MAST10007 Lecture Slides 2018 s1

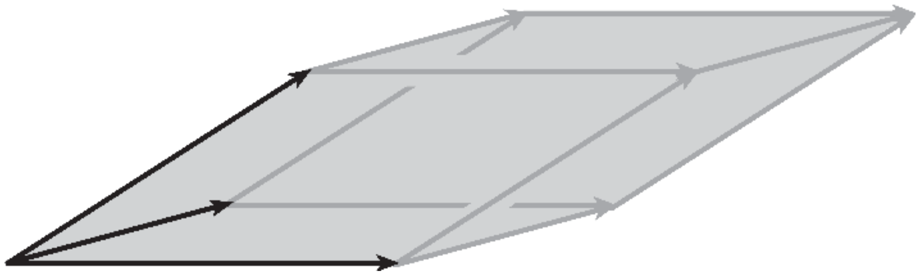
Linear Algebra (University of Melbourne)

School of Mathematics and Statistics

Semester 1, 2018

MAST10007 Linear Algebra

Lecture Slides



These notes are **not** intended as a textbook.
They are an accompaniment to the lectures.

You will need to take notes during the lectures.
This means more than merely filling in the blanks!

These slides have been printed on one side so that you can use the blank side for note-taking.

MAST10007 Linear Algebra

Topic 1: Linear equations [AR 1.1 and 1.2]

One of the major topics studied in linear algebra is systems of linear equations and their solutions.

We will study an efficient and systematic technique for solving simultaneous linear equations.

1.1. Systems of equations. Coefficient arrays. Row operations.

1.2. Reduction of systems to row-echelon form and reduced row-echelon form.

1.3. Consistent and inconsistent systems. Infinite solution sets.

1.1 Systems of equations, coefficient arrays, row operations

We begin with two examples to illustrate how linear equations can arise. We just set up the equations without solving them.

Example Let $V \subset \mathbb{R}^3$ be the set of solutions of the two equations

$$\begin{aligned} x - y - z &= 0 \qquad (1)\\ 6y - 3z &= 0 \end{aligned}$$

Linearity: if (x_1, y_1, z_1) and (x_2, y_2, z_2) are solutions of (1) then:

- $(x_1 + x_2, y_1 + y_2, z_1 + z_2)$ is a solution of (1)
- (ax_1, ay_1, az_1) is a solution of (1) for any $a \in \mathbb{R}$.

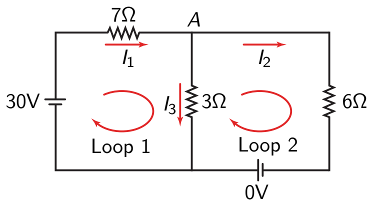
A *vector space* over $\mathbb{R}$ is a set V satisfying for any $u, v \in V$ and $a \in \mathbb{R}$:

- $u + v \in V$
- $au \in V$.

It satisfies other (rather easy) properties which we will meet later.

Example

The numbers I_1, I_2, I_3 give the current (in Amps) as indicated.



- The current through point A gives: $I_1 - I_2 - I_3 = 0.$
- The voltage drop around Loop 2 gives: $6I_2 - 3I_3 = 0.$
- The voltage drop around Loop 1 gives: $7I_1 + 3I_3 = 30.$

Solving these three simultaneous equations gives the current flowing through all parts of the circuit.

(We've used Kirchoff's Laws and Ohm's Law in writing down the equations in this example. You don't need to know what they are!)

Linear equations

In the previous example, the unknowns were I_1, I_2, I_3, and a typical equation obtained was $7I_1 + 3I_3 = 30$.

We call this a *linear equation* in the variables I_1 and I_3 because the coefficients are constants and the variables are raised to the first power only.

Definition (Linear equation and linear system)

A *linear equation* in n variables, $x_1, x_2, \ldots, x_n$, is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

where $a_1, \ldots, a_n$ and b are constants, and not all the a_i are zero.

A finite collection of linear equations in the variables $x_1, x_2, \ldots, x_n$ is called a *system of linear equations* or a *linear system*.

Examples

$$\left\{ \begin{aligned} x + 2y &= 7 \\ \frac{3}{8}x - 21y &= 0 \end{aligned} \right\}$$

$$\left\{ \begin{aligned} x_1 + 5x_2 + 6x_3 &= 100 \\ x_2 - x_3 &= -1 \\ -x_1 + x_3 &= 11 \end{aligned} \right\}$$

Definition (Solution of a system of linear equations)

A *solution to a system of linear equations* in the variables $x_1, \ldots, x_n$ is a set of values of these variables which satisfy *every* equation in the system.

How would you solve the following linear system?

$$\begin{Bmatrix} 2x - y = 3 \\ x + y = 0 \end{Bmatrix}$$

Graphically

- Need accurate sketch!
- Not practical for three or more variables.

Elimination

- Will always give a solution, but is too adhoc, particularly in higher dimensions (meaning three or more variables).

Coefficient arrays

As an example, consider the previously encountered linear system

$$\left\{\begin{array}{rcrcrl} I_1 & - & I_2 & - & I_3 & = 0 \\ 7I_1 & + & (0\times)I_2 & + & 3I_3 & = 30 \\ (0\times)I_1 & + & 11I_2 & - & 3I_3 & = 50 \end{array}\right\}.$$

The coefficients of the unknowns I_1, I_2, I_3 can be written as a 3×3 array of numbers:

$$\begin{bmatrix} 1 & -1 & -1 \\ 7 & 0 & 3 \\ 0 & 11 & -3 \end{bmatrix}.$$

Definition (Matrix)

A *matrix* is a rectangular array of numbers.
The numbers in the array are called the *entries* of the matrix.
A *$p \times q$ matrix* has p rows and q columns.

(Matrices are discussed in further detail in Topic 2.)

Definition (Augmented matrix of a linear system)

The *augmented matrix* for a linear system is the matrix formed from the coefficients in the equations and the constant terms.

Example

The augmented matrix for the previous set of equations is:

Example

Write the following system of linear equations as an augmented matrix:

$$\left\{ \begin{array}{r} 2x - y = 3 \\ x + y = 0 \end{array} \right\}.$$

Note

The number of rows is equal to the number of equations.
Each column, except the last, corresponds to a variable.
The last column contains the constant term from each equation.

Row Operations

Our aim is to use matrices to assist us in finding a solution to a system of equations.

We will perform operations on the augmented matrix. An essential condition is that whichever operations we perform, we must be able to recover the solution to the original system from the new matrix we obtain.

Definition (Elementary row operations)

The *elementary row operations* are:

1. Interchanging two rows.
2. Multiplying a row by a non-zero constant.
3. Adding a multiple of one row to another.

Example

Back to our simple system:

$$\left\{\begin{aligned} 2x - y &= 3 \\ x + y &= 0 \end{aligned}\right\}.$$

Let's apply some elementary row operations to the corresponding augmented matrix:

$$\left[\begin{array}{cc|c} 2 & -1 & 3 \\ 1 & 1 & 0 \end{array}\right] \sim$$

Note

The matrices are not equal, but are equivalent meaning the solution set is the same for each system represented by each augmented matrix.

1.2 Reduction of systems to reduced row-echelon form

Gaussian elimination

Using a sequence of elementary row operations, we can get to a simpler matrix that represents an equivalent linear system.
The leftmost non-zero element in each row is called the *leading entry*.

Definition (Row-echelon form)

A matrix is in *row-echelon form* if:

1. For any row with a leading entry, all elements below that entry and in the same column as it, are zero.
2. For any two rows, the leading entry of the lower row is further to the right than the leading entry in the higher row.
3. Any row that consists solely of zeros is lower than any row with non-zero entries.

Examples

$$\begin{bmatrix} 1 & -2 & 3 & 4 & 5 \end{bmatrix}, \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 4 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{array}\right] \qquad \text{r.e. form}$$

$$\begin{bmatrix} 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 3 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & -3 & 6 & -4 & 9 \end{bmatrix} \qquad \text{not r.e. form}$$

Gaussian elimination is an algorithm with input a matrix and output a matrix in row-echelon form which describes an equivalent linear system.

Gaussian elimination

1. Interchange rows, if necessary, to bring a non-zero number to the top of the first column with a non-zero entry.
2. Add suitable multiples of the top row to lower rows so that all entries below the leading entry are 0.
3. Start again at Step 1 applied to the matrix minus the first row.

Example

Use Gaussian elimination to reduce the augmented matrix which represents the linear system

$$\left\{\begin{array}{rcrcrcr} 3x &+& 2y &-& z &=& -15 \\ x &+& y &-& 4z &=& -30 \\ 3x &+& y &+& 3z &=& 11 \\ 3x &+& 3y &-& 5z &=& -41 \end{array}\right\}$$

to row-echelon form.

By reducing a matrix to row-echelon form, Gaussian elimination allows us to easily solve a system of linear equations.

Example

From the row-echelon matrix of the previous example we can calculate the solutions to the original system.

Note

This procedure relies on the fact that the new row-echelon matrix gives a linear system with exactly the same set of solutions as the original linear system.

Definition (Reduced row-echelon form)

A matrix is in *reduced row-echelon form* if the following three conditions are satisfied:

1. It is in row-echelon form.
2. Each leading entry is equal to 1 (called a *leading 1*).
3. In each column containing a leading 1, all other entries are zero.

Examples

$$\left[\begin{array}{cccc|c} 1 & -2 & 3 & -4 & 5 \end{array}\right] \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{are in r.r.e form}$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 3 \end{array}\right] \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 2 & 4 \\ 0 & 1 & 3 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -4 & 9 \end{bmatrix} \quad \text{are not in r.r.e form}$$

Gauss-Jordan elimination

Gauss-Jordan elimination is a systematic way to reduce a matrix to reduced row-echelon form using row operations.

Gauss-Jordan elimination

1. Use Gaussian elimination to reduce matrix to row-echelon form.
2. Use row operations (of type 3) to create zeros above the leading entries.
3. Multiply rows by appropriate numbers (type 2 row ops) to create the leading 1's.

The order of operations is not unique, however the reduced row-echelon form of a matrix is!

Example

Use Gauss-Jordan elimination to find a solution to the linear system

$$\left\{\begin{array}{rcrcrcr} 3x & + & 2y & - & z & = & -15 \\ x & + & y & - & 4z & = & -30 \\ 3x & + & y & + & 3z & = & 11 \\ 3x & + & 3y & - & 5z & = & -41 \end{array}\right\}.$$

1.3 Consistent and inconsistent systems

Recall that a solution of a system of linear equations is a set of values of the variables which satisfy every equation in the system.

In the preceding example we saw a linear system having a unique solution. Not all linear systems have a unique solution.

Example

Find all solutions of the system

$$\left\{\begin{array}{rcl} x - y + z &=& 3 \\ x - 7y + 3z &=& -11 \\ 2x + y + z &=& 16 \end{array}\right\}.$$

What's going on here?

Another example

Find all solutions of the system

$$\left\{\begin{array}{rcl} x + y + z &=& 4 \\ 2x + y + 2z &=& 9 \\ 3x + 2y + 3z &=& 13 \end{array}\right\}.$$

Is this the same behaviour as the previous example?

Types of solution sets

As seen in the above examples there are different types of solutions possible for systems of equations.

- The system can have no solution.
  We say that the system is *inconsistent*.
- The system can have exactly one solution.
  We say that the system is *consistent*.
- The system can have infinitely many solutions.
  We (again) say that the system is *consistent*.

Inconsistent systems

We can determine the type (consistent or inconsistent) of a system by reducing its augmented matrix to row-echelon form.

The system is inconsistent if there is at least one row in the row-echelon matrix having all values equal to zero on the left and a non-zero entry on the right. For example,

$$\left[\begin{array}{cccc|c} 0 & 0 & \cdots & 0 & 5 \end{array}\right]$$

Why is this inconsistent?

If we try to recover the equation represented by this row, it says:

$$0 \times x_1 + 0 \times x_2 + \cdots + 0 \times x_n = 5$$

and of course this is not satisfied for any values of $x_1, \ldots, x_n$.

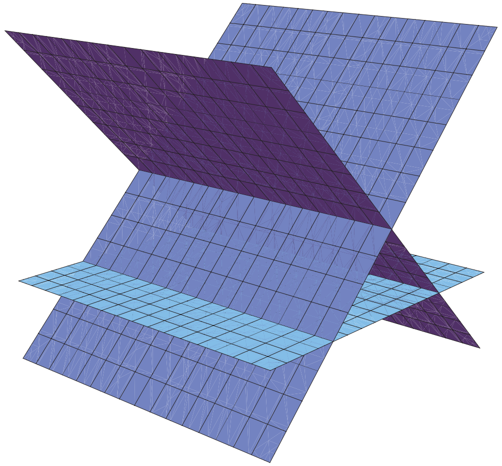
Example

$$\left[\begin{array}{ccc|c} 1 & 2 & 0 & 4 \\ 2 & 1 & 1 & 2 \\ 4 & 2 & 2 & 3 \end{array}\right]$$

If a system of equations is inconsistent, then the row-echelon form of its augmented matrix will have a row of the form $\left[\begin{array}{cccc|c} 0 & 0 & \cdots & 0 & a \end{array}\right]$ with $a \neq 0$.

Example (inconsistent)

Geometrically, an inconsistent system is one for which there is no common point of intersection for the planes determined by the system.



Consistent systems

Recall that a *consistent* system has either a unique solution or infinitely many solutions.

Unique solution:

For a consistent system of equations with n variables, a unique solution exists precisely when the row reduced augmented matrix has n non-zero rows. In this case we can read off the solution straight from the reduced row-echelon form of the matrix.

Example

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 5 \\ 1 & 2 & 1 & 4 \\ 2 & 1 & 1 & 4 \end{array}\right] \sim$$

and the solution is

$$x_1 = \qquad\qquad x_2 = \qquad\qquad x_3 =$$

Infinitely many solutions:

Example

Find all solutions to the equations with r.r.e. form:

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 5 & 1 \\ 0 & 0 & 1 & 0 & 6 & 2 \\ 0 & 0 & 0 & 1 & 7 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The corresponding (non-zero) equations are
$x_1 + 2x_2 + 5x_5 = 1, \quad x_3 + 6x_5 = 2, \quad x_4 + 7x_5 = 3.$

We can choose x_2 and x_5 as we wish. Say $x_2 = s, x_5 = t$. Then the other variables must be given by
$x_1 = 1 - 2s - 5t, \quad x_3 = 2 - 6t, \quad x_4 = 3 - 7t.$

In this way, we can describe every possible solution.

In general

Suppose we have a consistent linear system with n variables.

- If the row-reduced augmented matrix has $< n$ non-zero rows, then the system has infinitely many solutions.
- If r is the number of non-zero rows in the row-echelon form, then $n - r$ parameters are needed to specify the solution set.

More precisely, in the row-echelon form of the matrix, there will be $n - r$ columns which contain no leading entry. We can choose the variable corresponding to such a column arbitrarily. The values for the remaining variables will then follow.

Example

Suppose that the r.r.e. matrix for a system of equations is

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

We can represent the values of x_2 and x_4 by *parameters* s and t, respectively.

Then the values of the other variables are

$$x_1 = \qquad\qquad x_3 =$$

Example

Solve the linear system:

$$\left\{ \begin{array}{rl} v - 2w + z &= 1 \\ 2u - v - z &= 0 \\ 4u + v - 6w &= 3 \end{array} \right\}.$$

Example

Find the values of k for which the system

$$\left\{ \begin{array}{rl} x_1 + 3x_2 + 4x_3 &= 6 \\ 4x_1 + 9x_2 - x_3 &= 4 \\ 6x_1 + 9x_2 + kx_3 &= 8 \end{array} \right\}$$

has

(i) no solution

(ii) a unique solution

(iii) an infinite number of solutions

Topic 2: Matrices and Determinants [AR 1.3–1.7, 2.1–2.4]

In studying linear systems we introduced the idea of a matrix. Next we see that matrices are not only useful tools for solving systems of equations but that they have their own algebraic structure and have many other interesting properties.

