

or:

$$A_{jk} + B_{jk} = C_{jk} \qquad \begin{cases} j = 1, 2, 3, \ldots, M \\ k = 1, 2, 3, \ldots, N \end{cases} \tag{1.60}$$

9.2 Scalar Multiplication

A scalar α can be multiplied into each component of a matrix A to form a new matrix B such that

$$\alpha A = B \; ,$$

or

$$\alpha A_{jk} = B_{jk} \qquad \begin{cases} j = 1, 2, 3, \ldots, M \\ k = 1, 2, 3, \ldots, N \end{cases} \tag{1.61}$$

9.3 Matrix Multiplication

Two matrices A, B, of equal or different dimensions can be multiplied to form a new matrix C:

$$AB = C$$

The main 'rule' is that the number of *columns* in A must equal the number of *rows* in B:

$$A_{(M,K)} \times B_{(K,N)} = C_{(M,N)}$$

Non-Commutativity

If you're paying attention, the commutated product BA may violate the above, and no product is defined. In any case, we should assume that the multiplication of two matrices is not commutative:

$$AB \neq BA \tag{1.62}$$

Multiplication Formula

To derive the formula for matrix multiplication, begin with the following 'unified' representation (1.59) of the respective matrices:

$$A = \sum_{m=1}^{M} \sum_{k=1}^{K} \hat{e}_m \left(A_{mk}\right) \hat{e}_k \qquad B = \sum_{k'=1}^{K} \sum_{n=1}^{N} \hat{e}_{k'} \left(B_{k'n}\right) \hat{e}_n$$

Then, the product AB reads

$$\begin{aligned} AB &= \sum_{m=1}^{M} \sum_{k=1}^{K} \hat{e}_m \left(A_{mk}\right) \hat{e}_k \sum_{k'=1}^{K} \sum_{n=1}^{N} \hat{e}_{k'} \left(B_{k'n}\right) \hat{e}_n \\ &= \sum_{m=1}^{M} \sum_{k=1}^{K} \sum_{k'=1}^{K} \sum_{n=1}^{N} \hat{e}_m \left(A_{mk}\right) \hat{e}_k \, \hat{e}_{k'} \left(B_{k'n}\right) \hat{e}_n \end{aligned}$$

Note that the quantity $\hat{e}_m \, \hat{e}_k \, \hat{e}_{k'} \, \hat{e}_n$ is the juxtaposition of two projectors, readily translating to $\hat{e}_m \, (\hat{e}_k \cdot \hat{e}_{k'}) \, \hat{e}_n$. Note further that the parenthesized product obeys (1.37)-(1.38), namely

$$\hat{e}_k \cdot \hat{e}_{k'} = \begin{cases} 1 & k = k' \\ 0 & k \neq k' \end{cases} ,$$

which has the effect of equating $k = k'$ in the above, eliminating one of the sums. So far then, we have

$$\begin{aligned} AB = C &= \sum_{m=1}^{M} \sum_{k=1}^{K} \sum_{n=1}^{N} (A_{mk} B_{kn}) \, \hat{e}_m \, \hat{e}_n \\ C &= \sum_{m=1}^{M} \sum_{n=1}^{N} \left(\sum_{k=1}^{K} A_{mk} B_{kn} \right) \hat{e}_m \, \hat{e}_n \, . \end{aligned}$$

The symbol C has replaced the quantity AB on the left. Comparing the right side to (1.59), we conclude that the component C_{mn} of matrix C is given by the famed *matrix multiplication* formula:

$$C_{mn} = \sum_{k=1}^{K} A_{mk} B_{kn} \qquad \begin{cases} m = 1, 2, 3, \dots, M \\ n = 1, 2, 3, \dots, N \end{cases} \tag{1.63}$$

Equation (1.63) reminds that it's only required that the number of columns in A match the number of rows in B. For instance, the operation $A_{(2,4)} \times B_{(4,3)} = C_{(2,3)}$, explicitly written as

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \\ b_{41} & b_{42} & b_{43} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix}$$

is perfectly valid, whereas the commuted product $B_{(4,3)} \times A_{(2,4)}$ is undefined.

Associativity

A direct consequence of matrix multiplication is the associativity rule:

$$(AB) \, C = A \, (BC) \tag{1.64}$$

9.4 Change of Basis

A square matrix A with components A_{jk} in the basis $\{\hat{e}_j\}$ can be represented by (1.59):

$$A = \sum_{j=1}^{N} \sum_{k=1}^{N} \hat{e}_j \, (A_{jk}) \, \hat{e}_k \, .$$

Under a change of basis $\{\hat{e}_j\} \to \{\hat{u}_j\}$, we can use (1.49)

$$\hat{e}_j = \sum_{k=1}^{N} \tilde{U}_{jk}\, \hat{u}_k$$

to replace the unit vectors, leading to

$$A' = \sum_{m=1}^{N} \sum_{n=1}^{N} \hat{u}_m \left(\sum_{j=1}^{M} \sum_{k=1}^{N} U_{mj} A_{jk} \tilde{U}_{kn} \right) \hat{u}_n \,,$$

where the (first) term $\tilde{U}_{jm}$ has been replaced by U_{mj} due to (1.52). The parenthesized quantity is precisely the formula for the component A'_{mn} of the transformed matrix

$$A'_{mn} = \sum_{j=1}^{N} \sum_{k=1}^{N} U_{mj} A_{jk} \tilde{U}_{kn} \,, \tag{1.65}$$

or in symbolic form,

$$A' = U A \tilde{U} \,.$$

Note that the above verifies the associativity rule (1.64) for matrix multiplication. The order in which the sums are taken directly corresponds to which matrices are multiplied first. As a bonus, (1.65) tells us exactly how to take the product of three square matrices.

10 Linear Systems

10.1 Order-Two Formalism

Motivation

Consider the linear system of two equations containing two unknowns x and y,

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \,, \end{aligned}$$

where all coefficients are assumed nonzero. One way to solve the system begins by eliminating y, which means to multiply the top and bottom equations by d, b, respectively, and add the results:

$$x\,(ad - bc) = de - bf$$

Similarly, we can eliminate x to end up with

$$y\,(ad - bc) = af - ec \,,$$

and it is now trivial to solve for x and y.