Much of the material will have been seen already in your previous studies.

2.1 Some notation

We denote by A_{ij} the entry in the i-th row and j-th column of A:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix} \quad \text{or} \quad A = [A_{ij}]\,.$$

A matrix that has m rows and n columns has *size* $m \times n$. If $m = n$ then the *trace* of A is defined to be

$$Tr(A) = A_{11} + A_{22} + \dots + A_{nn}.$$

Example

The matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ \pi & e & 27.1 \end{bmatrix}$ has size 2×3.

Some entries are: $A_{12} = 2, \quad A_{21} = \pi, \quad A_{23} = 27.1.$

Some special matrices

- A matrix with the same number of rows as columns is a *square matrix*.
- A matrix with only one row is called a *row matrix*.
- A matrix with only one column is called a *column matrix*.
- A matrix with all elements equal to zero is a *zero matrix*,

  eg. $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.
- A square matrix with $A_{ij} = 0$ for $i \neq j$ is called a *diagonal matrix*,

  eg. $\begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$.
- A square matrix A satisfying $A_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ is called an *identity matrix*. The identity matrix of size $n \times n$ is denoted by I_n,

  eg. $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

2.2 Matrix operations

Some operations on matrices that we will look at are:
addition, scalar multiplication, multiplication and transposing.

Definition (Scalar multiple)

Let A be a matrix and $c \in \mathbb{R}$. The product *cA* is the matrix obtained by multiplying all entries of A by c. It is called a *scalar multiple* of A:

$$(cA)_{ij} = c \times A_{ij}.$$

Definition (Addition of matrices)

Let A and B be matrices of the same size. The *sum $A + B$* is the matrix obtained by adding corresponding entries of A and B. It has the same size as A and B.

$$(A + B)_{ij} = A_{ij} + B_{ij}$$

Be Careful: Matrices of different sizes cannot be added.

Notation: We write $A - B$ in place of $A + (-1)B$

Example

Let

$$A = \begin{bmatrix} 2 & 0 & -3 \\ 1 & -1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & -1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} -1 & 1 \\ 2 & 0 \end{bmatrix}.$$

Calculate (where possible) $A + B$ and $(A + B) + C$.

Properties of Matrix Addition

For matrices A, B and C, all of the same size, the following properties hold:

1. $A + B = B + A$. (commutativity)
2. $A + (B + C) = (A + B) + C$. (associativity)
3. $A - A = 0$.
4. $A + 0 = A$.

Here 0 denotes the zero matrix of the same size as A, B and C.

All these properties follow from the corresponding properties of the scalars (real or complex numbers).

Definition (Matrix multiplication)

Let A be an $m \times n$ matrix and B be a $n \times q$ matrix. The *product AB* of A and B is a matrix of size $m \times q$. The entry in position ij of the matrix product is obtained by taking row i of A, and column j of B, then multiplying together the entries in order and adding.

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

Note: The matrix product AB is only defined if the number of columns of A is equal to the number of rows of B.

Example

Let $A = \begin{bmatrix} 1 & -1 \\ 3 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 0 \\ -7 & 1 \end{bmatrix}$

Calculate AB and BA (if they exist).

Example

Let $A = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}$.

Calculate AB and BA.

Note: In this example $AB \neq BA$, even though both are defined.

Matrix multiplication is not commutative (in general).

Properties of matrix multiplication

The following properties hold whenever the matrix products and sums are defined:

1. $A(B + C) = AB + AC$. (left distributivity)
2. $(A + B)C = AC + BC$. (right distributivity)
3. $A(BC) = (AB)C$. (associativity)
4. $A(\alpha B) = \alpha(AB)$.
5. $AI_n = I_m A = A$ (where A has size $m \times n$).
6. $A0 = 0$ and $0A = 0$.

Here α is a scalar and 0 denotes zero matrix of the appropriate size.

Matrix powers

If A is a square matrix and $n \geqslant 1$ is an integer we define $A^n = AA \cdots A$ as the product of n copies of A.

Matrix transpose

Definition (Transpose of a matrix)

Let A be an $m \times n$ matrix. The *transpose of A*, denoted by A^T, is defined to be the $n \times m$ matrix whose entries are given by interchanging the rows and columns of A:

$$(A^T)_{ij} = A_{ji}.$$

Example

Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$. Then $A^T =$

Properties of the transpose

1. $(A^T)^T = A$.
2. $(A + B)^T = A^T + B^T$ (whenever $A + B$ is defined).
3. $(\alpha A)^T = \alpha A^T$ (where α is a scalar).
4. $(AB)^T = B^T A^T$ (whenever AB is defined).

Parts 1, 2 and 3 follow easily from the definition.

To prove part 4, we also use the definition of matrix multiplication given previously.

Exercise

Prove these!

A non-property of matrix multiplication

If two real numbers a and b satisfy $ab = 0$, then at least one of a and b is equal to 0. This is not always true for matrices.

Example

Let $A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$.
Then $AB = 0$, but $A \neq 0$ and $B \neq 0$.

What is going on here? The point is that the matrices A and B in the above example do not have inverses.

2.3 Matrix inverses

Definition (Matrix inverse)

A matrix A is called *invertible* if there exists a matrix B such that $AB = I$ and $BA = I$, where I is an identity matrix.
The matrix B is called the *inverse* of A and is denoted by A^{-1}. If A is not invertible, we say that A is *singular*.

We can prove from the definition that:

- For A to be invertible it must be square.
- (If it exists) A^{-1} has the same size as A.
- (If it exists) A^{-1} is unique.
- If A is invertible, then A^{-1} is invertible and $(A^{-1})^{-1} = A$.
- $I^{-1} = I$, 0 has no inverse.
- A matrix that has a row consisting entirely of zeros is singular.

Properties of the matrix inverse

If A and B are invertible matrices of the same size, and α is a non-zero scalar, then

1. $(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$.
2. $(AB)^{-1} = B^{-1}A^{-1}$.
3. $(A^n)^{-1} = \left(A^{-1}\right)^n \qquad$ (for all $n \in \mathbb{N}$).
4. $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

Exercise

Prove these!

Inverse of a 2×2 matrix

In general, for a 2×2 matrix A$= \begin{bmatrix} a & b \\ c & d \end{bmatrix}$:

1. A is invertible iff $(ad - bc) \neq 0$.
2. If $(ad - bc) \neq 0$, then $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Example

Find the inverse of $A = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$.

Calculating the inverse of a matrix

Algorithm: input: $n \times n$ matrix A
output: A^{-1} or "A is not invertible".

1. Construct the ("grand augmented") matrix $[A \mid I]$, where I is the $n \times n$ identity matrix.
2. Apply row operations to $[A \mid I]$ to get the block corresponding to A into reduced row-echelon form. This gives

$$[A \mid I] \sim [R \mid B]$$

where R is in reduced row-echelon form.

3. If $R = I$, then A is invertible and $A^{-1} = B$.
If $R \neq I$, then A is singular , i.e. A^{-1} does not exist.

We will see shortly why this method works.

Example

Find the inverse of $A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & -1 & 1 \\ 0 & 1 & 3 \end{bmatrix}$.

Row operations and matrix multiplication

The effect of a row operation can be achieved by multiplication on the left by a suitable matrix.

Definition (Elementary matrix)

An $n \times n$ matrix is an *elementary matrix* if it can be obtained from I_n by performing a single elementary row operation.

Examples

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}.$$

Let E_p be the elementary matrix obtained by applying a row operation p to the identity matrix I_n.

If A is a matrix such that the product E_pA is defined,
then the product E_pA is equal to the result of performing p on A.

We can perform a sequence of elementary row operations using a corresponding sequence of elementary matrices.
(Be careful about the order!)

Example

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \xrightarrow{p_1} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} \xrightarrow{p_2} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = B,$$

$$E_1 = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \qquad E_2 = \begin{bmatrix} 1 & 0 \\ 0 & \frac{-1}{2} \end{bmatrix},$$

$$E_2E_1A = \begin{bmatrix} 1 & 0 \\ 0 & \frac{-1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = B.$$

If $A \sim I$, then there is a sequence of elementary matrices $E_1, E_2, \ldots, E_n$ such that $E_n E_{n-1} \cdots E_2 E_1 A = I$.

This can be used to prove the following:

Theorem (AR 1.5.3)

1. *Let A be an $n \times n$ matrix. Then,*

$$A \text{ is invertible} \iff A \sim I_n.$$

2. *If A and B are $n \times n$ matrices such that $AB = I_n$, then A is invertible and $B = A^{-1}$.*
3. *Every invertible matrix can be written as a product of elementary matrices.*

This theorem justifies why our procedure for finding inverses using row operations actually works.... Can you see why?

Linear systems revisited

Any linear system in the variables $x_1, \ldots, x_n$ can be written in the form $A\mathbf{x} = \mathbf{b}$ where A is an $m \times n$ matrix, $\mathbf{b}$ is an $m \times 1$ matrix and $\mathbf{x} = \begin{bmatrix} x_1 \cdots x_n \end{bmatrix}^T$.

In the special case that A is invertible, the solution can be found using the inverse of A.

Theorem

If A is an invertible matrix, then a linear system of the form $A\mathbf{x} = \mathbf{b}$ has a unique solution. It is given by $\mathbf{x} = A^{-1}\mathbf{b}$

Proof: $A\mathbf{x} = \mathbf{b} \implies A^{-1}A\mathbf{x} = A^{-1}\mathbf{b} \implies \mathbf{x} = A^{-1}\mathbf{b}$ $\square$

Example

Use a matrix inverse to solve the linear system

$$\left\{\begin{array}{rcrcrcr} x & + & 2y & + & z & = & -3 \\ -x & - & y & + & z & = & 11 \\ & & y & + & 3z & = & 21 \end{array}\right\}.$$

(Note that we've previously calculated the inverse of the matrix $\begin{bmatrix} 1 & 2 & 1 \\ -1 & -1 & 1 \\ 0 & 1 & 3 \end{bmatrix}$).

2.4 Rank of a matrix

Definition (Rank of a matrix)

The *rank* of a matrix A is the number of non-zero rows in the reduced row-echelon form of A.

Note

- This is the same as the number of non-zero rows in a row-echelon form of A.
- If A has size $m \times n$, then clearly $\operatorname{rank}(A) \leqslant m$.
  We will see later that $\operatorname{rank}(A^T) = \operatorname{rank}(A)$, from which it follows that $\operatorname{rank}(A) \leqslant n$.

Example

Find the rank of each of the following matrices:

$$\begin{bmatrix} 1 & 2 & 1 \\ -1 & -1 & 1 \\ 0 & 1 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & 1 & -2 \\ 1 & -3 & 0 & 5 \end{bmatrix}$$

Theorem

The linear system $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix, has:

1. *No solution if* $\operatorname{rank}(A) < \operatorname{rank}([A \,|\, \mathbf{b}])$.
2. *A unique solution if* $\operatorname{rank}(A) = \operatorname{rank}([A \,|\, \mathbf{b}])$ *and* $\operatorname{rank}(A) = n$.
3. *Infinitely many solutions if* $\operatorname{rank}(A) = \operatorname{rank}([A \,|\, \mathbf{b}])$ *and* $\operatorname{rank}(A) < n$.

Proof: This is just a restatement of the results in section 1.3 using the idea of rank. □

Note

It is always the case that $\operatorname{rank}(A) \leqslant \operatorname{rank}([A \,|\, \mathbf{b}])$ and $\operatorname{rank}(A) \leqslant n$.

Theorem

If A is an $n \times n$ matrix, the following conditions are equivalent:

1. *A is invertible.*
2. *$A\mathbf{x} = \mathbf{b}$ has a unique solution for any $\mathbf{b}$.*
3. *The rank of A is n.*
4. *The reduced row-echelon form of A is I_n.*

Proof:

$1 \Rightarrow 2$ We've seen before (slide 54).

$2 \Rightarrow 3$ Follows from what we already knew about linear systems.

$3 \Rightarrow 4$ Immediate from the definition of rank, and that fact that A is square.

$4 \Rightarrow 1$ Let R be the r.r.e. form of A, so $R = EA$ where $E = E_k E_{k-1} \dots E_1$ is a product of elementary matrices. So $I = EA$. We have already noted that this implies that A is invertible (slide 53). $\square$

2.5 Determinants [AR 2.1–2.3]

When we calculate the inverse of the 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ we see that the number $ad - bc$ is important:

- If $ad - bc \neq 0$ then we can find the inverse of A.
- If $ad - bc = 0$, then A is not invertible.

So this number plays an important role when we study A. We call it the *determinant* of A and write it as $\det(A)$.

The determinant also has an important *geometrical meaning*: the *area of the parallelogram* spanned by the vectors (a, b) and (c, d) is the absolute value of $\det(A)$.

We now examine how the determinant extends to a function that associates a real number to any square matrix (not just 2×2 matrices).

Defining the determinant

Definition (Determinant)

Let A be an $n \times n$ matrix. The *determinant* of A, denoted $\det(A)$ or $|A|$, can be defined as the signed sum of all the ways to multiply together n entries of the matrix, with all chosen from different rows and columns.

To determine the sign of the products, imagine all but the elements in the product in question are set to zero in the matrix. Now swap columns until a diagonal matrix results. If the number of swaps required is even, then the product has a $+$ sign, while if it is odd, it is to be given a $-$ sign.

Determinant of a 3×3 matrix

Suppose that

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

We can form a table of all the products of 3 entries taken from different rows and columns, together with the signs:

product	sign
$a_{11}a_{22}a_{33}$	$+$
$a_{11}a_{23}a_{32}$	$-$
$a_{12}a_{21}a_{33}$	$-$
$a_{12}a_{23}a_{31}$	$+$
$a_{13}a_{21}a_{32}$	$+$
$a_{13}a_{22}a_{31}$	$-$

Hence,

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

The formula for 3×3 matrices is complicated and it quickly becomes worse as the size of the matrix increases (there are $n!$ terms for an $n \times n$ matrix). But there are better ways to calculate determinants.

Cofactors

Here is one way to calculate (small) determinants practically.

Definition

Let A be a square matrix. The *(i, j)-cofactor* of A, denoted by C_{ij}, is the number given by

$$C_{ij} = (-1)^{i+j} \det (A(i, j))$$

where $A(i, j)$ is the matrix obtained from A by deleting the ith row and jth column.

Example

If $A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & -1 & 1 \\ 0 & 1 & 3 \end{bmatrix}$, then $A(2,3) = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $C_{23} =$

Cofactor Expansion

We can write the determinant of a 3×3 matrix A in terms of cofactors, e.g.

$$\det(A) = a_{11} C_{11} + a_{12} C_{12} + a_{13} C_{13}.$$

This is called the *cofactor expansion* along the first row of A.

Example

Calculate $\det \begin{bmatrix} 1 & 2 & 1 \\ -1 & -1 & 1 \\ 0 & 1 & 3 \end{bmatrix}$

Theorem (Cofactor expansion)

The determinant of an $n \times n$ matrix A can be computed by choosing any row (or column) of A and multiplying the entries in that row (or column) by their cofactors and then adding the resulting products.

That is, for each $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant n$,

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

(this is called cofactor expansion along the ith row)

and

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

(this is called cofactor expansion along the jth column).

Proof: We can prove this in the $n = 3$ case using the formula on slide 62. The proof for the general case is essentially the same, but gets more technical... □

How do you remember the sign of the cofactor?

The $(1, 1)$-cofactor always has sign $+$. Starting from there, imagine walking to the square you want using either horizontal or vertical steps. The appropriate sign will change at each step.

We can visualise this arrangement with the following matrix:

$$\begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

So, for example, C_{13} is assigned $+$ but C_{32} is assigned $-$.

Example

Calculate $\begin{vmatrix} 1 & -2 & 0 & 1 \\ 3 & 1 & 2 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & -2 & 1 & 2 \end{vmatrix}$.

Some properties of determinants:

Suppose A and B are square matrices. Then,

1. $\det\left(A^T\right) = \det(A)$.
2. If A has a row or column of zeros, then $\det(A) = 0$.
3. $\det(AB) = \det(A)\det(B)$.
4. If A is invertible, then $\det(A) \neq 0$ and $\det\left(A^{-1}\right) = \frac{1}{\det(A)}$.
5. If A is singular, then $\det(A) = 0$.
6. If $A = \left[\begin{array}{c|c} C & * \\ \hline 0 & D \end{array}\right]$, with C and D both square, then
$$\det(A) = \det(C)\det(D).$$

Idea of proof: 2: Cofactor expansion. 4: Follows from 3. 5: Follows from the theorem on slide 59, 2 and 3. 1, 3, 6: Need to use the definition... $\square$

Row operations and determinants

Calculating determinants via cofactors becomes a very large calculation as the size of the matrix increases.

We need a better way to calculate larger determinants.

First observe that for some types of matrix it is easy to write down their determinant.

Definition (Triangular Matrix)

A matrix is said to be *upper triangular* (respectively *lower triangular*) if all the elements below (respectively above) the main diagonal are zero.

Examples

$$\begin{bmatrix} 2 & -1 & 9 \\ 0 & 3 & 2 \\ 0 & 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 & 0 \\ 1 & 3 & 0 \\ 2 & -3 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Theorem

If A is an $n \times n$ triangular matrix, then $\det(A)$ is the product of the entries on the main diagonal of A.

Idea of proof: (Repeated) cofactor expansion along the first column. $\square$

Example

Let $A = \begin{bmatrix} 2 & -10 & 92 & -117 \\ 0 & 3 & 28 & -31 \\ 0 & 0 & -1 & 27 \\ 0 & 0 & 0 & 2 \end{bmatrix}$.

What is $\det(A)$?

We can use row operations to transform a matrix into triangular form. The effect on the determinant of each of the three types of elementary row operations are as follows.

Theorem

Let A be a square matrix.

1. *If B is obtained from A by swapping two rows (or two columns) of A, then $\det(B) = -\det(A)$.*
2. *If B is obtained from A by multiplying a row (or column) of A by the scalar α, then $\det(B) = \alpha \det(A)$.*
3. *If B is obtained from A by replacing a row (or column) of A by itself plus a multiple of another row (column), then $\det(B) = \det(A)$.*

Proof: The corresponding elementary matrices have determinants -1, α and 1 (respectively). Then use $\det(B) = \det(E)\det(A)$. □

Example

Calculate $\begin{vmatrix} 1 & 2 & 1 \\ -1 & -1 & 1 \\ 0 & 1 & 3 \end{vmatrix}$ and $\begin{vmatrix} 2 & -4 & 1 \\ 3 & -6 & 3 \\ 2 & 1 & 4 \end{vmatrix}$.

Example

Calculate $\begin{vmatrix} 1 & -2 & 0 & 1 \\ 3 & 1 & 2 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & -2 & 1 & 2 \end{vmatrix}$.

We collect here some properties of the determinant function, most of which we've already noted.

Theorem

Let A be an $n \times n$ matrix. Then,

1. $\det\left(A^T\right) = \det(A)$.
2. $\det(AB) = \det(A)\det(B)$.
3. $\det(\alpha A) = \alpha^n \det(A)$.
4. *If A is a triangular matrix, then its determinant is the product of the elements on the main diagonal.*
5. *If A has a row (or column) of zeros, then $\det(A) = 0$.*
6. *If A has a row (or column) which is a scalar multiple of another row (or column) then $\det(A) = 0$.*
7. *A is singular iff $\det(A) = 0$ (so A is invertible iff $\det(A) \neq 0$).*

Topic 3: Euclidean Vector Spaces

3.1 Vectors in $\mathbb{R}^n$ [AR 3.1]

Geometrically, a pair of real numbers (a, b) can be thought of as representing a *directed line segment* from the origin in the plane.

These can be added together, and multiplied by a real number $\alpha \in \mathbb{R}$.

Algebraically, these operations are given by:

$$\begin{aligned} \text{vector addition:} \quad & (a, b) + (c, d) = (a + c, b + d). \\ \text{scalar multiplication:} \quad & \alpha(a, b) = (\alpha a, \alpha b). \end{aligned}$$

Notation:

$$\begin{aligned} \mathbb{R}^2 &= \{(a, b) \mid a, b \in \mathbb{R}\} \\ &= \text{ the set of all ordered pairs of real numbers.} \end{aligned}$$

The algebraic approach to vector addition and scalar multiplication extends to 3 dimensions or more.

$$\begin{aligned}\mathbb{R}^n &= \{(x_1, x_2, \ldots, x_n) \mid x_i \in \mathbb{R} \text{ for } i = 1, 2, \ldots, n\} \\ &= \text{the set of all } \textit{n-tuples} \text{ of real numbers.}\end{aligned}$$

We will refer to elements of $\mathbb{R}^n$ as *vectors*.

We can add two vectors together and multiply a vector by a scalar:

$$\begin{aligned}\text{vector addition}: &\quad (x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n). \\ \text{scalar multiplication}: &\quad \alpha(x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n).\end{aligned}$$

Notation

We often denote by $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ the vectors in $\mathbb{R}^3$ given by

$$\mathbf{i} = (1, 0, 0), \qquad \mathbf{j} = (0, 1, 0), \qquad \mathbf{k} = (0, 0, 1).$$

Any vector in $\mathbb{R}^3$ can be written in terms of $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$:

$$\mathbf{u} = (u_1, u_2, u_3) = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}.$$

3.2 Dot product [AR 3.2]

Let

$$\mathbf{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n \quad \text{and} \quad \mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$$

be two vectors in $\mathbb{R}^n$.

Definition (Dot product)

We define the *dot product* (or *scalar product* or *Euclidean inner product*) by

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

This is the definition for real scalars. For complex scalars, see Slide 205.

Examples

$(3, -1) \cdot (1, 2) =$

$(\mathbf{i} + \mathbf{j} + \mathbf{k}) \cdot (-\mathbf{j} + \mathbf{k}) =$

Definition (magnitude)

The *length* (or *magnitude* or *norm*) of a vector $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$ is given by

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}.$$

It follows from Pythagoras' theorem that this corresponds to our geometric idea of length for vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Example

Let $\mathbf{u} = 2\mathbf{i} - \mathbf{j} + 2\mathbf{k}$. Then, $\|\mathbf{u}\| =$

A vector having length equal to 1 is called a *unit vector.*

Example

Find a unit vector parallel to $\mathbf{u}$.

The distance between points P and Q is the length of the vector $P - Q$.

Example

Find the distance between the points $P(1,3,-1)$ and $Q(2,1,-1)$.

The angle between two vectors is *defined* in terms of the dot product.

Definition (Angle)

The *angle* θ between two non-zero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is given by

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta, \quad \text{where} \quad 0 \leqslant \theta \leqslant \pi.$$

The angle defined in this way is exactly the usual angle between two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$.

That our definition of angle makes sense relies on the following fact.

Theorem (Cauchy-Schwarz Inequality for $\mathbb{R}^n$)

Let $\mathbf{u}$, $\mathbf{v}$ *be vectors in* $\mathbb{R}^n$. *Then,*

$$|\mathbf{u} \cdot \mathbf{v}| \leqslant \|\mathbf{u}\|\|\mathbf{v}\|,$$

with equality holding precisely when $\mathbf{u}$ *is a multiple of* $\mathbf{v}$.

Proof: Consider the quadratic polynomial given by
$(\mathbf{u} + t\mathbf{v}) \cdot (\mathbf{u} + t\mathbf{v})$... $\square$

Properties of the dot product

1. $\mathbf{u} \cdot \mathbf{v}$ is a scalar.
2. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.
3. $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$.
4. $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$.
5. $(\alpha \mathbf{u}) \cdot \mathbf{v} = \alpha(\mathbf{u}.\mathbf{v})$.
6. $\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$, where θ is the angle between $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$.

All of these can be proved directly from the definitions.

Note

Suppose $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$ are two vectors in $\mathbb{R}^n$.
If we write each as a row matrix $U = [u_1 \ \cdots \ u_n]$ $V = [v_1 \ \cdots \ v_n]$, then

$$\mathbf{u} \cdot \mathbf{v} = UV^T.$$

Example

Find the angle between the vectors $\mathbf{u} = 2\mathbf{i} + \mathbf{j} - \mathbf{k}$ and $\mathbf{v} = \mathbf{i} + 2\mathbf{j} + \mathbf{k}$.

Important special case:

If $\mathbf{u}, \mathbf{v}$ are non-zero vectors and $\mathbf{u} \cdot \mathbf{v} = 0$, then the angle θ between $\mathbf{u}, \mathbf{v}$ is $\pi/2$, i.e., the vectors are *perpendicular* or *orthogonal*.

3.3 Cross product of vectors in $\mathbb{R}^3$ [AR 3.5]

Let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ be two vectors in $\mathbb{R}^3$.

Definition (Cross product)

The *cross product* (or *vector product*) of $\mathbf{u}$ and $\mathbf{v}$ is the vector given by

$$\mathbf{u} \times \mathbf{v} = (u_2 v_3 - u_3 v_2)\mathbf{i} + (u_3 v_1 - u_1 v_3)\mathbf{j} + (u_1 v_2 - u_2 v_1)\mathbf{k}.$$

A convenient way to remember this is as a "determinant":

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{k}$$

(using cofactor expansion along the first row).

Algebraic properties of the cross product

1. $\mathbf{v} \times \mathbf{u} = -(\mathbf{u} \times \mathbf{v})$.
2. $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) + (\mathbf{u} \times \mathbf{w})$.
3. $(\alpha \mathbf{u}) \times \mathbf{v} = \alpha(\mathbf{u} \times \mathbf{v})$.
4. $\mathbf{u} \times \mathbf{0} = \mathbf{0}$.
5. $\mathbf{u} \times \mathbf{u} = \mathbf{0}$.
6. $\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = 0$.

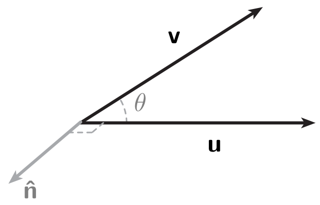
All of these can be proved directly from the definition.

Note

The cross product is defined *only* for $\mathbb{R}^3$. Unlike the dot product and many of the other properties we are considering, it does not extend to $\mathbb{R}^n$ in general.

Geometry of the cross product

$$\mathbf{u} \times \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \sin(\theta) \, \hat{\mathbf{n}}$$



- $\hat{\mathbf{n}}$ is a unit vector (i.e., $\|\hat{\mathbf{n}}\| = 1$).
- $\hat{\mathbf{n}}$ is perpendicular to both $\mathbf{u}$ and $\mathbf{v}$ (i.e., $\hat{\mathbf{n}} \cdot \mathbf{u} = 0$ and $\hat{\mathbf{n}} \cdot \mathbf{v} = 0$).
- $\hat{\mathbf{n}}$ points in the direction given by the right-hand rule.
- $\theta \in [0, \pi]$ is the angle between $\mathbf{u}$ and $\mathbf{v}$.
- If $\theta = 0$ or $\theta = \pi$, then $\mathbf{u} \times \mathbf{v} = \mathbf{0}$.

Example

Find a vector perpendicular to both $(2, 3, 1)$ and $(1, 1, 1)$.

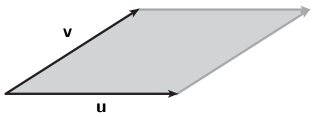
3.4 Geometric applications [AR 3.4]

Basic applications

Suppose that $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ are two vectors. Then:

1. $\mathbf{u}$ and $\mathbf{v}$ are *perpendicular* precisely when $\mathbf{u} \cdot \mathbf{v} = 0$.
2. The *area of the parallelogram* defined by $\mathbf{u}$ and $\mathbf{v}$ is equal to

$$\|\mathbf{u} \times \mathbf{v}\|.$$



Note

If $\mathbf{u}$ and $\mathbf{v}$ are elements of $\mathbb{R}^2$ with $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$, then

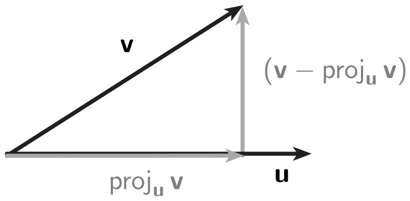
$$\text{area of parallelogram} = \text{absolute value of } \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}.$$

Example

Find the area of the triangle with vertices $(2,-5,4)$, $(3,-4,5)$ and $(3,-6,2)$.

3. Assuming $\mathbf{u} \neq \mathbf{0}$, the *projection of* $\mathbf{v}$ *onto* $\mathbf{u}$ is given by

$$\operatorname{proj}_{\mathbf{u}} \mathbf{v} = \frac{(\mathbf{u} \cdot \mathbf{v})}{\|\mathbf{u}\|^2} \mathbf{u}$$



Notice that:

- If we set $\hat{\mathbf{u}} = \mathbf{u}/\|\mathbf{u}\|$, then $\operatorname{proj}_{\mathbf{u}} \mathbf{v} = (\hat{\mathbf{u}} \cdot \mathbf{v})\, \hat{\mathbf{u}}$.
- $\mathbf{u} \cdot (\mathbf{v} - \operatorname{proj}_{\mathbf{u}} \mathbf{v}) =$

Example

Let $\mathbf{w} = (2, -1, -2)$ and $\mathbf{v} = (2, 1, 3)$.
Find vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ such that

- $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$,
- $\mathbf{v}_1$ is parallel to $\mathbf{w}$, and
- $\mathbf{v}_2$ is perpendicular to $\mathbf{w}$.

Scalar triple product

If $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are vectors in $\mathbb{R}^3$, then their *scalar triple product* is the real number $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ and is given by

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

Notice that

$$\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = 0.$$

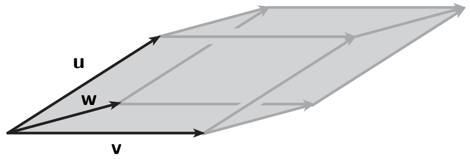
Similarly

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{u}) = \mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot (\mathbf{v} \times \mathbf{u}) = 0.$$

Geometric interpretation of the scalar triple product

Suppose that $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ are three vectors.

Then, the *parallelepiped* defined by $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$



has volume equal to the absolute value of the scalar triple product of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$.

$$\text{volume of parallelepiped} = |\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})| = \text{absolute value of } \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

Example

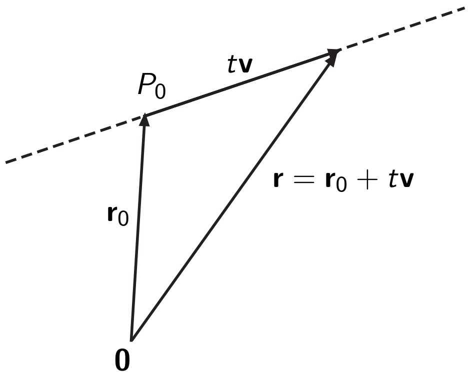
Find the volume of the parallelepiped with adjacent edges $\overrightarrow{PQ}$, $\overrightarrow{PR}$, $\overrightarrow{PS}$, where the points are $P(2,-1,1)$, $Q(4,6,7)$, $R(5,9,7)$ and $S(8,8,8)$.