If it just so happens that $ad - bc = 0$, the equations for x and y become indeterminate, meanwhile implying $de = bf$ and $af = ec$. To visualize this, treat each equation as a separate line

$$y_1 = -\frac{a}{b}x + \frac{e}{b} \qquad\qquad y_2 = -\frac{c}{d}x + \frac{f}{d} \, ,$$

having respective slopes $m_1 = -a/b$, $m_2 = -c/d$. Take the difference of slopes to find

$$m_1 - m_2 = -\frac{a}{b} + \frac{c}{d} = \frac{1}{bd}(-ad + bc) = 0 \, ,$$

implying the lines are parallel. Moreover, $de = bf$ causes the lines to have the same y-intercept, thus the two lines y_1, y_2 are *identical*. This reduces the number of equations in the system back down to one equation that has an infinite number of solutions on the line $y_{1,2}$. In the special case that $e = 0$ or $f = 0$, the lines $y_{1,2}$ are parallel but non-overlapping, in which case no solutions exist at all.

Only when $ad - bc$ is *nonzero* does the line y_1 cross y_2 somewhere in the Cartesian plane at one point (x_0, y_0). The intersection point is the 'solution' to the system of equations.

Matrix Formulation

Start with the same two-dimensional system and relabel all coefficients such that

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 &= b_1 \\ A_{21}x_1 + A_{22}x_2 &= b_2 \, , \end{aligned} \tag{1.66}$$

admitting a clean matrix representation

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

where

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \, .$$

Determinant

The classification of solutions $\vec{x}$ depends on the quantity $A_{11}A_{22} - A_{12}A_{21}$, called the *determinant* of the matrix A:

$$\det A = \det \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = A_{11}A_{22} - A_{12}A_{21} \tag{1.67}$$

If $\det A$ is nonzero, the vector $\vec{x}$ solves the system. On the other hand, any matrix with $\det A = 0$ is called *singular*, having a non-obvious number of solutions (zero or infinite, depending on $\vec{b}$). For the non-singular case $\det A \neq 0$, the solution to the system is given by:

$$x_1 = \frac{1}{\det A}\left(A_{22}b_1 - A_{12}b_2\right)$$
$$x_2 = \frac{1}{\det A}\left(A_{11}b_2 - A_{21}b_1\right)$$

Cramer's Rule

In the above solutions for x_1, x_2, observe that the quantities

$$A_{22}b_1 - A_{12}b_2 \qquad\qquad A_{11}b_2 - A_{21}b_1$$

are themselves the determinants of new matrices C_1, C_2 such that

$$x_1 = \frac{1}{\det A}\det\begin{bmatrix} b_1 & A_{12} \\ b_2 & A_{22}\end{bmatrix} = \frac{\det C_1}{\det A}$$
$$x_2 = \frac{1}{\det A}\det\begin{bmatrix} A_{11} & b_1 \\ A_{21} & b_2\end{bmatrix} = \frac{\det C_2}{\det A}$$

That is, the solution to the two-dimensional linear system (1.66) with nonzero determinant is

$$x_j = \frac{\det C_j}{\det A} \qquad\qquad j = 1, 2\,, \tag{1.68}$$

with

$$C_1 = \begin{bmatrix} b_1 & A_{12} \\ b_2 & A_{22}\end{bmatrix} \qquad\qquad C_2 = \begin{bmatrix} A_{11} & b_1 \\ A_{21} & b_2\end{bmatrix}\,. \tag{1.69}$$

The 'recipe' that got us to this point is called *Cramer's Rule*: if $\det A \neq 0$, there's a solution to the system.

10.2 Order-N Formalism

Generalizing the 2×2 linear system to have M equations and N unknowns, we begin with

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + \cdots + A_{1N}x_N &= b_1 \\ A_{21}x_1 + A_{22}x_2 + A_{23}x_3 + \cdots + A_{2N}x_N &= b_2 \\ A_{31}x_1 + A_{32}x_2 + A_{33}x_3 + \cdots + A_{3N}x_N &= b_3 \\ &\cdots \\ A_{M1}x_1 + A_{M2}x_2 + A_{M3}x_3 + \cdots + A_{MN}x_N &= b_M\,, \end{aligned} \tag{1.70}$$

admitting the matrix representation $A\vec{x} = \vec{b}$:

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1N} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{2N} \\ A_{31} & A_{32} & A_{33} & \cdots & A_{3N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{M1} & A_{M2} & A_{M3} & \cdots & A_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_N \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \cdots \\ b_M \end{bmatrix} \tag{1.71}$$

At this stage, the relationship between M and N indicates the quality of solutions (if any) to the system. If $N > M$, the system is said to be *under-determined*, and there is not enough information to choose a solution. On the other hand, if $M > N$, the system is *over-determined*, and there may be zero, or perhaps infinite solutions.

In order to proceed, the matrix A is taken to be *square* with $M = N$. Then, the recipe for solving a two-dimensional linear system extrapolates to N dimensions. Although its not (yet) straightforward how to calculate the determinant of A, we can still use Cramer's rule to write down the components of the solution vector $\vec{x}$, namely

$$x_j = \frac{\det C_j}{\det A} \qquad\qquad j = 1, 2, 3, \ldots, N \tag{1.72}$$

The matrix C_j is constructed by starting with A and replacing the jth column with $\vec{b}$. That is:

$$C_j = \begin{bmatrix} A_{11} & A_{12} & \cdots & b_{1j} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & b_{2j} & \cdots & A_{2N} \\ A_{31} & A_{32} & \cdots & b_{3j} & \cdots & A_{3N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{N1} & A_{N2} & \cdots & b_{Nj} & \cdots & A_{NN} \end{bmatrix} \tag{1.73}$$

10.3 Row Operations

An important tool set called *row operations* can be applied to matrices. To illustrate these, consider again the N-dimensional linear system (1.70). The 'game' we play is to find ways to manipulate the system of equations without losing any information. Without much trouble, one finds the allowed operations to be:

- Exchange two rows.
- Multiply a row by a (nonzero) scalar.
- Replace a row by the sum of itself and another row.

For the sake of assigning symbols to the above row operations, let us denote row exchanges as E, scalar multiplication as M, and a row replacement as R. Using a four-dimensional square matrix as an example, row operations explicitly look like:

$$E^{23}A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix}$$

$$M^2_\alpha A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ \alpha A_{31} & \alpha A_{32} & \alpha A_{33} & \alpha A_{34} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix}$$

$$R_3^2 A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{31}+A_{21} & A_{32}+A_{22} & A_{33}+A_{23} & A_{34}+A_{24} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix}$$

Note that the subscripts and superscripts on the symbols E, M, R are mere convenience of notation, often-omitted.

11 Determinants

The *determinant* is a scalar calculated from the components of a square $(N \times N)$ matrix A. Of the many things the determinant can tell us, we've already seen that $\det A$ indicates the 'quality' of solutions to a linear system. Namely, if the determinant is nonzero, the linear system has a solution given by Cramer's rule. The determinant of a two-dimensional square matrix is given by (1.67).