Vector equation of a line

The *vector equation* of a line through a point P_0 in the direction determined by a vector $\mathbf{v}$ is

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}, \qquad t \in \mathbb{R},$$

where $\mathbf{r}_0 = \overrightarrow{OP_0}$.



Each point on the line is given by a unique value of t.

Letting $\mathbf{r} = (x, y, z)$, $\mathbf{r}_0 = (x_0, y_0, z_0)$ and $\mathbf{v} = (a, b, c)$, the equation becomes

$$(x, y, z) = (x_0, y_0, z_0) + t(a, b, c).$$

Parametric equations for a line

Equating components gives the *parametric equations* for the line:

$$\left\{ \begin{array}{l} x = x_0 + ta \\ y = y_0 + tb \\ z = z_0 + tc \end{array} \right\}, \qquad t \in \mathbb{R}.$$

Example

What is the parametric form of the line passing through the points $P(-1, 2, 3)$ and $Q(4, -2, 5)$?

Cartesian equations for a line

If $a \neq 0$, $b \neq 0$ and $c \neq 0$, we can solve the parametric equations for t and equate. This gives the *cartesian form* of the straight line:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

Example

What is the cartesian form of the line passing through the points $P(-1, 2, 3)$ and $Q(4, -2, 5)$?

Example

Find the vector equation of the line whose cartesian form is

$$\frac{x + 1}{5} = \frac{y - 3}{-1} = \frac{z - 4}{2}.$$

Definition

Two lines are said to:

- *intersect* if there is a point lying in both.
- be *parallel* if their direction vectors are parallel.
- be *skew* if they do not intersect and are not parallel.

The *angle* between two lines is the angle between their direction vectors.

Example

Find the vector equation of the line through the point $P(0, 0, 1)$ that is parallel to the line given by

$$\frac{x-1}{1} = \frac{y+2}{2} = \frac{z-6}{2}.$$

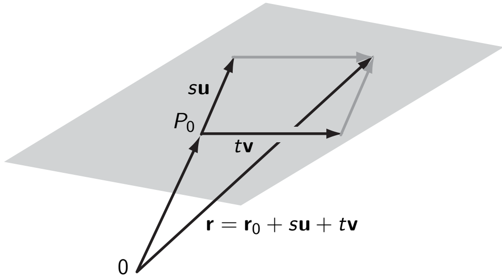
Planes

A plane is determined by a point P_0 on it and any two non-parallel vectors $\mathbf{u}$ and $\mathbf{v}$ lying in the plane.

Vector equation of a plane

Letting $\mathbf{r}_0 = \overrightarrow{OP_0}$, the *vector equation* of the plane is

$$\mathbf{r} = \mathbf{r}_0 + s\mathbf{u} + t\mathbf{v} \quad s, t \in \mathbb{R}.$$



Every value of s and t determines a point on the plane. Conversely, every point on the plane has position vector given by some s and t.

If $\mathbf{u}$ and $\mathbf{v}$ are two non-parallel vectors lying in a plane, then

$$\mathbf{n} = \mathbf{u} \times \mathbf{v}$$

is a non-zero vector that is perpendicular to the plane.

Such a vector is called a *normal vector* to the plane.

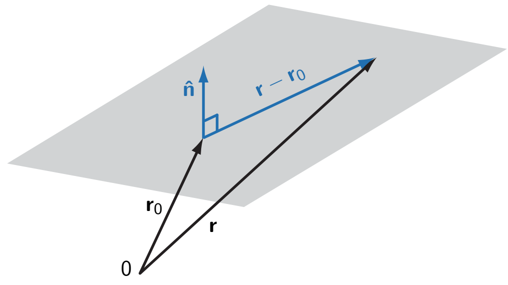
The angle between two planes is given by the angle between their normal vectors.

If P is a point with position vector $\mathbf{r}$, then

$$\begin{aligned} P \text{ lies on the plane} &\iff \mathbf{r} - \mathbf{r}_0 \text{ is parallel to plane} \\ &\iff \mathbf{r} - \mathbf{r}_0 \text{ is perpendicular to } \mathbf{n}. \end{aligned}$$

It follows that the equation of the plane can also be written as

$$(\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{n} = 0.$$



Writing $\mathbf{r} = (x, y, z)$ and $\mathbf{n} = (a, b, c)$, we obtain the *Cartesian equation* of the plane

$$ax + by + cz = d,$$

where $d = \mathbf{r}_0 \cdot \mathbf{n}$.

Examples

1. The plane perpendicular to the direction $(1, 2, 3)$ and through the point $(4, 5, 6)$ is given by $x + 2y + 3z = d$ where $d = 1 \times 4 + 2 \times 5 + 3 \times 6$:

$$x + 2y + 3z = 32.$$

2. Find Cartesian and vector equations for the plane perpendicular to $(1, 0, -2)$ and containing the point $(1, -1, -3)$.

3. The plane through $(1, 1, 1)$ containing vectors parallel to $(1, 0, 1)$ and $(0, 1, 2)$ is the set of all vectors of the form

$$(1, 1, 1) + s(1, 0, 1) + t(0, 1, 2) \qquad s, t \in \mathbb{R}.$$

4. Find the Cartesian equation of the plane with vector form

$$(x, y, z) = (1, 1, 1) + s(1, 0, 1) + t(0, 1, 2).$$

5. Find the vector and Cartesian equations of the plane containing the three points $P(2,1,-1)$, $Q(3,0,1)$, and $R(-1,1,-1)$.

Intersection of a line and a plane

Example

Where does the line

$$\frac{x-1}{1} = \frac{y-2}{2} = \frac{z-3}{3}$$

meet the plane $3x + 2y + z = 20$?

A typical point of the line is

Putting this into the equation for the plane, we get

This gives the point of intersection as

Intersection of two planes

Example Find a vector equation for the line of intersection of the two planes $x + 3y + 2z = 6$ and $3x + 2y + z = 11$.

Solving the two equations (for the planes) simultaneously gives

As a check: the direction of the line is perpendicular to both normals and so is given by

Distance from a point to a line

Example Find the (shortest) distance from the point $P(2, 1, 1)$ to the line with cartesian equation

$$\frac{x-2}{1} = \frac{y-1}{1} = \frac{z}{2}.$$

Topic 4: General Vector Spaces

Vectors in $\mathbb{R}^n$ have some basic properties shared by many other mathematical systems. For example,

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} \qquad \text{for all } \mathbf{u}, \mathbf{v}$$

is true for many other systems.

Key Idea:

Write down these basic properties and look for other systems which share these properties. Any system that does share these properties will be called a *vector space*.

4.1 The vector space axioms [AR 4.1]

Let's start trying to write down the basic properties that we want 'vectors' to satisfy.

A *vector space* is a set V with two operations defined:

1. addition.
2. scalar multiplication.

We want these two operations to satisfy the kind of algebraic properties that we are used to from vectors in $\mathbb{R}^n$.

For example, we want our vector operations to satisfy

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} \quad \text{and} \quad \alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}.$$

This leads to a list of ten properties (or axioms) that we will then take as our definition.

The scalars are members of a number system $\mathbb{F}$ called a *field* in which we have addition, subtraction, multiplication and division.

In many cases, the field will be $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{F}_2$ (the two element field).

In this subject, we will mainly concentrate on the case $\mathbb{F} = \mathbb{R}$.

However, $\mathbb{F} = \mathbb{C}$ is also very important both for theoretical considerations and applications.

Definition (Vector Space)

A *vector space* is a non-empty set V with two operations: addition and scalar multiplication.
These operations are required to satisfy the following rules.
For any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$:

Addition behaves well:

- A1 $\mathbf{u} + \mathbf{v} \in V$. (closure of vector addition)
- A2 $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$. (associativity)
- A3 $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. (commutativity)

There must be a zero and inverses:

- A4 There exists a vector $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$. (existence of zero vector)
- A5 For all $\mathbf{v} \in V$, there exists a vector $-\mathbf{v}$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$. (additive inverses)

Definition (Vector Space continued)

For all $\mathbf{u}, \mathbf{v}, \in V$ and $\alpha, \beta \in \mathbb{F}$:

Scalar multiplication behaves well:

- M1 $\alpha\mathbf{v} \in V$. (closure of scalar multiplication)
- M2 $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$. (associativity of scalar multiplication)
- M3 $1\mathbf{v} = \mathbf{v}$. (multiplication by unit scalar)

Addition and scalar multiplication combine well :

- D1 $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$. (distributivity 1)
- D2 $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$. (distributivity 2)

Remark

It follows from the axioms that for all $\mathbf{v} \in V$:

1. $0\mathbf{v} = \mathbf{0}$.
2. $(-1)\mathbf{v} = -\mathbf{v}$.

Can you prove these?

We are going to list some systems that obey these rules.

We are not going to show that the axioms hold for these systems.
If, however, you would like to get a feel for how this is done, read AR4.1.

4.2 Examples of vector spaces

1. $\mathbb{R}^n$ is a vector space with scalars $\mathbb{R}$.

After all, this was what we based our definition on! Vector spaces with $\mathbb{R}$ as the scalars are called *real vector spaces*.

2. Vector space of matrices

Denote by M_{mn}(or $M_{mn}(\mathbb{R})$, $M_{m,n}(\mathbb{R})$ or $M_{m\times n}(\mathbb{R})$) the set of all $m \times n$ matrices with real entries.
M_{mn} is a real vector space with the following familiar operations:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_21+b_{21} & a_{22}+b_{22} \end{bmatrix},$$

$$\alpha \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} \\ \alpha a_{21} & \alpha a_{22} \end{bmatrix}.$$

What is the zero vector in this vector space?

Note: Matrix multiplication is not a part of this vector space structure.

3. Vector space of polynomials

For a fixed n, denote by $\mathcal{P}_n$ (or $\mathcal{P}_n(\mathbb{R})$) the set of all polynomials with degree at most n:

$$\mathcal{P}_n = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \mid a_0, a_1, \ldots, a_n \in \mathbb{R}\}.$$

If we define vector addition and scalar multiplication by

$$(a_0 + \cdots + a_nx^n) + (b_0 + \cdots + b_nx^n) = (a_0 + b_0) + \cdots + (a_n + b_n)x^n,$$
$$\alpha(a_0 + a_1x + \cdots + a_nx^n) = (\alpha a_0) + (\alpha a_1)x + \cdots + (\alpha a_n)x^n, \ \alpha \in \mathbb{R},$$

then $\mathcal{P}_n$ is a vector space.

4. Vector space of functions

Let S be a set.

Denote by $\mathcal{F}(S, \mathbb{R})$ the set of all functions from S to $\mathbb{R}$.

Given $f, g \in \mathcal{F}(S, \mathbb{R})$ and $\alpha \in \mathbb{R}$, let $f + g$ and αf be the functions defined by

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

$$(\alpha f)(x) = \alpha \times f(x).$$

Equipped with these operations, $\mathcal{F}(S, \mathbb{R})$ is a vector space.

Remark
The '$+$' on the left of the first equation is not the same as the '$+$' on the right!

Why not?

Example
Let $f, g \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ be defined by

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \sin x \qquad and \qquad g : \mathbb{R} \to \mathbb{R}, \quad g(x) = x^2.$$

What do $f + g$ and $3f$ mean?

What is the zero vector in $\mathcal{F}(\mathbb{R}, \mathbb{R})$?

$\mathbf{0} : \mathbb{R} \to \mathbb{R}$ is the function defined by

$$\mathbf{0}(x) = 0.$$

That is, $\mathbf{0}$ is the function that maps all numbers to zero.

Complex vector spaces

A vector space that has $\mathbb{C}$ as the scalars is called a *complex vector space*.

Example

$$\mathbb{C}^2 = \{(a_1, a_2) \mid a_1, a_2 \in \mathbb{C}\}$$

with the operations

$$\begin{aligned} (a_1, a_2) + (b_1, b_2) &= (a_1 + b_1, a_2 + b_2) \\ \alpha(a_1, a_2) &= (\alpha a_1, \alpha a_2) \end{aligned} \qquad \text{(where } a_1, a_2, b_1, b_2, \alpha \in \mathbb{C}\text{)}$$

is a complex vector space.

Remark

All the examples of real vector spaces given above ($\mathbb{R}^n$, $\mathcal{P}_n(\mathbb{R})$, $\mathcal{F}(S, \mathbb{R})$) have complex analogues denoted by $\mathbb{C}^n$, $\mathcal{P}_n(\mathbb{C})$, $\mathcal{F}(S, \mathbb{C})$.

Why consider general vector spaces?

4.3 Linear Combinations, Linear Dependence [AR 4.2]

We have seen that a plane through the origin in $\mathbb{R}^3$ can be built up by starting with two vectors and taking all scalar multiples and sums.

In this way, we can build up lines, planes and their higher-dimensional versions.

Let V be a vector space with scalars $\mathbb{F}$.

Definition

A *linear combination* of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in V$ is a sum

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k,$$

where each α_i is a scalar.

Examples

1. $\mathbf{w} = (2,3)$ is a linear combination of $\mathbf{e}_1 = (1,0)$ and $\mathbf{e}_2 = (0,1)$.

2. $\mathbf{w} = (2,3)$ is a linear combination of $\mathbf{v}_1 = (1,2)$ and $\mathbf{v}_2 = (3,1)$.

3. $\mathbf{w} = (1,2,3)$ is *not* a linear combination of $\mathbf{v}_1 = (1,1,1)$ and $\mathbf{v}_2 = (0,0,1)$.

4.4 Linear Dependence [AR 4.3]

By taking all linear combinations of a given set of vectors, we can build up lines, planes, etc.

We can make this more efficient by removing any redundant vectors from the set.

Definition (Linear dependence)

A set of vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is *linearly dependent* if there are scalars $\alpha_1, \dots, \alpha_k$, at least one of which is non-zero, such that

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}.$$

A set which is not linearly dependent is called *linearly independent*.

Note that we can easily generalise this definition to allow for an infinite set S.

Remember

The zero vector $\mathbf{0} = (0, \dots, 0)$ is *not* the same as the number zero.

So, the set of vectors is linearly dependent if and only if some *non-trivial* linear combination gives the zero vector.

Theorem (AR Thm 4.3.1)

Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *(where* $k \geqslant 2$*) are linearly dependent iff one vector is a linear combination of the others.*

Proof:

$\square$

Geometrically:
Two vectors are linearly dependent iff one is a multiple of the other.

Three vectors in $\mathbb{R}^3$ are linearly dependent iff they lie in a plane containing the origin.

Examples

1. The vectors $(2, -1)$ and $(-6, 3)$ are linearly dependent.
2. $(2, -1)$ and $(4, 1)$ are not linearly dependent.

Definition (Linear independence)

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is *linearly independent* if it is not linearly dependent. Equivalently, if

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0},$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k$ are scalars, then

$$\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_k = 0$$

is the only solution.

Examples

1. The vectors $(1, 0, 0)$, $(1, 1, 0)$ and $(1, 1, 1)$ are linearly independent.

2. $(1, 1, 2), (1, -1, 2), (3, 1, 6)$ are linearly dependent.

3. $(1, 2, 3), (1, 0, 0), (0, 1, 0), (0, 1, 1)$ are linearly dependent.

To decide if vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ are linearly independent

1. Form the matrix A having the vectors as columns (we'll denote it by $A = [\mathbf{v}_1 \cdots \mathbf{v}_k]$).
2. Reduce to row-echelon form R.
3. Count the number of non-zero rows in R (this is rank(A)):

$$\mathbf{v}_1, \ldots, \mathbf{v}_k \text{ are linearly independent} \iff \operatorname{rank}(A) = k$$

Why does this method work?