11.1 Three Dimensions

Consider the three-dimensional linear system

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + A_{13}x_3 &= b_1 \\ A_{21}x_1 + A_{22}x_2 + A_{23}x_3 &= b_2 \\ A_{31}x_1 + A_{32}x_2 + A_{33}x_3 &= b_3 \; , \end{aligned} \tag{1.74}$$

and we're interested in the determinant of the matrix

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} .$$

Labeling each row R_1, R_2, R_3, respectively, we deploy row operations to (i) multiply R_2 by a factor of A_{11}/A_{21}, and then (ii) replace R_2 with $R_2 - R_1$:

$$A \to \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & \frac{A_{11}A_{22}}{A_{21}} - A_{12} & \frac{A_{11}A_{23}}{A_{21}} - A_{13} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

Next, (iii) multiply R_3 by a factor of A_{11}/A_{31}, and (iv) replace R_3 with $R_3 - R_1$:

$$A \to \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & \frac{A_{11}A_{22}}{A_{21}} - A_{12} & \frac{A_{11}A_{23}}{A_{21}} - A_{13} \\ 0 & \frac{A_{11}A_{32}}{A_{31}} - A_{12} & \frac{A_{11}A_{33}}{A_{31}} - A_{13} \end{bmatrix}$$

With the matrix configured as such, observe that the 'important' information is crammed into the lower 2×2 portion of the transformed matrix. Accordingly, we deploy the determinant formula (1.67) to write

$$\det A \propto \left(\frac{A_{11}A_{22}}{A_{21}} - A_{12}\right)\left(\frac{A_{11}A_{33}}{A_{31}} - A_{13}\right) - \left(\frac{A_{11}A_{23}}{A_{21}} - A_{13}\right)\left(\frac{A_{11}A_{32}}{A_{31}} - A_{12}\right) ,$$

which after simplifying, becomes

$$\det A \left(\frac{A_{21}A_{31}}{A_{11}}\right) \propto A_{11}\left(A_{22}A_{33} - A_{32}A_{23}\right) - A_{12}\left(A_{21}A_{33} - A_{31}A_{23}\right) + A_{13}\left(A_{21}A_{32} - A_{31}A_{22}\right) .$$

Keeping in mind that the determinant of A is a single number that characterizes the solutions to the system, it follows that the right-side quantity in the above contains all of the required information. It's also easy (enough) to see that exchanging two rows in the original A will lead to the same final form of $\det A$, up to numerical factors and/or negative signs. In conclusion, we take the the order-three determinant to be

$$\det A = A_{11}\left(A_{22}A_{33} - A_{32}A_{23}\right) - A_{12}\left(A_{21}A_{23} - A_{31}A_{33}\right) + A_{13}\left(A_{21}A_{32} - A_{31}A_{22}\right) . \tag{1.75}$$

11.2 Four Dimensions

Having witnessed the trick of performing row operations on an order-three square matrix A to condense all relevant information into a 2×2 square sub-matrix, this should also work for higher-order matrices. Indeed, the order-four matrix

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix} ,$$

permits a similar process to reduce the order of the problem. Doing so, the order-four determinant becomes the sum of four terms:

$$\det A = A_{11} \det \begin{bmatrix} A_{22} & A_{23} & A_{24} \\ A_{32} & A_{33} & A_{34} \\ A_{42} & A_{43} & A_{44} \end{bmatrix} - A_{12} \det \begin{bmatrix} A_{21} & A_{23} & A_{24} \\ A_{31} & A_{33} & A_{34} \\ A_{41} & A_{43} & A_{44} \end{bmatrix} + A_{13} \det \begin{bmatrix} A_{21} & A_{22} & A_{24} \\ A_{31} & A_{32} & A_{34} \\ A_{41} & A_{42} & A_{44} \end{bmatrix} - A_{14} \det \begin{bmatrix} A_{12} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \\ A_{41} & A_{42} & A_{43} \end{bmatrix} \tag{1.76}$$

11.3 Sub-Matrix and Matrix Minor

Taking another look at the three- and four-dimensional determinant formulas, we see the right side contains the sum of several 'cross sections' of the original matrix, each having dimension $N - 1$.

Sub-Matrix

The *sub-matrix* S_{jk} removes the jth row and the kth column from the original matrix A.

Matrix Minor

The *matrix minor*, denoted M_{jk} is the determinant of the sub-matrix S_{jk}. With matrix minor notation, equations (1.75), (1.76) can be written:

$$\det A_{(3)} = A_{11}M_{11} - A_{12}M_{12} + A_{13}M_{13}$$
$$\det A_{(4)} = A_{11}M_{11} - A_{12}M_{12} + A_{13}M_{13} - A_{14}M_{14}$$

11.4 N Dimensions

Using matrix minor notation, the three and four-dimensional determinant formulas are suggest of how to handle the N-dimensional case. Formally, the procedure is to use row operations to condense down-andd-right all of the information on the matrix. After the dust settles, the general formula for the determinant of an order-N matrix is remarkably simple:

$$\begin{aligned} &\text{if } N = 1: \quad \det A = \det \left[A_{11}\right] \\ &\text{if } N > 1: \quad \det A = \sum_{\substack{j \leq N \\ k=1}}^{N} (-1)^{j+k} A_{jk}M_{jk} \end{aligned} \tag{1.77}$$

Note too that the summation can take place over rows *or* columns, meaning

$$\text{if } N > 1: \quad \det A = \sum_{\substack{j \leq N \\ k=1}}^{N} (-1)^{j+k} A_{kj}M_{kj}$$

also holds. Note that the variable j is fixed at some integer less than N. Only the k-variable is summed over.

11.5 Properties

Multiplication Rules

Readily shown from the properties of determinants are various multiplication rules. A scalar α multiplied by a matrix A of dimension N has the result

$$\det(\alpha A) = \alpha^N \det A \,.$$

Meanwhile, for the product of two matrices A, B:

$$\det(AB) = \det(A)\,\det(B)$$

Row Operations

The row operations E (row exchange), M (multiply by scalar), R (combine rows) have the following effect on the determinant:

$$\begin{aligned}\det(EA) &= -\det A\\ \det(MA) &= \alpha\det A\\ \det(RA) &= \det A\end{aligned}$$

12 Inverse Matrix

12.1 Definition

Given a square matrix A of dimension N, there *may* exist a special matrix A^{-1} that obeys the property

$$A^{-1}A = AA^{-1} = I\ , \tag{1.78}$$

where A^{-1} is called the *inverse matrix* to A, and I is the identity matrix of dimension N. The product of A and its inverse, or vise versa, results in the identity matrix.