It follows from what we know about the number of solutions of a linear system (slide 58).

Examples

1. $(1,0,0,0),(1,1,0,0),(1,1,1,0) \in \mathbb{R}^4$ are linearly independent.

2. $(1,1),(2,3),(1,0) \in \mathbb{R}^2$ are linearly dependent.

An important observation

If A and B are row-equivalent matrices (i.e., $A \sim B$), then the columns of A satisfy the *same linear relations* as the columns of B. In other words, two systems of equations saying that a linear combination of columns is zero have the same set of solutions.

This implies that relations between the (original) vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ can be read off the reduced row-echelon form of the matrix $[\mathbf{v}_1 \cdots \mathbf{v}_k]$.

Example

If

$$[\mathbf{v}_1 \cdots \mathbf{v}_5] \sim \begin{bmatrix} 1 & 3 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

then $\mathbf{v}_2 = 3\mathbf{v}_1$ and $\mathbf{v}_4 = 2\mathbf{v}_1 + \mathbf{v}_3$.

Example

The vectors $\mathbf{v}_1 = (1, 2, 1, 3)$, $\mathbf{v}_2 = (2, 4, 2, 6)$, $\mathbf{v}_3 = (0, -1, 3, -1)$ and $\mathbf{v}_4 = (1, 3, -2, 4)$ satisfy $\mathbf{v}_2 = 2\mathbf{v}_1$ and $\mathbf{v}_4 = \mathbf{v}_1 - \mathbf{v}_3$.

Useful Facts:

From the method above for deciding whether vectors are linearly dependent, we can derive the following.

Theorem (AR Thm 4.3.3)

If $k > n$, then any k vectors in $\mathbb{R}^n$ are linearly dependent.

Idea of proof: Because then the r.r.e. form of the matrix $[\mathbf{v}_1 \cdots \mathbf{v}_k]$ must contain columns which do not have a leading entry. So these columns will be dependent on other columns which do have a leading entry. $\square$

Theorem

Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ *are linearly independent iff the matrix* $A = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ *has* $\det(A) \neq 0$.

Idea of proof:

$$\text{linearly independent} \iff \text{rank}(A) = n \iff \det(A) \neq 0.$$

$\square$

Example

Decide whether the following vectors in $\mathbb{R}^3$ are linearly dependent or independent:

$$(1, 2, -1), \ (0, 3, 4), \ (2, 1, -6), \ (0, 0, 2).$$

If they are dependent, write one as a linear combination of the others.

Example

Are the following elements of $M_{2,2}$ linearly independent?

$$\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -2 & 1 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 0 & 4 \end{bmatrix}.$$

4.5 Subspaces of vector spaces

Certain subsets S of a vector space V have the nice property that any linear combination of vectors from S still lies in S.

Definition (Subspace)

A *subspace* of a vector space V is a subset $S \subseteq V$ that is itself a vector space (using the operations from V).

The following theorem allows us to avoid checking all 10 axioms.

Theorem (Subspace Theorem, AR Thm 4.2.1)

Let V be a vector space. A subset $W \subseteq V$ is a subspace of V iff

0. *W is non-empty.*
1. *W is closed under vector addition.*
2. *W is closed under scalar multiplication.*

Examples

1. The xy-plane $S = \{(x, y, z) \in \mathbb{R}^3 \mid z = 0\}$ is a subspace of $\mathbb{R}^3$.

2. The plane S in $\mathbb{R}^3$ through the origin and parallel to the vectors $\mathbf{u} = (1, 1, 1)$ and $\mathbf{v} = (1, 2, 3)$ is a subspace of $\mathbb{R}^3$.

3. The set $H = \{(x, y) \in \mathbb{R}^2 \mid x \geqslant 0, y \geqslant 0\}$ is not a subspace of $\mathbb{R}^2$.

4. *Any* line or plane in $\mathbb{R}^3$ that contains the origin is a subspace.

Example

Let $V = M_{2,2}$, the vector space of real 2×2 matrices, and $H \subset V$ be those matrices in V whose trace is equal to 0. The 'trace' is defined to be the sum of the diagonal entries, so

$$H = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a + d = 0 \right\}.$$

Show that H is a subspace of V.

Another example

Let

$$V = \mathcal{P}_2 = \{a_0 + a_1x + a_2x^2 \mid a_0, a_1, a_2 \in \mathbb{R}\}$$

and

$$W = \{a_0 + a_1x + a_2x^2 \mid a_1a_2 \geqslant 0\} \subseteq V.$$

Is W a subspace of V?

More examples

1. $\{\mathbf{0}\}$ is always a subspace of V.
2. V is always a subspace of V.
3. The set of diagonal matrices $\left\{ \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \middle| a, b, c \in \mathbb{R} \right\}$ is a subspace of $M_{3,3}$.

4. The subset of continuous functions $\{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$.

5. $S = \{2 \times 2 \text{ matrices with determinant equal to } 0\}$
$= \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| ad - bc = 0 \right\}$ is not a subspace of $M_{2,2}$.

6. $\{f : [0,1] \to \mathbb{R} \mid f(0) = 2\}$ is not a subspace of $\mathcal{F}([0,1], \mathbb{R})$.

Useful Fact

A subspace W of V must necessarily contain the zero vector $\mathbf{0} \in V$.

So the line $y = 2x + 1$ is *not* a subspace of $\mathbb{R}^2$.

In fact, a line or plane in $\mathbb{R}^3$ is a subspace if and only if it contains the origin.

An important class of examples of subspaces is illustrated by the following.

Example

The set of all solutions to the homogeneous linear system

$$\left\{\begin{aligned} x + y + z &= 0 \\ x - y - z &= 0 \end{aligned}\right\}$$

is a subspace of $\mathbb{R}^3$.

Remember that a linear system is called *homogeneous* if *all* the constant terms are equal to zero.

In general, we have the following:

Proposition (AR Thm 4.2.4)

The set of solutions of a system of m homogeneous linear equations in n variables is a subspace of $\mathbb{R}^n$.

Proof:

We check the 3 conditions in the definition of subspace:

4.6 Spanning sets

Generating a subspace

Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a set of vectors in a vector space V defined over $\mathbb{F}$.

Definition

The *span* of the set $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is the set of all linear combinations of vectors from S

$$\text{Span}(S) = \{\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k \mid \mathbf{v}_1, \dots, \mathbf{v}_k \in S \text{ and } \alpha_1, \dots, \alpha_k \in \mathbb{F}\}.$$

It is sometimes also denoted by $\langle \mathbf{v}_1, \dots, \mathbf{v}_k \rangle$.

Examples

1. In $\mathbb{R}^2$, $\text{Span}\{(3,2)\}$ is the line through the origin in the direction given by $(3,2)$, i.e., the line $y = \frac{2}{3}x$.

2. In $\mathbb{R}^3$, $\text{Span}\{(1,1,1),(3,2,1)\}$ is the plane $x - 2y + z = 0$.

3. In $\mathbb{R}^3$, $\text{Span}\{(1,1,1),(3,3,3)\}$ is the line $x = y = z$.

Lemma (AR Thm 4.2.3)

1. $\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ *is a subspace of* V.
2. *Any subspace of* V *that contains the vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *contains* $\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$.

Proof:

$\square$

Remark

The subspace spanned by a set of vectors is the 'smallest' subspace that contains those vectors.

More examples

In $\mathbb{R}^3$:

1. $\mathrm{Span}\{(1,1,0)\}$ is the line through the origin containing the point $(1,1,0)$.
2. $\mathrm{Span}\{(1,0,0),(1,1,0)\}$ is the xy-plane.
3. $\mathrm{Span}\{(1,0,0),(-3,7,0),(1,1,0)\}$ is the xy-plane.

4. Span$\{(1,0,0),(2,3,-4),(1,1,0)\}$ is the whole of $\mathbb{R}^3$.

Spanning sets

Definition (Spanning set)

Let V be a vector space. The vectors $\mathbf{v}_1,\ldots,\mathbf{v}_k \in V$ are said to *span V* if Span$\{\mathbf{v}_1,\ldots,\mathbf{v}_k\} = V$. The set $\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$ is a *spanning set* for V.

Equivalently, V contains all the vectors $\mathbf{v}_1,\ldots,\mathbf{v}_k$ and all vectors in V are linear combinations of the vectors $\mathbf{v}_1,\ldots,\mathbf{v}_k$.

Examples

1. $(1,0)$ and $(1,1)$ span $\mathbb{R}^2$.

2. $(1,1,2)$ and $(1,0,1)$ do not span $\mathbb{R}^3$.

3. $(1,1,2)$, $(1,0,1)$ and $(2,1,3)$ do not span $\mathbb{R}^3$.

4. In $\mathcal{P}_2$ the element $p(x) = 2 + 2x + 5x^2$ is a linear combination of $p_1(x) = 1 + x + x^2$ and $p_2(x) = x^2$, but $q(x) = 1 + 2x + 3x^2$ is not. So $\{p_1, p_2\}$ is not a spanning set for $\mathcal{P}_2$.

Example

Show that the following vectors span $\mathbb{R}^4$:

$$\{(1,0,-1,0),(1,1,1,1),(3,0,0,0),(4,1,-3,-1)\}.$$

We need to show that for *any* vector $(a,b,c,d) \in \mathbb{R}^4$, the equation

$$x(1,0,-1,0) + y(1,1,1,1) + z(3,0,0,0) + w(4,1,-3,-1) = (a,b,c,d)$$

has a solution.

Writing this linear system in matrix form gives

$$\underbrace{\begin{bmatrix} 1 & 1 & 3 & 4 \\ 0 & 1 & 0 & 1 \\ -1 & 1 & 0 & -3 \\ 0 & 1 & 0 & -1 \end{bmatrix}}_{A} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$

This has augmented matrix

$$\left[\begin{array}{cccc|c} 1 & 1 & 3 & 4 & a \\ 0 & 1 & 0 & 1 & b \\ -1 & 1 & 0 & -3 & c \\ 0 & 1 & 0 & -1 & d \end{array}\right].$$

We know that this linear system is consistent *for all possible values of a, b, c, d* if and only if $\operatorname{rank}(A) = 4$...

which is the case in this example.

To decide if $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ span $\mathbb{R}^n$

1. Form the matrix $A = [\mathbf{v}_1 \cdots \mathbf{v}_k]$ having columns given by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$.
2. Calculate $\operatorname{rank}(A)$ as before:
   a. Reduce to row-echelon form
   b. Count the number of non-zero rows

Then,

$$\mathbf{v}_1, \ldots, \mathbf{v}_k \quad \text{span} \quad \mathbb{R}^n \iff \operatorname{rank}(A) = n.$$

Why does this method work?

Since $A = [\mathbf{v}_1 \cdots \mathbf{v}_k]$ has k columns, we know that $\operatorname{rank}(A) \leqslant k$.

It follows that:

Proposition

If $k < n$, then $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ does not span $\mathbb{R}^n$.

4.7 Bases and dimension [AR 4.4–4.5, 4.7]

In general, spanning sets can be (too) big.

For example, $(1,0), (0,1), (2,3), (-7,4)$ span $\mathbb{R}^2$, but the last two are not needed since they can be expressed as linear combinations of the first two. The vectors are not linearly independent.

Bases

Definition (Basis)

A *basis* for a vector space V is a set of vectors from V which:

1. spans V.
2. is linearly independent.

Examples

1. $\{(1,0),(0,1)\}$ is a basis for $\mathbb{R}^2$.

2. $\{(2,0),(-1,1)\}$ is a basis for $\mathbb{R}^2$.

3. $\{(2,-1,-1),(1,2,-3)\}$ is a basis for the plane $x+y+z=0$ in $\mathbb{R}^3$.

4. $\{(2,3,7)\}$ is a basis for the line in $\mathbb{R}^3$ given by $\dfrac{x}{2}=\dfrac{y}{3}=\dfrac{z}{7}$.

Note

A vector space V can have many bases. For example, any two vectors in $\mathbb{R}^2$ which are not collinear will form a basis of $\mathbb{R}^2$.

Notation/example:

The vectors

$$\mathbf{e}_1=(1,0,0,\dots,0),\mathbf{e}_2=(0,1,0,\dots,0),\dots,\mathbf{e}_n=(0,0,\dots,0,1)$$

form a basis of $\mathbb{R}^n$, called the *standard basis.*

The following is an important and very useful theorem about bases:

Theorem (AR Thms 4.5.1, 4.5.2)

Let V be a vector space and let $\{v_1, \ldots, v_k\}$ be a basis for V.

1. *A subset of V with more than k vectors is linearly dependent.*
2. *A subset of V with fewer than k vectors does not span V.*
3. *Any two bases of V have the same number of elements.*

Proof: First prove for $V \subset \mathbb{R}^n$ using what we already know about (homogeneous) linear systems. Then generalise to general V.
Note that the theorem can be generalised to V having a basis with infinitely many elements

So, in particular, every basis of $\mathbb{R}^n$ has exactly n elements.

Dimension

The above theorem tells us that although there can be many different bases for the same space, they will all have the same number of vectors.

Definition (Dimension)

The *dimension* of a vector space V is the number of vectors in a basis for V. This is denoted by $\dim(V)$. We call V *finite-dimensional* if it admits a finite basis, and *infinite-dimensional* otherwise.

Examples

1. The dimension of $\mathbb{R}^2$ is
2. $\mathbb{R}^n$ has dimension
3. The line $\{(\alpha, \alpha, \alpha) \mid \alpha \in \mathbb{R}\} \subseteq \mathbb{R}^3$ has dimension
4. The plane $\{(x, y, z) \mid x + y + z = 0\} \subseteq \mathbb{R}^3$ has dimension
5. $\{1, x, x^2, \ldots, x^n\}$ is a basis for $\mathcal{P}_n$. So $\dim(P_n(\mathbb{R})) =$.
6. $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis for $M_{2\times 2}$, so
7. $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$ is a basis for the subspace S of 2×2 matrices with trace equal to zero, so $\dim S =$.

In the special case where $V = \{\mathbf{0}\}$ it is convenient to say that $\dim(V) = 0$.

An infinite-dimensional example

Let $\mathcal{P}$ be the set of *all* polynomials:

$$\mathcal{P} = \{a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n \mid n \in \mathbb{N}, a_0, a_1, \ldots, a_n \in \mathbb{R}\}.$$

Then $\mathcal{P}$ is a vector space and the set $\mathcal{B} = \{1, x, x^2, \ldots\}$ is a basis for $\mathcal{P}$.

So $\mathcal{P}$ is an infinite-dimensional vector space.

Can you see why $\mathcal{B}$ is a basis?

Examples

1. The polynomials

$$\{2 + x + x^2, \quad 1 + x, \quad -1 - 7x^2, \quad x - x^2\}$$

   are linearly dependent, since

2. The matrices

$$\left\{ \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}, \quad \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 6 & 7 \\ 4 & 5 \end{bmatrix} \right\}$$

   do not span $M_{2\times 2}$ since

Standard Bases

It is useful to fix names for certain bases.

- The *standard basis for $\mathbb{R}^n$* is

$$\{(1, 0, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, 0, 0, \dots, 0, 1)\}.$$

  The dimension of $\mathbb{R}^n$ is n.

- The *standard basis for $M_{m,n}$* is

$$\left\{ \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \right\}.$$

  The dimension of $M_{m,n}$ is mn.

- The *standard basis for $\mathcal{P}_n$* is

$$\left\{1, x, x^2, \ldots, x^n\right\}.$$

The dimension of $\mathcal{P}_n$ is $n+1$.