Existence

For the notion of the inverse to make sense, the matrix A must perform a one-to-one mapping of a vector $\vec{x}$ to a vector $\vec{b}$ as in $A\vec{x} = \vec{b}$. By multiplying A^{-1} into both sides of $A\vec{x} = \vec{b}$, we end up with

$$A^{-1}A\vec{x} = A^{-1}\vec{b}\,,$$

effectively 'solving' for the vector $\vec{x}$:

$$\vec{x} = A^{-1}\vec{b} \tag{1.79}$$

12.2 Formula

To come up with a formula for the inverse of A, consider another matrix B defined from the components A_{jk} such that

$$B_{jk} = (-1)^{j+k}\,M_{kj}\ , \tag{1.80}$$

where M_{kj} are the matrix minors of A. As innocent as it looks, (1.80) is quite 'computationally expensive', which means as N grows, it requires preposterous efforts to calculate the B-matrix by hand.

To proceed, calculate the product AB by matrix multiplication. Starting with the formula (1.63), and replacing the components of B using (1.80), we find

$$(AB)_{mn} = \sum_{k=1}^{N} A_{mk}B_{kn} = \sum_{k=1}^{N} A_{mk}\,(-1)^{k+n}\,M_{nk}\ .$$

In the case $m = n$, the above reduces to the formula for the determinant of A, namely (1.77). Any other case $m \neq n$ causes the right side to resolve to zero:

$$(AB)_{mn} = \begin{cases} \det A & m = n \\ 0 & m \neq n \end{cases}$$

In symbolic terms, the above reads

$$AB = (\det A)\, I\,,$$

where by comparison to (1.78), suggests the combination $B/\det A$ is equal to the inverse of A:

$$A^{-1} = \frac{1}{\det A} B \tag{1.81}$$

12.3 Products

The inverse of the product of two matrices is equal to the product of the individual inverses in reversed order:

$$(AB)^{-1} = B^{-1} A^{-1} \tag{1.82}$$

12.4 Cramer's Rule Derived

For a linear system of N dimensions, we start with the dichotomy

$$A\vec{x} = \vec{b} \qquad\qquad \vec{x} = A^{-1}\vec{b}\,,$$

where accounting for (1.81), the 'solution' vector $\vec{x}$ is written

$$\vec{x} = \frac{1}{\det A} B\vec{b}\,,$$

or using index notation,

$$x_j = \frac{1}{\det A} \sum_{k=1}^{N} B_{jk} b_k = \frac{1}{\det A} \sum_{k=1}^{N} (-1)^{j+k} M_{kj} b_k\,.$$

The product $M_{kj} b_k$ has the k, j indices in 'reverse' order, in the sense that the calculation $M\vec{b}$ does *not* represent this situation. Instead, the sum constitutes the determinant of a matrix modified from A such that the kth column is replaced by $\vec{b}$. If this situation sounds familiar, it precisely describes the matrix introduced as equation (1.73)

$$C_j = \begin{bmatrix} A_{11} & A_{12} & \cdots & b_{1j} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & b_{2j} & \cdots & A_{2N} \\ A_{31} & A_{32} & \cdots & b_{3j} & \cdots & A_{3N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{N1} & A_{N2} & \cdots & b_{Nj} & \cdots & A_{NN} \end{bmatrix}\,,$$

and the above reduces to Cramer's rule (1.72) for the soltuion of the system:

$$x_j = \frac{\det C_j}{\det A} \qquad\qquad j = 1, 2, 3, \ldots, N$$

12.5 Two Dimensions

Consider the two-dimensional matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

whose determinant is given by (1.67). To calculate the inverse, begin with the B matrix given by (1.80), coming out to

$$B = \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix} .$$

Then, by the inverse formula (1.81), the inverse of the 2×2 matrix reads:

$$A^{-1} = \frac{1}{A_{11}A_{22} - A_{12}A_{21}} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix} \tag{1.83}$$

12.6 Three Dimensions

The three-dimensional matrix with

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

has a total of nine minors M_{jk}, readily readable from A. Constructing the matrix B using

$$B_{jk} = (-1)^{j+k} M_{kj} ,$$

we find

$$B_{11} = (-1)^2 (A_{22}A_{33} - A_{32}A_{23}) \qquad B_{12} = (-1)^3 (A_{12}A_{33} - A_{32}A_{13})$$

$$B_{13} = (-1)^4 (A_{12}A_{23} - A_{22}A_{13}) \qquad B_{21} = (-1)^3 (A_{21}A_{33} - A_{31}A_{23})$$

$$B_{22} = (-1)^4 (A_{11}A_{33} - A_{31}A_{13}) \qquad B_{23} = (-1)^5 (A_{11}A_{23} - A_{21}A_{13})$$

$$B_{31} = (-1)^4 (A_{21}A_{32} - A_{31}A_{22}) \qquad B_{32} = (-1)^5 (A_{11}A_{32} - A_{31}A_{12})$$

$$B_{33} = (-1)^6 (A_{11}A_{22} - A_{21}A_{12}) ,$$

or in matrix form:

$$B = \begin{bmatrix} (A_{22}A_{33} - A_{32}A_{23}) & -(A_{12}A_{33} - A_{32}A_{13}) & (A_{12}A_{23} - A_{22}A_{13}) \\ -(A_{21}A_{33} - A_{31}A_{23}) & (A_{11}A_{33} - A_{31}A_{13}) & -(A_{11}A_{23} - A_{21}A_{13}) \\ (A_{21}A_{32} - A_{31}A_{22}) & -(A_{11}A_{32} - A_{31}A_{12}) & (A_{11}A_{22} - A_{21}A_{12}) \end{bmatrix}$$

With the matrix B fully specified in terms of A, the inverse A^{-1} is given by (1.81), namely

$$A^{-1} = \frac{1}{\det A} B .$$

Note that $\det A$ was already calculated as equation (1.75).

13 Matrix Forms

13.1 Transpose and Symmetry

Transpose Matrix

Given a matrix A, there always exists the *transpose* of A, which swaps all rows for columns and vice versa. The transpose of a matrix A is denoted A^T, particularly

$$A^T_{jk} = A_{kj} \, . \tag{1.84}$$

Symmetric Matrix

A square matrix is said to be *symmetric* if the original matrix A is equal to the transposed matrix A^T:

$$A = A^T \qquad\qquad A_{jk} = A_{kj} \tag{1.85}$$

Anti-symmetric Matrix

A square matrix is said to be *anti-symmetric* if the original matrix A is equal to the negative transposed matrix A^T:

$$A = -A^T \qquad\qquad A_{jk} = -A_{kj} \tag{1.86}$$

13.2 Role of Row Operations

A general $M \times N$ matrix

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{bmatrix}$$

can be *reduced* by any operations E, M, R to produce a different matrix A' that contains the same information or similar information to A. This process can be applied sequentially to achieve various matrix forms cataloged below.