Calculating bases for subspaces of $\mathbb{R}^n$

How can we find a basis for a given subspace of $\mathbb{R}^n$?

The method will depend on how the subspace is described.
Often we know (or can easily calculate) a spanning set.

Example
Find a basis for the subspace

$$S = \{(a+2b, b, -a+b, a+3b) \mid a, b \in \mathbb{R}\} \subseteq \mathbb{R}^4.$$

What is the dimension of S?

Another example

Find a basis for the subspace of $\mathbb{R}^4$ spanned by the vectors $(1, -1, 2, 1)$, $(-2, 2, -4, -2)$, $(1, 0, 3, 0)$.

To find a basis for the span of a set of vectors

To find a basis for the subspace V spanned by vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$:

1. Form the matrix $A = [\mathbf{v}_1 \cdots \mathbf{v}_k]$.
2. Reduce to row-echelon form B.
3. The columns of A corresponding to the steps (leading entries) in B give a basis for V.

Why does this method work?

Recall that if $A \sim B$, then the columns of A and B satisfy the same linear relationships. The question then reduces to one about matrices in reduced row-echelon form.

Remarks

- Don't forget that it is the columns of A (not its row-echelon form) that we use for the basis.
- This method gives a basis that is a subset of the original set of vectors. Later we will give a second method which gives a basis of different, but usually simpler, vectors

Example

Let $S = \{(1, -1, 2, 1), (0, 1, 1, -2), (1, -3, 0, 5)\}$.
Find a subset of S that is a basis for $\langle S \rangle$.

Column space of a matrix

Definition (Column space)

Let A be an $m \times n$ matrix. The subspace of $\mathbb{R}^m$ spanned by the columns of A is called the *column space* of A.

Suppose $A \sim B$. In general, the column space of B is *not* equal to the column space of A, although they do have the same dimension.

Example

$$A = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & -3 \\ 2 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix} = B.$$

The column space of A is

The column space of B is

Are they equal?

What is the dimension of each subspace?

Suppose that $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subseteq \mathbb{R}^n$ is a set of vectors.

Remember that we saw a method for obtaining a basis for $\mathrm{Span}(S)$ that started with the matrix $[\mathbf{v}_1 \cdots \mathbf{v}_k]$ having the vectors of S as columns.

This method gave us a basis of $\mathrm{Span}(S)$ that is a subset of S.

So, if V is a subspace of $\mathbb{R}^n$:

Every spanning set for V contains a basis for V.

It also follows from the above method that:

Every linearly independent subset of V extends to a basis for V.

More generally:

Theorem

Let V be a vector space.

1. *Any spanning set for V contains a basis for V.*
2. *Any linearly independent set in V can be extended to a basis of V.*

Theorem

Let V be a vector space, and suppose that $\dim(V) = m$.

1. *If a spanning set of V has exactly m elements, then it is a basis.*
2. *If a linearly independent subset of V has exactly m elements, then it is a basis.*

Proof: We first prove the theorem for V a subspace of $\mathbb{R}^n$ using the methods above. Then we generalise.

Example

Given that

$$\{(1, -\pi, \sqrt{2}), (23, 1, 100), (\frac{3}{2}, 7, 1)\}$$

is a spanning set for $\mathbb{R}^3$, it is a basis for $\mathbb{R}^3$.

Another method for finding a basis: The 'row method'

We first illustrate by repeating a previous example, but this time using a matrix that has the vectors as rows (not columns).

Example
Let $S = \{(1,-1,2,1),(8,1,1,-2),(4,5,-7,6)\}$
Find a basis for $\langle S \rangle$.

Write the vectors as rows in a matrix:

$$\begin{bmatrix} 1 & -1 & 2 & 1 \\ 8 & 1 & 1 & -2 \\ 4 & 5 & -7 & -6 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 9 & -15 & -10 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

It follows that a basis for $\langle S \rangle$ is

Can you see why?

To find a basis for the span of a set of vectors: the 'row method'

To find a basis for the subspace V spanned by vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$.

1. Form the matrix A whose rows are the given vectors.
2. Reduce to row-echelon form B.
3. The non-zero rows of B are a basis for V.

Note: this basis will usually not be a *subset* of $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$.

To explain why this works, let's introduce some more terminology.

Row space of a matrix

Definition (Row Space)

Let A be a $m \times n$ matrix. The subspace of $\mathbb{R}^n$ spanned by the rows of A is called the *row space* of A.

Suppose $A \sim B$. The above method relies on the fact that (unlike the situation with the column space):

The row space of B *is always* equal to the row space of A.

Why? Can you prove it?

Example

$$A = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & -3 \\ 2 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix} = B.$$

The row space of A is Span

The row space of B is Span

Note

For any matrix A:

$$\mathrm{rank}(A) = \dim(\text{row space of } A) = \dim(\text{column space of } A).$$

This is because the number of non-zero rows in the row-echelon form is equal to the number of leading entries.

Example

Let

$$A = \begin{bmatrix} 1 & -1 & 2 & -2 \\ 2 & 0 & 1 & 0 \\ 5 & -3 & 7 & -6 \\ 1 & 1 & -1 & 3 \end{bmatrix}.$$

Find a basis and the dimension for:

1. the column space of A.
2. the row space of A.

$$A = \begin{bmatrix} 1 & -1 & 2 & -2 \\ 2 & 0 & 1 & 0 \\ 5 & -3 & 7 & -6 \\ 1 & 1 & -1 & 3 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & -1 & 2 & -2 \\ 0 & 2 & -3 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So...

Solution space of a matrix

We have seen how to find a basis for a subspace given a spanning set for the subspace.

Another way in which subspaces frequently arise is as the set of solutions to a homogenous linear system. This can be written in matrix form

$$A\mathbf{x} = \mathbf{0}.$$

We saw previously that such a solution set is always a subspace of $\mathbb{R}^n$ where n is the number of variables.

Definition (Solution Space)

Let A be a $m \times n$ matrix. The subspace of $\mathbb{R}^n$ given by $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$ is called the *solution space* or *nullspace* of A.

Bases for solution spaces

How can we find a basis for the solution space?

We first illustrate with an example.

Example

Find a basis for the subspace of $\mathbb{R}^4$ defined by the equations

$$\left\{\begin{aligned} x_1 + x_3 + x_4 &= 0 \\ 3x_1 + 2x_2 + 5x_3 + x_4 &= 0 \\ x_2 + x_3 - x_4 &= 0 \end{aligned}\right\}.$$

Finding a basis for the solution space

To find a basis for the solution space of a system of homogeneous linear equations:

1. Write the system in matrix form $A\mathbf{x} = \mathbf{0}$.
2. Reduce A to row-echelon form B.
3. Solve the system $B\mathbf{x} = \mathbf{0}$ as usual, and write the solution in the form

$$\mathbf{x} = t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k,$$

where $t_1, \ldots, t_k \in \mathbb{R}$ are parameters and $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$.

Then $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a basis for the solution space.

Why does this work?

The set $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a spanning set for the solution space, since every solution can be written as a linear combination of the vectors in S.

The fact that the vectors are linearly independent results from the way in which they were defined.

Suppose we chose the parameters to be the variables whose corresponding column in B had no leading entry. Say $t_i = x_{m_i}$.

Then the m_j coordinate of $\mathbf{v}_i$ is 1 if $i = j$ and 0 if $i \neq j$.

It follows that the $\mathbf{v}_i$ are linearly independent.

Another example

Find a basis for the subspace of $\mathbb{R}^4$ given by

$$V = \{(x_1, x_2, x_3, x_4) \mid x_1 + 2x_2 + x_3 + x_4 = 0, 3x_1 + 6x_2 + 4x_3 + x_4 = 0\}.$$

The matrix A of coefficients is

$$A = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 3 & 6 & 4 & 1 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix}.$$

The solutions are given by

$$(x_1, x_2, x_3, x_4) = t_1(\quad , \quad , \quad , \quad) + t_2(\quad , \quad , \quad , \quad) \qquad t_1, t_2 \in \mathbb{R}.$$

So a basis for V is:

Rank-nullity theorem

Given a $m \times n$ matrix A there are three associated subspaces:

1. The *row space* of A is the subspace of $\mathbb{R}^n$ spanned by the rows. Its dimension is equal to $\operatorname{rank}(A)$.

2. The *column space* of A is the subspace of $\mathbb{R}^m$ spanned by the columns. Its dimension is equal to $\operatorname{rank}(A)$.

3. The *solution space* of A is the subspace of $\mathbb{R}^n$ given by

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}\}.$$

   The solution space is also called the *nullspace* of A.
   Its dimension is called the *nullity* of A, denoted by $\operatorname{nullity}(A)$.

We have seen techniques to find bases for the row space, column space and solution space of a matrix.

As a result of these techniques we observe the following:

Theorem (cf. AR Thm 4.8.2)

Suppose A is an $m \times n$ matrix. Then

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n = \text{number of columns in } A.$$

Given what we know about finding the rank and the solution space of a matrix, this is simply the statement that every column in the row-echelon form either contains a leading entry or doesn't contain a leading entry.

Note

If you are asked to find the solution space, the column space and the row space of A, you only need to find the reduced row-echelon form of A once.

Remember

If A and B are row equivalent matrices, then:

- the row space of A is equal to the row space of B.
- the solution space of A is equal to the solution space of B.
- the column space if A is *not* necessarily equal to the column space of B.
- the columns of A satisfy the same linear relations as the columns of B.
- the dimension of the column space of A is equal to the dimension of the column space of B.

4.8 Coordinates relative to a basis

Lemma

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for a vector space V, then every vector $\mathbf{v} \in V$ can be written uniquely in the form

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \ldots \alpha_n \mathbf{v}_n$$

, where $\alpha_1, \ldots, \alpha_n$ are scalars.

Proof:

$\square$

Definition (Coordinates)

Suppose $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an *ordered* basis for a vector space V. For $\mathbf{v} \in V$ we can write

$$\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k \quad \text{for some scalars} \quad \alpha_1, \ldots, \alpha_k \in \mathbb{R}.$$

The scalars $\alpha_1, \ldots, \alpha_k$ are uniquely determined by $\mathbf{v}$ and are called the *coordinates of* $\mathbf{v}$ *relative to* $\mathcal{B}$.

The column matrix

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix}$$

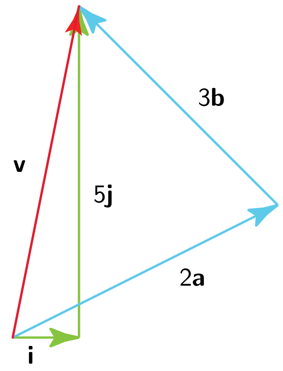
is called the *coordinate matrix of v with respect to* $\mathcal{B}$.

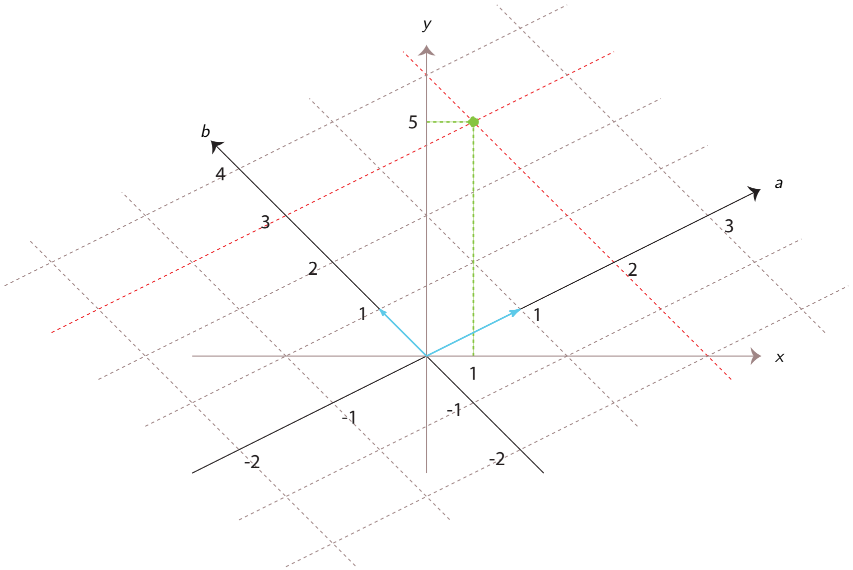
Examples

1. If we consider $\mathbb{R}^2$ with the standard basis $\mathcal{B} = \{\mathbf{i}, \mathbf{j}\}$, then the vector $\mathbf{v} = (1,5)$ has coordinates $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$.

2. Now consider $\mathbb{R}^2$ with basis $\mathcal{B}' = \{\mathbf{a}, \mathbf{b}\}$, where $\mathbf{a} = (2,1)$ and $\mathbf{b} = (-1,1)$.

   Then $\mathbf{v} = (1,5)$ has coordinates $[\mathbf{v}]_{\mathcal{B}'} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.





The point $(1,5)$ shown on a grid formed by a and b.

3. Consider the subspace $V = \{(x,y,z) \in \mathbb{R}^3 \mid x+y+z=0\}$ with basis $\mathcal{B} = \{(1,-1,0),(1,0,-1)\}$. Then $\mathbf{v} = (1,2,-3) \in V$ has coordinates $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$.

4. In $\mathcal{P}_2$ with basis $\mathcal{B} = \{1, x, x^2\}$, the polynomial $p = 2 + 7x - 9x^2$ has coordinates $[p]_{\mathcal{B}} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$.

5. In $M_{2\times 2}$ with basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$,

   the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ has coordinates $[A]_{\mathcal{B}} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$.

Having fixed a basis, there is a one-to-one correspondence between vectors and their coordinate matrices.

If $\mathbf{u}, \mathbf{v}$ are vectors in a vector space V, and α is a scalar, then

$$\begin{aligned}[\mathbf{u}+\mathbf{v}]_{\mathcal{B}} &= [\mathbf{u}]_{\mathcal{B}} + [\mathbf{v}]_{\mathcal{B}},\\ [\alpha\mathbf{v}]_{\mathcal{B}} &= \alpha[\mathbf{v}]_{\mathcal{B}}.\end{aligned}$$

A consequence of this is that vectors are linearly independent if and only if their coordinate matrices are (using *any* basis).
(Exercise: Prove this!)
Coordinates can often be used to reduce problems about general vector spaces to problems in $\mathbb{R}^n$.

For example, let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be vectors in a vector space V with basis $\mathcal{B}$. Then,

$$\mathbf{v}_1, \ldots, \mathbf{v}_k \text{ are linearly dependent}$$

if and only if their coordinate matrices $[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_k]_{\mathcal{B}}$ are linearly dependent.

Topic 5: Inner Product Spaces [AR Chap 6]

5.1 Definition of inner products

The *Euclidean length* of a vector in $\mathbf{v} \in \mathbb{R}^n$ is defined as

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

The *angle* θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ was defined by

$$\cos\theta = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{v}\|\|\mathbf{u}\|} \qquad \text{with} \quad 0 \leqslant \theta \leqslant \pi.$$

The *projection of* $\mathbf{v}$ *onto* $\mathbf{u}$ was given in terms of the dot product by

$$\text{proj}_{\mathbf{u}}\,\mathbf{v} = \frac{(\mathbf{u} \cdot \mathbf{v})}{\|\mathbf{u}\|^2}\mathbf{u}.$$

We will generalise key properties of the dot product to define an inner product.

Definition (Inner Product)

Let V be a vector space over the real numbers. An *inner product* on V is a function that associates with every pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a real number, denoted by $\langle \mathbf{u}, \mathbf{v} \rangle$, satisfying the following properties.