13.3 Triangular Forms

Square matrices with $M = N$ admit two special reduced forms called *triangular forms*.

Upper Triangular Form

If (by row operations or otherwise) a square matrix has $A_{jk} = 0$ when $j > k$, the form is called *upper triangular*:

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ 0 & A_{22} & \cdots & A_{2n} \\ 0 & 0 & \cdots & \cdots \\ 0 & 0 & 0 & A_{nn} \end{bmatrix} \tag{1.87}$$

Lower Triangular Form

If a square matrix has $A_{jk} = 0$ when $j < k$, the form is called *lower triangular*:

$$A = \begin{bmatrix} A_{11} & 0 & 0 & 0 \\ A_{21} & A_{22} & 0 & 0 \\ \cdots & \cdots & \cdots & 0 \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \tag{1.88}$$

13.4 Diagonal Form

If a square matrix has $A_{jk} = 0$ when $j \neq k$, the form is called *diagonal*:

$$A = \begin{bmatrix} A_{11} & 0 & 0 & 0 \\ 0 & A_{22} & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & A_{nn} \end{bmatrix} \tag{1.89}$$

For any triangular or diagonal matrix A, the determinant is equal to the product of its diagonal entries:

$$\det A = A_{11}A_{22}\cdots A_{NN} = \prod_{j=1}^{N} A_{jj}$$

13.5 Augmented Matrix

For linear systems characterized by $A\vec{x} = \vec{b}$, where A is an $M \times N$ matrix, we can construct the *augmented* matrix by appending the components of $\vec{b}$ as an extra column:

$$A|b = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} & b_1 \\ A_{21} & A_{22} & \cdots & A_{2N} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{M1} & A_{M2} & \cdots & A_{MN} & b_M \end{bmatrix} \tag{1.90}$$

In the general case, $\vec{b}$ can be replaced with any matrix with M rows.

13.6 Row-Reduced Echelon Form

If (by any means) the a square matrix and a vector $\vec{x}$ can be written as

$$I|x = \begin{bmatrix} 1 & 0 & \cdots & 0 & x_1 \\ 0 & 1 & \cdots & 0 & x_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & x_N \end{bmatrix} , \tag{1.91}$$

this is called the *row-reduced echelon form.*

14 Elimination

14.1 Linear Systems

Consider an N-dimensional linear system $A\vec{x} = \vec{b}$, represented by the $M = N$-case of (1.70):

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + \cdots + A_{1N}x_N &= b_1 \\ A_{21}x_1 + A_{22}x_2 + A_{23}x_3 + \cdots + A_{2N}x_N &= b_2 \\ A_{31}x_1 + A_{32}x_2 + A_{33}x_3 + \cdots + A_{3N}x_N &= b_3 \\ &\cdots \\ A_{N1}x_1 + A_{N2}x_2 + A_{N3}x_3 + \cdots + A_{NN}x_N &= b_N \end{aligned}$$

Equivalently, the above is represented by an augmented matrix $A|b$ of the form (1.90). Next, imagine having done all of the hard work to solve the system

$$\begin{aligned} x_1 + 0 + 0 + \cdots + 0 &= x_1 \\ 0 + x_2 + 0 + \cdots + 0 &= x_2 \\ 0 + 0 + x_3 + \cdots + 0 &= x_3 \\ &\cdots \\ 0 + 0 + 0 + \cdots + x_N &= x_N \ , \end{aligned}$$

which appears like a tautological thing to write, but is in fact the row-reduced echelon form, $I|x$ cataloged as equation (1.91). Written this way, the solutions x_j to the system are readily exportable as the right side of each equation.

The natural question is, how can we start with $A|b$ and somehow end up with with $I|x$ using matrix trickery? The answer is called *elimination*, which is a sequence of row operations E, M, R that we carry out on the augmented matrix $A|b$ to bring it the form $I|x$. Representing the exact sequence of row operations as one 'operator' $\tilde{O}\,(E, M, R)$ or simply $\tilde{O}$, one writes

$$\tilde{O}\,(A|b) = I|x \ . \tag{1.92}$$

One may think of $\tilde{O}$ as a sequential list of procedures to carry out on $A|b$, much as a program receives input and returns output.

Example

Consider a linear system represented by the augmented matrix

$$A|b = \begin{bmatrix} 1 & 1 & 1 & 5 \\ 2 & 3 & 5 & 8 \\ 4 & 0 & 5 & 2 \end{bmatrix} .$$

Denoting the rows of $A|b$ as R_j, the first three operations may go as follows: (i) Subtract $2R_1$ from R_2. (ii) Subtract $4R_1$ from R_3. (iii) Add $4R_2$ to R_3.

$$A|b \xrightarrow{\text{(i)}} \begin{bmatrix} 1 & 1 & 1 & 5 \\ 0 & 1 & 3 & -2 \\ 4 & 0 & 5 & 2 \end{bmatrix} \xrightarrow{\text{(ii)}} \begin{bmatrix} 1 & 1 & 1 & 5 \\ 0 & 1 & 3 & -2 \\ 0 & -4 & 1 & -18 \end{bmatrix} \xrightarrow{\text{(iii)}} \begin{bmatrix} 1 & 1 & 1 & 5 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 13 & -26 \end{bmatrix}$$

Note the new matrix has zeros down and left of the diagonal, i.e. upper triangular form. Don't stop here though: (iv) Divide R_3 by 13 and subtract $3R_3$ from R_2. (v) Subtract R_3 from R_1. (vi) Subtract R_2 from R_1.

$$\xrightarrow{\text{(iv)}} \begin{bmatrix} 1 & 1 & 1 & 5 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -2 \end{bmatrix} \xrightarrow{\text{(v)}} \begin{bmatrix} 1 & 1 & 0 & 7 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -2 \end{bmatrix} \xrightarrow{\text{(vi)}} \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -2 \end{bmatrix} = I|x$$

Elimination halts when the 'matrix part' of the above reduces to the identity. Reading off the right-hand column, we see the solution to the system of equations is

$$x_1 = 3 \qquad\qquad x_2 = 4 \qquad\qquad x_3 = -2 \,.$$

Reconciling what just happened with equation (1.92), we see the operator $\tilde{O}$ is comprised of steps (i)-(vi), each being one particular E, M, R operation.