For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\alpha \in \mathbb{R}$:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.
2. $\alpha\langle \mathbf{u}, \mathbf{v} \rangle = \langle \alpha\mathbf{u}, \mathbf{v} \rangle$.
3. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$.
4. a. $\langle \mathbf{u}, \mathbf{u} \rangle \geqslant 0$
   b. $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \Rightarrow \mathbf{u} = \mathbf{0}$

A real vector space V together with an inner product is called a *real inner product space*.

For $V = \mathbb{R}^n$, $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$ defines an inner product.

But this is not the only inner product on the vector space $\mathbb{R}^n$.

Example

Show that, in $\mathbb{R}^2$, if $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ then

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + 2u_2 v_2$$

defines an inner product.

More examples

1. Show that in $\mathbb{R}^3$, $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 - u_2 v_2 + u_3 v_3$ does not define an inner product by showing that axiom 4a does not hold.

2. Show that in $\mathbb{R}^2$,

$$\langle \mathbf{u}, \mathbf{v} \rangle = [\mathbf{u}]^T \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} [\mathbf{v}] = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

defines an inner product.

Another example

In the case of the vector space $\mathcal{P}_n$ of polynomials of degree less than or equal to n, one possible choice of inner product is

$$\langle p, q \rangle = \int_0^1 p(x)q(x)\,dx.$$

To verify this, we would need to check, for any polynomials $p(x), q(x), r(x)$ and any scalar $\alpha \in \mathbb{R}$, that

1. $\int_0^1 p(x)q(x)\,dx = \int_0^1 q(x)p(x)\,dx$.
2. $\alpha \int_0^1 p(x)q(x)\,dx = \int_0^1 (\alpha p(x))q(x)\,dx$.
3. $\int_0^1 (p(x) + q(x))\,r(x)\,dx = \int_0^1 p(x)r(x)\,dx + \int_0^1 q(x)r(x)\,dx$.
4. $\int_0^1 p(x)^2\,dx \geqslant 0$.
5. $\int_0^1 p(x)^2\,dx = 0$ only when $p(x)$ is the zero polynomial.

Complex vector spaces

Over the complex vector space $\mathbb{C}^n$, an inner product is defined slightly differently.

Definition (Hermitian product on $\mathbb{C}^n$)

For complex scalars, define the *Hermitian product* of $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ by

$$\mathbf{u} \cdot \mathbf{v} = u_1\overline{v_1} + u_2\overline{v_2} + \cdots + u_n\overline{v_n},$$

where a bar indicates complex conjugation.

The conjugation is necessary to have real (and positive) lengths:

$$\mathbf{v} \cdot \mathbf{v} = v_1\overline{v_1} + v_2\overline{v_2} + \cdots + v_n\overline{v_n} = |v_1|^2 + |v_2|^2 + \cdots + |v_n|^2 \geqslant 0.$$

Example

Let $\mathbf{u} = (1+i, 1-i)$ and $\mathbf{v} = (i, 1)$ then $\mathbf{u} \cdot \mathbf{v} =$

We adjust only property 1 of a real inner product to get a Hermitian inner product on a general complex vector space.

Definition (Inner Product over $\mathbb{C}$)

A Hermitian *inner product* on a complex vector space V is a function that associates with every pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a complex number, denoted by $\langle \mathbf{u}, \mathbf{v} \rangle$, satisfying the following properties.

For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\alpha \in \mathbb{C}$:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$.
2. $\alpha\langle \mathbf{u}, \mathbf{v} \rangle = \langle \alpha\mathbf{u}, \mathbf{v} \rangle$.
3. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$.
4. a. $\langle \mathbf{u}, \mathbf{u} \rangle \geqslant 0$.
   b. $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \Rightarrow \mathbf{u} = \mathbf{0}$.

The inner product on $\mathcal{P}_n(\mathbb{R})$ defined on Slide 204 is similarly generalised to the complex polynomials in $\mathcal{P}_n(\mathbb{C})$ simply by replacing the second polynomial in the definition by its complex conjugate.

5.2 Geometry from inner products

In a general vector space, how can we "measure angles" and "find lengths"?

If we fix an inner product on the vector space, we can then *define* length and angle using the same equations that we saw above for $\mathbb{R}^n$. We simply replace the dot product by the chosen inner product.

Definition (Length)

For a real or complex vector space with an inner product $\langle \cdot, \cdot \rangle$, define the *length* (or *norm*) of a vector $\mathbf{v}$ by $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

Definition (Angle)

For a real vector space with an inner product $\langle \cdot, \cdot \rangle$, we define:

- the *angle* θ between $\mathbf{v}$ and $\mathbf{u}$ using

$$\cos\theta = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{v}\|\|\mathbf{u}\|}, \qquad 0 \leqslant \theta \leqslant \pi.$$

In the complex case, we replace $\langle \mathbf{v}, \mathbf{u} \rangle$ in this formula by its real part.

Note

In order for these definitions of angle to make sense, we need

$$-1 \leqslant \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{v}\|\|\mathbf{u}\|} \leqslant 1 \quad \text{or} \quad -1 \leqslant \frac{\mathrm{Re}\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{v}\|\|\mathbf{u}\|} \leqslant 1.$$

We will see shortly that this is always the case.

Definition (Orthogonal vectors)

Two vectors $\mathbf{v}$ and $\mathbf{u}$ are said to be *orthogonal* if $\langle \mathbf{v}, \mathbf{u} \rangle = 0$

We present several examples of inner products.

Example

$\langle (u_1, u_2), (v_1, v_2) \rangle = u_1 v_1 + 2u_2 v_2$ defines an inner product on $\mathbb{R}^2$.

If $\mathbf{u} = (3, 1)$ and $\mathbf{v} = (-2, 3)$, then

$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \|(5, -2)\| = \sqrt{\langle (5, -2), (5, -2) \rangle} = \sqrt{25 + 8} = \sqrt{33}$

and

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle (3, 1), (-2, 3) \rangle = 3 \times (-2) + 2 \times 1 \times 3 = 0,$$

so $\mathbf{u}$ and $\mathbf{v}$ are orthogonal (using this inner product).

Example (an inner product for functions)

The set of all real-valued continuous functions on $[a, b]$

$$C[a, b] = \{f : [a, b] \to \mathbb{R} \mid f \text{ is continuous}\}$$

is a vector space.

It's a subspace of $\mathcal{F}([a, b], \mathbb{R})$, the vector space of all functions.

For $f, g \in C[a, b]$, define

$$\langle f, g \rangle = \int_a^b f(x) g(x) dx.$$

This is an inner product.

Example

Consider $C[0, 2\pi]$ with the inner product $\langle f, g \rangle = \int_0^{2\pi} f(x)g(x)dx$.

The norms of the functions $s(x) = \sin(x)$ and $c(x) = \cos(x)$ are

$$\begin{aligned} \|s\|^2 = \langle s, s \rangle = \int_0^{2\pi} \sin^2(x)dx &= \int_0^{2\pi} \frac{1}{2}(1 - \cos(2x))dx \\ &= \left[\frac{x}{2} - \frac{1}{4}\sin(2x) \right]_0^{2\pi} = \pi. \end{aligned}$$

So, $\|s\| = \sqrt{\pi}$ and we similarly find that $\|c\| = \sqrt{\pi}$. Now,

$$\langle s, c \rangle = \int_0^{2\pi} \sin(x)\cos(x)dx = \int_0^{2\pi} \frac{1}{2}\sin(2x)dx = \left[-\frac{1}{4}\cos(2x) \right]_0^{2\pi} = 0.$$

$\sin(x)$ and $\cos(x)$ are thus orthogonal

This is used in the study of periodic functions (using 'Fourier series') in (for example) signal analysis, speech recognition, music recording etc.

5.3 Cauchy-Schwarz inequality

Theorem (Cauchy-Schwarz inequality)

Let V be an inner product space. Then, for all $\mathbf{u}, \mathbf{v} \in V$,

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leqslant \|\mathbf{u}\| \|\mathbf{v}\|.$$

Equality holds if and only if one vector is a multiple of the other.

Proof: The proof is as for the case $V = \mathbb{R}^n$ on slide 82. $\square$

We defined the angle between two vectors using

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|} \quad \text{(real)} \qquad \cos\theta = \frac{\mathrm{Re}\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|} \quad \text{(complex)}.$$

Since $|\mathrm{Re}\langle \mathbf{u}, \mathbf{v} \rangle| \leqslant |\langle \mathbf{u}, \mathbf{v} \rangle|$, the Cauchy-Schwarz inequality gives

$$|\cos\theta| \leqslant 1.$$

This implies that the definition of angle is okay!

Application

For two continuous functions $f, g \colon [a, b] \to \mathbb{R}$, it follows directly from the Cauchy-Schwarz inequality that

$$\left(\int_a^b f(x)g(x)dx\right)^2 \leqslant \left(\int_a^b f(x)^2 dx\right)\left(\int_a^b g(x)^2 dx\right).$$

And we don't need to do any (more) calculus to prove it!

Set $f(x) = \sqrt{x}$ and $g(x) = 1/\sqrt{x}$; also $a = 1, b = t > 1$.

We obtain

$$\left(\int_1^t 1\, dx\right)^2 \leqslant \left(\int_1^t x\, dx\right)\left(\int_1^t \frac{1}{x}\, dx\right)$$

which becomes

$$(t-1)^2 \leqslant \left(\frac{t^2-1}{2}\right)\ln t \quad \text{or} \quad \ln t \geqslant \frac{2(t-1)}{t+1}, \quad t \geqslant 1.$$

5.4 Orthogonality and the Gram-Schmidt algorithm

Recall that $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Definition (Orthogonal set of vectors)

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is called *orthogonal* if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$.

Examples

1. $\{(1,0,0), (0,1,0), (0,0,1)\}$ is orthogonal in $\mathbb{R}^3$ with the dot product as inner product.
2. So is $\{(1,1,1), (1,-1,0), (1,1,-2)\}$ using the dot product.
3. $\{\sin(x), \cos(x)\}$ is orthogonal in $C[0, 2\pi]$ equipped with the inner product defined (using a definite integral) on slide 211.

Proposition

Every orthogonal set of non-zero vectors is linearly independent.

Proof:

$\square$

Definition

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is called *orthonormal* if it is orthogonal and each vector has length one. That is

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \quad \text{is orthonormal} \quad \iff \quad \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases}$$

Note

Any orthogonal set of non-zero vectors can be made orthonormal by dividing each vector by its length.

Examples

1. In $\mathbb{R}^3$ with the dot product:
   - $\{(1,0,0), (0,1,0), (0,0,1)\}$ is orthonormal.
   - $\{(1,1,1), (1,-1,0), (1,1,-2)\}$ is not (though it is orthogonal).
   - $\{\frac{1}{\sqrt{3}}(1,1,1), \frac{1}{\sqrt{2}}(1,-1,0), \frac{1}{\sqrt{6}}(1,1,-2)\}$ is orthonormal.

2. In $C[0,2\pi]$ with the inner product

$$\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)dx :$$

- the set $\{\sin(x), \cos(x)\}$ is orthogonal but not orthonormal.
- the set $\{\frac{1}{\sqrt{\pi}}\sin(x), \frac{1}{\sqrt{\pi}}\cos(x)\}$ is orthonormal.
- The (infinite) set

$$\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\sin(x), \frac{1}{\sqrt{\pi}}\cos(x), \frac{1}{\sqrt{\pi}}\sin(2x), \frac{1}{\sqrt{\pi}}\cos(2x), \dots\}$$

is orthonormal.

Orthonormal bases

Bases that are orthonormal are particularly convenient to work with. For example, we have the following.

Lemma

If $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis for V and $\mathbf{x} \in V$, then

$$\mathbf{x} = \langle \mathbf{x}, \mathbf{v}_1\rangle \mathbf{v}_1 + \cdots + \langle \mathbf{x}, \mathbf{v}_n\rangle \mathbf{v}_n.$$

Proof: Exercise! □

Gram-Schmidt procedure

Say we have a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ for a vector space V. How can we convert this to an orthonormal basis for V? Here is the method:

Step (1) $\mathbf{u}_1 =$

Step (2) $\mathbf{w}_2 =$

and $\mathbf{u}_2 =$

Step (3) $\mathbf{w}_3 =$

$\mathbf{u}_3 =$

$\vdots$

Step (k) $\mathbf{w}_k =$

$\mathbf{u}_k =$

Gram-Schmidt procedure

Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for V.

1. Let $\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1$.

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

2b. $\mathbf{u}_2 = \frac{1}{\|\mathbf{w}_2\|}\mathbf{w}_2$.

3a. $\mathbf{w}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{v}_3, \mathbf{u}_2 \rangle \mathbf{u}_2$.

3b. $\mathbf{u}_3 = \frac{1}{\|\mathbf{w}_3\|}\mathbf{w}_3$.

$\vdots$ $\quad\vdots$

ka. $\mathbf{w}_k = \mathbf{v}_k - \langle \mathbf{v}_k, \mathbf{u}_1 \rangle \mathbf{u}_1 - \cdots - \langle \mathbf{v}_k, \mathbf{u}_{k-1} \rangle \mathbf{u}_{k-1}$.

kb. $\mathbf{u}_k = \frac{1}{\|\mathbf{w}_k\|}\mathbf{w}_k$.

Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal basis for V.

Note: This method will also produce an orthonormal basis starting from any *spanning set*: just discard any zero vectors $\mathbf{w}_i$ and continue.

Example

Find an orthonormal basis for the subspace W of $\mathbb{R}^4$ (with dot product) spanned by

$$\{(1,1,1,1),(2,4,2,4),(1,5,-1,3)\}.$$

Answer

$$\left\{\frac{1}{2}(1,1,1,1), \frac{1}{2}(-1,1,-1,1), \frac{1}{2}(1,1,-1,-1)\right\}.$$

5.5 Orthogonal projection

Orthogonal projection [AR 6.2]

Let V be a real vector space with inner product $\langle \cdot, \cdot \rangle$.
Let $\mathbf{u} \in V$ be a unit vector (i.e., $\|\mathbf{u}\| = 1$).

Definition

The *(orthogonal) projection* of $\mathbf{v}$ onto $\mathbf{u}$ is

$$\mathbf{p} = \langle \mathbf{v}, \mathbf{u} \rangle \mathbf{u}.$$

Note

- $\mathbf{p} = \|\mathbf{v}\| \cos\theta \, \mathbf{u}$.
- $\mathbf{v} - \mathbf{p}$ is orthogonal to $\mathbf{u}$.

Example

Consider $\mathcal{P}_2$ with the inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$.
The orthogonal projection of $\mathbf{v} = 1 + 2x + 3x^2$ onto $\mathbf{u} = \sqrt{3}\,x$ is

$$\mathbf{p} = \langle \mathbf{v}, \mathbf{u} \rangle \mathbf{u} = 3x \int_0^1 (x + 2x^2 + 3x^3)dx = \frac{23}{4}x.$$

More generally, we can project onto a subspace W of V as follows.

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthonormal basis for W.

Definition

The *(orthogonal) projection* of $\mathbf{v} \in V$ onto the subspace W is

$$\operatorname{proj}_W(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{v}, \mathbf{u}_k \rangle \mathbf{u}_k.$$

Properties of $\mathbf{p} = \operatorname{proj}_W(\mathbf{v})$

- $\mathbf{p} \in W$.
- if $\mathbf{w} \in W$, then $\operatorname{proj}_W(\mathbf{w}) = \mathbf{w}$ (by the lemma on Slide 22).
- $\mathbf{v} - \mathbf{p}$ is orthogonal to W (i.e., orthogonal to every vector in W).
- $\mathbf{p}$ is the vector in W that is closest to $\mathbf{v}$.
  That is, for all $\mathbf{w} \in W$, $\|\mathbf{v} - \mathbf{w}\| \geqslant \|\mathbf{v} - \mathbf{p}\|$.
- $\mathbf{p}$ does not depend on the choice of orthonormal basis for W.