14.2 Matrix Inverse

Looking again at equation (1.92), note that the sequence of row operations $\tilde{O}$ applies to A and $\vec{b}$ separately:

$$\tilde{O}A = I \qquad\qquad \tilde{O}\vec{b} = \vec{x}$$

While $\tilde{O}$ is not established as a matrix, it does precisely same job as A^{-1}, and must contain the same information as A^{-1}. As a point of comparison, note the similarity between the above versus familiar relations

$$A^{-1}A = I \qquad\qquad A^{-1}\vec{b} = \vec{x} \,.$$

Going with the hunch that $\tilde{O}$ can be treated as an operator that obeys the associativity rule of matrix multiplication, we would be able to do the following:

$$\begin{aligned} \tilde{O}A &= I \\ \left(\tilde{O}A\right) A^{-1} &= IA^{-1} \\ \tilde{O}\left(AA^{-1}\right) &= A^{-1} \\ \tilde{O}I &= A^{-1} \end{aligned}$$

Once again, we see the sequence $\tilde{O}$ is doing the same job as A^{-1}. Rounding up the circumstancial evidence, we see the set of steps $\tilde{O}$ that carries $A \to I$ is the *same* set of steps that carries $I \to A^{-1}$. In the language of augmented matrices, this is summarized by

$$\tilde{O}\,(A|I) = I|A^{-1}\,. \tag{1.93}$$

This conspiracy of mathematics is otherwise known as *Gauss-Jordan elimination.*

Two Dimensions

Demonstrating on a 2×2 matrix, begin with $A|I$ as

$$A|I = \begin{bmatrix} A_{11} & A_{12} & 1 & 0 \\ A_{21} & A_{22} & 0 & 1 \end{bmatrix}\,,$$

and perform row operations until form $I|A^{-1}$ is attained. In brief detail, the augmented matrix develops as:

$$\begin{aligned}
A|I &\to \begin{bmatrix} A_{12}A_{21} - A_{22}A_{11} & 0 & -A_{22} & A_{21} \\ A_{21} & A_{22} & 0 & 1 \end{bmatrix} \\
A|I &\to \frac{1}{\det A}\begin{bmatrix} 1 & 0 & -A_{22} & A_{21} \\ 0 & -A_{22}/A_{21} & -A_{22} & A_{12} - \left(A_{12}A_{21} - A_{22}A_{11}\right)/A_{21} \end{bmatrix} \\
A|I &\to \frac{1}{\det A}\begin{bmatrix} 1 & 0 & A_{22} & -A_{12} \\ 0 & 1 & -A_{21} & A_{11} \end{bmatrix}
\end{aligned}$$

The final result is none other than (1.83), the formula for the inverse of a 2×2 square matrix:

$$A^{-1} = \frac{1}{A_{11}A_{22} - A_{12}A_{21}}\begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}$$

15 Eigenvectors and Eigenvalues

An important situation that arises in mathematics and physics is the so-called *eigenvalue problem*

$$A\vec{u} = \lambda\vec{u}\,. \tag{1.94}$$

The matrix A is taken to be square and N-dimensional. The vectors $\vec{u}^{(j)}$ that satisfy (1.94) are called *eigenvectors*, and the corresponding scalar $\lambda^{(j)}$ is called an *eigenvalue*.

15.1 Calculating Eigenvalues

The eivenvalue problem (1.94) can be equivalently framed as

$$\left(A - \lambda I\right)\vec{u} = 0\,, \tag{1.95}$$

where I is the identity matrix to match the dimension of A. Equation (1.95) is a special linear system like $A\vec{x} = \vec{b}$ with $\vec{b} = 0$, and can be solved accordingly.

Two Dimensions

Taking a two-dimensional case as an example, we have

$$A - \lambda I = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} A_{11} - \lambda & A_{12} \\ A_{21} & A_{22} - \lambda \end{bmatrix} ,$$

which, as a set of equations, looks like

$$\begin{aligned} (A_{11} - \lambda)\, x_1 &= -A_{12} x_2 \\ (A_{22} - \lambda)\, x_2 &= -A_{12} x_1 \end{aligned}$$

Multiply the pair of equations and cancel the product $x_1 x_2$ to get

$$(A_{11} - \lambda)(A_{22} - \lambda) - A_{12} A_{21} = 0 \,. \tag{1.96}$$

The only unknown in the equation is λ, which can be isolated using the quadratic formula:

$$\lambda_{\pm} = \frac{A_{11} + A_{22}}{2} \pm \frac{1}{2} \sqrt{(A_{11} - A_{22})^2 + 4 A_{12} A_{21}} \tag{1.97}$$

Note there are two solutions for λ, which we label λ_+, and λ_-, respectively.

15.2 Characteristic Equation

When confronted with the eigenvalue problem (1.95), the first order of business, usually, is to calculate the eigenvalues λ. As we've seen for the two-dimensional case, this process boiled down to equation (1.96). Pausing on this result for a moment, note that a quicker way to get there is to write

$$\det (A - \lambda I) = 0 \,, \tag{1.98}$$

which is in fact true in any number of dimensions. Equation (1.98) is called the *characteristic equation* of the system.

Characteristic Polynomial

The characteristic equation always 'simplifies' to the *characteristic polynomial*, a single equation embedding λ:

$$P_N(\lambda) = C_0 + C_1 \lambda + C_2 \lambda^2 + C_3 \lambda^3 + \cdots + C_N \lambda^N = 0 \tag{1.99}$$

The characteristic polynomial is suggestive of the *fundamental theorem of algebra*, stating that there are exactly N (complex) roots of a polynomial of degree N.

15.3 Calculating Eigenvectors

Once the eigenvalues λ are known, the components of each eigenvector $\vec{u}$ are readily calculated directly from

$$A \vec{u}^{(j)} = \lambda_j \vec{u}^{(j)} \qquad\qquad j = 1, 2, 3, \ldots, N \,.$$

Two Dimensions

Developing the eigenvalue problem in two dimensions, there are two eigenvalues $\lambda_\pm$ given by (1.97), and let us label the two corresponding eigenvectors $\vec{u}$, $\vec{v}$ such that

$$A\vec{u} = \lambda_+\vec{u} \qquad\qquad A\vec{v} = \lambda_-\vec{v}\,.$$

Working with the left equation first, it expands into two equations

$$\begin{aligned} A_{11}u_1 + A_{12}u_2 &= \lambda_+u_1 \\ A_{12}u_1 + A_{22}u_2 &= \lambda_+u_2\,, \end{aligned}$$

which gives us two ways to solve for the ratio u_1/u_2:

$$\frac{u_1}{u_2} = \frac{-A_{12}}{A_{11}-\lambda_+} \qquad\qquad \frac{u_1}{u_2} = \frac{-\left(A_{22}-\lambda_+\right)}{A_{21}} \tag{1.100}$$

As a sanity check, we may eliminate the ratio u_1/u_2 and recover the characteristic equation (1.96). A similar set of steps isolates the ratio v_1/v_2 for the second eigenvalue/eigenvector

$$\frac{v_1}{v_2} = \frac{-A_{12}}{A_{11}-\lambda_-} \qquad\qquad \frac{v_1}{v_2} = \frac{-\left(A_{22}-\lambda_-\right)}{A_{21}}\,, \tag{1.101}$$

which also combine to reproduce the characteristic equation, so we're on the right track.

Example: Symmetric Matrix

Suppose the matrix A is given as

$$A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}\,.$$

The eigenvalues of A are given by (1.97), and simplify very nicely:

$$\lambda_\pm = a \pm b$$

Denoting the respective eigenvectors $\vec{u}$, $\vec{v}$, we apply (1.100) directly to find

$$\frac{u_1}{u_2} = \frac{-b}{-b} = 1\,.$$

Meanwhile, (1.101) similarly tells us

$$\frac{v_1}{v_2} = \frac{-b}{b} = -1\,,$$

and we're done. Evidently, the two eigenvectors are

$$\vec{u} = \langle 1, 1\rangle \qquad\qquad \vec{v} = \langle 1, -1\rangle\,,$$

or in normalized form,

$$\hat{u} = \frac{1}{\sqrt{2}}\langle 1, 1\rangle \qquad\qquad \hat{v} = \frac{1}{\sqrt{2}}\langle 1, -1\rangle\,.$$

15.4 Diagonalization

For the eigenvalue problem (1.94)

$$A\vec{u} = \lambda\vec{u}$$

of dimension N, suppose we already have the list of N eigenvalues λ and N eigenvectors $\vec{u}$.

Modal Matrix

It's instructive to condense all of the eigenvector information into a new object called the *modal matrix*, denoted C, whose jth column is comprised of the components of the jth eigenvector:

$$C = \begin{bmatrix} u_1^{(1)} & u_1^{(2)} & \cdots & u_1^{(N)} \\ u_2^{(1)} & u_2^{(2)} & \cdots & u_2^{(N)} \\ \cdots & \cdots & \cdots & \cdots \\ u_N^{(1)} & u_N^{(2)} & \cdots & u_N^{(N)} \end{bmatrix} = \begin{bmatrix} \vec{u}^{(1)} & \vec{u}^{(2)} & \cdots & \vec{u}^{(N)} \end{bmatrix} \tag{1.102}$$

Then, the matrix product AC can be written

$$AC = \begin{bmatrix} \lambda_1\vec{u}^{(1)} & \lambda_2\vec{u}^{(2)} & \cdots & \lambda_N\vec{u}^{(N)} \end{bmatrix} = \begin{bmatrix} \lambda_1 u_1^{(1)} & \lambda_2 u_1^{(2)} & \cdots & \lambda_N u_1^{(N)} \\ \lambda_1 u_2^{(1)} & \lambda_2 u_2^{(2)} & \cdots & \lambda_N u_2^{(N)} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_1 u_N^{(1)} & \lambda_2 u_N^{(2)} & \cdots & \lambda_N u_N^{(N)} \end{bmatrix}.$$

Diagonal Matrix

The product AC, especially in matrix form, looks like the product of C with another, much simpler matrix. Consider a *diagonal* matrix Λ (Greek uppercase lambda) whose off-diagonal entries are all zero, and the eigenvalues occupy the diagonal:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_N \end{bmatrix} \tag{1.103}$$

Indeed, the right result of AC is reproduced by the product $C\Lambda$, meaning the matrix products are equal:

$$AC = C\Lambda$$

Supposing the inverse of C can be attained, the diagonal matrix Λ can be isolated:

$$\Lambda = C^{-1}AC \tag{1.104}$$

The process of attaining Λ is called the *diagonalization* of the matrix A. If the columns of C happen to form an orthonormal basis, the inverse matrix C^{-1} may be replaced with its transpose C^T.

15.5 Eigenvectors as a Basis

It is no coincidence that a system of N dimensions has N eigenvectors. It makes sense to wonder if an arbitrary linear combinaion (1.40) can be expressed via the change-of-basis formula (1.43) for vectors.

$$\vec{V} = \sum_{j=1}^{N} V_j \, \hat{e}_j \qquad \overset{?}{\to} \qquad (\vec{V})' = \sum_{j=1}^{N} V_j' \, \hat{u}^{(j)}$$

In the above, the eigenvectors are assumed to be normalized (unit magnitude), which is always possible for nonzero vectors. However, we are *not* to assume that the eigenvectors $\left\{\hat{u}^{(j)}\right\}$ form an orthogonal basis. That is, it's not always the case that any two eigenvectors are orthogonal.

Hermitian Matrix

Consider two solutions to the eigenvalue problem (1.94),

$$A\vec{u}^{(j)} = \lambda_j \vec{u}^{(j)} \qquad\qquad A\vec{u}^{(k)} = \lambda_k \vec{u}^{(k)} \, ,$$

and multiply $\vec{u}^{(k)}$, $\vec{u}^{(j)}$, onto the left and right sides respectively into each:

$$\vec{u}^{(k)} \cdot A\vec{u}^{(j)} = \lambda_j \vec{u}^{(k)} \cdot \vec{u}^{(j)} \qquad\qquad A\vec{u}^{(k)} \cdot \vec{u}^{(j)} = \lambda_k \vec{u}^{(k)} \cdot \vec{u}^{(j)}$$

Looking at the left side of each equation, it appears as if

$$\vec{u}^{(k)} \cdot A\vec{u}^{(j)} = A\vec{u}^{(k)} \cdot \vec{u}^{(j)} \tag{1.105}$$

wants to be true, but simply isn't in the general case. The special that satisfies (1.105) is called a *Hermitian* matrix.

Non-equal Eigenvalues

Pursuing the case where A is Hermitian, the above condenses to:

$$\lambda_j \vec{u}^{(k)} \cdot \vec{u}^{(j)} = \lambda_k \vec{u}^{(k)} \cdot \vec{u}^{(j)}$$

Now, if we assume that $\lambda_j \neq \lambda_k$, the *only* way to reconcile this result is that *non-equal eigenvectors of a Hermitian matrix are orthogonal*:

$$\vec{u}^{(k)} \cdot \vec{u}^{(j)} = 0$$

Just as importantly, this reinforces that the eigenvectors of a non-Hermitian matrix may not be orthogonal.

Equal Eigenvalues

If m of the N eigenvalues are equal, one speaks of *m-fold degeneracy*. In this case, the corresponding eigenvectors form a *vector subspace* of the original vector space that might admit its own orthonormal basis.

16 Degenerate Systems

Concerning the eigenvalue problem (1.94)

$$A\vec{u} = \lambda\vec{u}\,,$$

it could turn out that two eigenvalues λ_j, λ_k are equal, in which case we *may* be able to construct N unique eigenvectors $\vec{u}^{(j)}$, but not always. Specifically, for each repeated eigenvalue λ_j of multiplicity m_j, there must be m_j linearly independent eigenvectors. The ability to successfully do this depends on the system on hand.

16.1 Dead-end Case

Consider the matrix

$$A = \begin{bmatrix} -2 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}\,,$$

having a characteristic polynomial

$$(-2-\lambda)\,(1-\lambda)\,(-2-\lambda) = 0\,.$$

Evidently we find three eigenvalues, with two identical:

$$\lambda_1 = 1 \qquad\qquad \lambda_2 = -2 \qquad\qquad \lambda_3 = -2$$

Handling the easy case first, the eigenvector corresponding to λ_1 is calculated from $A\vec{u} = 1\vec{u}$, resulting in

$$\vec{u}^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\,.$$

Proceeding to the repeated eigenvalue case, we solve $A\vec{u} = -2\vec{u}$ to get a single eigenvector

$$\vec{u}^{(2,3)} = \frac{1}{\sqrt{10}}\begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}\,.$$

Note that $\vec{u}^{(1)}$ is linearly independent from $\vec{u}^{(2,3)}$, but not orthogonal. Since there is no obvious way to 'peel apart' the eigenvectors $\vec{u}^{(2,3)}$, the show stops here. The matrix A cannot be diagonalized.

16.2 Salvageable Case

Consider the matrix

$$A = \begin{bmatrix} 5 & -4 & 4 \\ 12 & -11 & 12 \\ 4 & -4 & 5 \end{bmatrix}\,,$$

having a characteristic polynomial

$$0 = \lambda^3 + \lambda^2 - 5\lambda + 3 \, ,$$

which factors into

$$0 = (\lambda - 1)(\lambda - 1)(\lambda + 3) \; .$$

We have three eigenvalues, with two identical:

$$\lambda_1 = 1 \qquad\qquad \lambda_2 = 1 \qquad\qquad \lambda_3 = -3$$

Handling the easy case first, the eigenvector corresponding to λ_3 is calculated from $A\vec{u} = -3\vec{u}$, leading to the relations

$$\begin{aligned} 2u_1 - u_2 + u_3 &= 0 \\ u_1 - u_2 + 2u_3 &= 0 \\ 3u_1 - 2u_2 + 3u_3 &= 0 \, , \end{aligned}$$

telling us the corresponding eigenvector is

$$\vec{u}^{(3)} = \frac{1}{\sqrt{11}} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \, ,$$

or any multiple.

Proceeding to the repeated eigenvalue case, we solve $A\vec{u} = \vec{u}$ to generate three copies of

$$u_1 - u_2 + u_3 = 0 \, .$$

With one equation and three unknowns, we may choose any *two* values to be arbitrary. For instance, we may choose $u_1 = 1$ with $u_2 = 0$, causing $u_3 = -1$. We then construct an eigenvector from these numbers:

$$\vec{u}^{(1)} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

On the other hand, we may choose $u_1 = 0$, $u_2 = 1$, causing $u_3 = 1$, to create another eigenvector, linearly independent from the others:

$$\vec{u}^{(2)} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

With three eigenvectors in hand, a modal matrix can be defined such that

$$C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 3 \\ -1 & 1 & 1 \end{bmatrix} \, ,$$

allowing the matrix A to be diagonalized using $\Lambda = C^{-1}AC$.

16.3 Normalizable Case

Consider the matrix

$$A = \begin{bmatrix} 1 & 0 & \sqrt{2} \\ 0 & 2 & 0 \\ \sqrt{2} & 0 & 0 \end{bmatrix} ,$$

having a characteristic polynomial

$$0 = (2 - \lambda)\left(-\lambda^2 + \lambda + 2\right) ,$$

indicating three eigenvalues, with two identical:

$$\lambda_1 = 2 \qquad\qquad \lambda_2 = 2 \qquad\qquad \lambda_3 = -1$$

Handling the easy case first, the eigenvector corresponding to λ_3 is calculated from $A\vec{x} = -\vec{x}$, leading to

$$\vec{u}^{(3)} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 0 \\ -\sqrt{2} \end{bmatrix} .$$

Proceeding to the repeated eigenvalue case, we solve $A\vec{u} = 2\vec{u}$ to get a single eigenvector

$$\vec{u}^{(1,2)} = \frac{1}{\sqrt{3\alpha^2/2 + \beta^2}} \begin{bmatrix} \alpha \\ \beta \\ \alpha/\sqrt{2} \end{bmatrix} ,$$

for two arbitrary constants α, β. The aim here is to tease two mutually orthogonal eigenvectors from the above, which means to require

$$\vec{u}^{(1)} \cdot \vec{u}^{(2)} = 0 .$$

This amounts to finding pairs of α_j, β_j that satisfy

$$\frac{3}{2}\alpha_1\alpha_2 + \beta_1\beta_2 = 0 .$$

Choosing $\alpha_1 = 0$ begins a fast avalanche that requires $\beta_1 = 1$, and also $\beta_2 = 0$, with α_2 remaining arbitrary. The remaining eigenvectors therefore read

$$\vec{u}^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad\qquad \vec{u}^{(2)} = \frac{1}{\sqrt{3/2}} \begin{bmatrix} 1 \\ 0 \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} \sqrt{2/3} \\ 0 \\ 1/\sqrt{3} \end{bmatrix} .$$

With three eigenvectors in hand, a modal matrix can be defined such that

$$C = \begin{bmatrix} 1 & 0 & \sqrt{2/3} \\ 0 & 1 & 0 \\ -\sqrt{2} & 0 & 1/\sqrt{3} \end{bmatrix} ,$$

allowing the matrix A to be diagonalized via $\Lambda = C^{-1}AC$. However, since the set of eigenvectors $\left\{\vec{u}^{(j)}\right\}$ form an orthonormal basis, we may further simplify the above using $C^{-1} = C^T$