Example

Let $W = \{(x, y, z) \mid x + y + z = 0\}$ in $V = \mathbb{R}^3$ with the dot product.
The set

$$\{\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1, -1, 0),\ \mathbf{u}_2 = \frac{1}{\sqrt{6}}(1, 1, -2)\}$$

is an orthonormal basis for W.

For $\mathbf{v} = (1, 2, 3)$, we have

$$\begin{aligned} \text{proj}_W(\mathbf{v}) &= \langle \mathbf{v}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{v}, \mathbf{u}_2 \rangle \mathbf{u}_2 \\ &= \\ &= \end{aligned}$$

Note that $\mathbf{v} - \text{proj}_W(\mathbf{v}) =$ is orthogonal to W.

Example

Find the point in the subspace W of $\mathbb{R}^4$ spanned by

$$\{(1, 1, 1, 1), (2, 4, 2, 4), (1, 5, -1, 3)\}$$

closest to $\mathbf{v} = (2, 2, 1, 3)$. (Use dot product.)

Answer

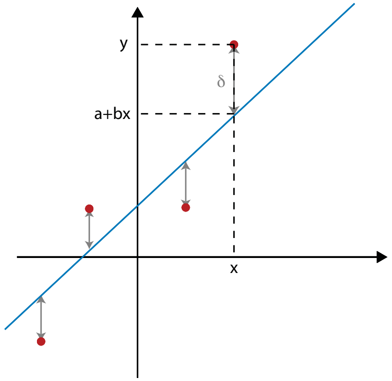
$\frac{1}{2}(3, 5, 3, 5)$.

5.6 Application: Curve Fitting (AR 6.4)

Problem: Given a set of data points (x_1, y_1), (x_2, y_2),..., (x_n, y_n), we want to find the straight line $y = a + bx$ which best approximates the data.

A common approach is to minimise the *least squares error* E

$$\begin{aligned} E^2 &= \text{sum of the squares of vertical errors} \\ &= \sum_{i=1}^{n} \delta_i^2 \\ &= \sum_{i=1}^{n} (y_i - (a + bx_i))^2. \end{aligned}$$



So given (x_1, y_1),..., (x_n, y_n), we want to find $a, b \in \mathbb{R}$ which minimise

$$\sum_{i=1}^{n} (y_i - (a + bx_i))^2.$$

This can be written as

$$\|\mathbf{y} - A\mathbf{u}\|^2,$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

To minimise $\|\mathbf{y} - A\mathbf{u}\|$, we want $A\mathbf{u}$ to be as close as possible to $\mathbf{y}$. We can use projection to find the closest point.

We seek the vector $A\mathbf{u}$ in

$$W = \{A\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^2\} \qquad (= \text{the column space of } A)$$

that is closest to $\mathbf{y}$.

This closest vector is precisely $\operatorname{proj}_W \mathbf{y}$.

So to find $\mathbf{u}$ we *could* project $\mathbf{y}$ to W to get $A\mathbf{u} = \operatorname{proj}_W \mathbf{y}$

and from this calculate $\mathbf{u}$.

However, we can calculate $\mathbf{u}$ directly (without finding an orthonormal basis for W) by noting that

$$\begin{array}{rrl}
 & \langle \mathbf{w}, \mathbf{y} - \operatorname{proj}_W \mathbf{y}\rangle = 0 & \text{for all } \mathbf{w} \in W \\
\Longrightarrow & \langle A\mathbf{v}, \mathbf{y} - A\mathbf{u}\rangle = 0 & \text{for all } \mathbf{v} \in \mathbb{R}^2 \\
\Longrightarrow & (A\mathbf{v})^T(\mathbf{y} - A\mathbf{u}) = 0 & \text{for all } \mathbf{v} \in \mathbb{R}^2 \\
\Longrightarrow & \mathbf{v}^T A^T(\mathbf{y} - A\mathbf{u}) = 0 & \text{for all } \mathbf{v} \in \mathbb{R}^2 \\
\Longrightarrow & A^T(\mathbf{y} - A\mathbf{u}) = \mathbf{0} & \\
\Longrightarrow & A^T\mathbf{y} - A^TA\mathbf{u} = \mathbf{0} & \\
\Longrightarrow & A^TA\mathbf{u} = A^T\mathbf{y}. &
\end{array}$$

From this we can calculate $\mathbf{u}$, given that we know A and $\mathbf{y}$.

Summary

Given data points (x_1, y_1), (x_2, y_2),..., (x_n, y_n), to find the straight line $y = a + bx$ of best fit, we find

$$\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$$

satisfying the "normal equation"

$$A^T A\mathbf{u} = A^T \mathbf{y}, \qquad (*)$$

$$\text{where } \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \text{ and } A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

If $A^T A$ is invertible (and it usually is), the solution to $(*)$ is given by

$$\mathbf{u} = (A^T A)^{-1} A^T \mathbf{y}.$$

Example

Find the straight line which best fits the data points $(-1, 1)$, $(1, 1)$, $(2, 3)$.

Answer

Line of best fit is: $y = \frac{9}{7} + \frac{4}{7}x$.

Extension

The same method works for finding quadratic fitting curves.
To find the quadratic $y = a + bx + cx^2$ which best fits data (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$, we take

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

and solve

$$A^T A\mathbf{u} = A^T\mathbf{y}$$

for

$$\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

Topic 6: Eigenvalues and Eigenvectors [AR Chapt 5]

6.1 Definition of eigenvalues and eigenvectors

The topic of eigenvalues and eigenvectors is fundamental to many applications of linear algebra. These include quantum mechanics in physics, image compression and reconstruction in computing and engineering, the analysis of high-dimensional data in statistics and algorithms for searching the web.

The key idea is to look for *non-zero* vectors $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^n$ that are *stretched* by a matrix $A \colon \mathbb{R}^n \to \mathbb{R}^n$.

In other words, $A(\mathbf{v}) = \lambda\mathbf{v}$ for some scalar λ.

Then $\mathrm{Span}\{\mathbf{v}\}$ is a 1-dimensional subspace of $\mathbb{R}^n$ that is preserved by A.

- If $\lambda = 0$, then A sends $\mathbf{v}$ to 0.
- If λ is real and positive, then A rescales $\mathbf{v}$.
- If λ is real and negative, then A also reverses the direction of $\mathbf{v}$.
- However, λ may be complex (see Slide 246).

In fact, the idea of eigenvalues and eigenvectors can be applied directly to square matrices.

Definition

Let A be an $n \times n$ matrix and let λ be a scalar. Then, a non-zero $n \times 1$ column matrix $\mathbf{v}$ with the property that

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

is called an *eigenvector*, while λ is called the *eigenvalue*.

To develop some geometric intuition, it is handy to think of A as being the standard matrix of a linear transformation.

Example

Consider the matrix $\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$ as the standard matrix of a linear transformation. What is the effect of this transformation on the vectors $(2, 1)$ and $(2, -1)$?

6.2 Finding eigenvalues

If I denotes the $n \times n$ identity matrix, the defining equation

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

for eigenvalues and eigenvectors can be rewritten in the form

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

The values of λ for which this equation has *non-zero solutions* are precisely the eigenvalues.

Theorem

The homogeneous linear system $(A - \lambda I)\mathbf{v} = \mathbf{0}$ has a non-zero solution if and only if $\det(A - \lambda I) = 0$. Consequently, the eigenvalues of A are the values of λ for which

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

Notation

The equation $\det(\lambda I - A) = 0$ is called the *characteristic equation*. From our study of determinants, we know that $\det(\lambda I - A)$ is a polynomial of degree n in λ. It is called the *characteristic polynomial*.

Example

Find the eigenvalues of $\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$.

Example

Find the eigenvalues of $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

If we use the standard basis, how can the corresponding linear transformation on $\mathbb{R}^2$ be described geometrically? How does this tell you that the matrix does not have any real eigenvalues?

Examples like this one show that it can be useful to consider a real matrix as a linear transformation on a *complex* vector space.

Example

Find the eigenvalues of the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$.

Note that this matrix has only three eigenvalues (over $\mathbb{C}$). However, one of these eigenvalues is a *repeated* root of the characteristic polynomial. We say that this eigenvalue has *multiplicity* 2.

6.3 Finding eigenvectors

If λ is an eigenvalue of an $n \times n$ matrix A, then $\{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = \lambda\mathbf{v}\}$ is a subspace of $\mathbb{R}^n$ (exercise!) called the *eigenspace* of λ.

To find the eigenvectors of a matrix

- For each eigenvalue λ, solve the homogeneous linear system $(A - \lambda I)\mathbf{v} = \mathbf{0}$.
- Use row reduction as usual.
- Note that $\operatorname{rank}(A - \lambda I) < n$, so you always obtain at least one row of zeros.

Note that, unlike eigenvalues, the eigenvectors of a matrix are not unique. e.g., any eigenvector $\mathbf{v}$ can always be multiplied by any non-zero scalar and it will still be an eigenvector.

Example

For each of the eigenvalues $3, -1$ of the matrix $\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$, find a corresponding eigenvector.

Example

For each of the eigenvalues $\lambda = -1, 8$ of the matrix $A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$, find a basis for the corresponding eigenspace.

The eigenvalue -1 has multiplicity 2, so it is quite satisfying that it corresponds to two linearly independent eigenvectors.

Example

Find the eigenvalues and a basis of eigenvectors for the matrix
$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$.

Note that in this case, we have an eigenvalue of multiplicity 2 but only one eigenvector. This may be unsatisfying but it is a fact of life.

Complex eigenvalues

Recall that any real root λ of $\det(A - \lambda I)$ gives rise to an eigenvector v which is stretched by a factor of λ, i.e. it satisfies $Av = \lambda v$.
What is the geometric interpretation of a complex root λ?

Example

The matrix that rotates $\mathbb{R}^2$ anticlockwise by angle θ is

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta. \end{bmatrix}$$

Its eigenvalues are $e^{i\theta}$ and $e^{-i\theta}$ which are not real if $\theta \neq n\pi$. R_θ thus has no real eigenvectors (though it does have complex ones).

This accords with geometry: R_θ does not rescale any subspace of $\mathbb{R}^2$.

In general, any complex eigenvalue $\lambda = |\lambda|e^{\pm i\theta}$ of an $n \times n$ real matrix gives rise to a plane in $\mathbb{R}^n$ that is rotated by θ and scaled by $|\lambda|$.

Theorem

A square matrix A is invertible if and only if 0 is not an eigenvalue of A.

Proof:
Let the characteristic polynomial of A be

$$p(\lambda) = \det(\lambda I - A) = a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n.$$

Then,

$$\begin{aligned} 0 \text{ is an eigenvalue of } A &\iff p(0) = 0 \\ &\iff \det(0I - A) = 0 \\ &\iff \det(A) = 0 \\ &\iff A \text{ is not invertible.} \end{aligned}$$

$\square$

6.4 Diagonalisation [AR 5.2]

We now take up the problem of studying the question of when the eigenvectors of an $n \times n$ matrix A form a basis for $\mathbb{R}^n$. Such bases are extremely important in the applications of eigenvalues and eigenvectors mentioned earlier.

Definition

A square matrix A is said to be *diagonalisable* if there is an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix. The matrix P is said to *diagonalise* A.

To test if A is diagonalisable, the following theorem can be used.

Theorem

An $n \times n$ matrix A is diagonalisable if and only if there is a basis of $\mathbb{R}^n$ (or $\mathbb{C}^n$) all of whose elements are eigenvectors of A.

Idea of proof:

If we can form a basis in which all the basis vectors are eigenvectors of T, then the new matrix for T will be diagonal.

$\square$

A simple case in which A is diagonalisable is when A has *n distinct eigenvalues*. This follows from the theorem above and the following

Lemma

Eigenvectors corresponding to distinct eigenvalues are linearly independent.

Idea of proof:

$\square$

Example

Decide whether or not the following matrices are diagonalisable.

$$A = \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

How to diagonalise a matrix

Suppose A is diagonalisable. How can we find an invertible matrix P and a diagonal matrix D with $D = P^{-1}AP$?

Theorem

Let A be a diagonalisable $n \times n$ matrix. Choose any basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of $\mathbb{R}^n$ (or $\mathbb{C}^n$) whose elements are eigenvectors of A. If λ_i is the eigenvalue of the eigenvector $\mathbf{v}_i$ and $P = \big[[\mathbf{v}_1] \cdots [\mathbf{v}_n]\big]$, then

$$P^{-1}AP = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

So once we have found a basis consisting of eigenvectors, we can write down the matrices P and D without further calculation.

Example

Check that $\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ are eigenvectors of the matrix

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$$

and read off the corresponding eigenvalues.

Write down an invertible matrix P such that $D = P^{-1}AP$ is diagonal, and write down the diagonal matrix D.

Check your answer by evaluating both sides of the equation $AP = PD$.

Orthogonal matrices

Because a matrix often represents a physical system, it can be important that a change-of-basis transformation does not affect shape. This will happen when the change-of-basis matrix is orthogonal.

Definition

A real $n \times n$ matrix P is *orthogonal* if the columns of P form an *orthonormal* basis of $\mathbb{R}^n$ (with respect to the dot product).

Examples

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ and } \frac{1}{\sqrt{6}} \begin{bmatrix} -\sqrt{3} & 1 & \sqrt{2} \\ 0 & -2 & \sqrt{2} \\ \sqrt{3} & \sqrt{6} & \sqrt{2} \end{bmatrix} \text{ are orthogonal,}$$

but $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ are not.

Orthogonal matrices have some good properties:

Theorem

If P is an orthogonal $n \times n$ matrix and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then

- $P^{-1} = P^T$.
- $\|P\mathbf{u}\| = \|\mathbf{u}\|$.
- $\langle P\mathbf{u}, P\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$ *(where $\langle \mathbf{u}, \mathbf{v}\rangle$ is the dot product on $\mathbb{R}^n$).*

We can summarise this by saying that orthogonal matrices are square matrices that *preserve* lengths and angles.

Geometrically, orthogonal matrices represent rotations, reflections and rotations composed with reflections.

The proof of the first property above also shows that a square matrix P is orthogonal if and only if $P^TP = I$.

It now follows from properties 1 and 3 on Slide 68 that the determinant of an orthogonal matrix is either 1 or -1. Indeed,

$$\det(P)^2 = \det(P^T)\det(P) = \det(P^TP) = \det(I) = 1.$$

Real symmetric matrices

Definition

An $n \times n$ matrix A is *symmetric* if $A^T = A$.

Examples

$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ and $\begin{bmatrix} 3 & -1 & 4 \\ -1 & 1 & 5 \\ 4 & 5 & 9 \end{bmatrix}$ are symmetric,

but $\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$ and $\begin{bmatrix} 3 & -1 & 4 \\ -1 & 1 & 5 \\ 4 & 6 & 9 \end{bmatrix}$ are not.

Symmetric matrices arise (for example) in:

- quadratic functions.
- inner products.
- maxima and minima of multivariable functions.

Theorem

Let A be an $n \times n$ real symmetric matrix. Then:

1. *all eigenvalues of A are real;*
2. *eigenvectors from distinct eigenvalues are orthogonal;*
3. *A is diagonalisable;*
   (In fact, there is an orthonormal basis of eigenvectors.)
4. *we can write $A = QDQ^{-1}$ where D is diagonal and Q is orthogonal.*

Note that, in this case the diagonalisation formula can be written $A = QDQ^T$.

Idea of proof of 2:

$\square$

General method to orthogonally diagonalise real, symmetric A

1. Find the eigenvalues λ_i and any basis of eigenvectors for A.
2. Find an orthonormal basis for each eigenspace, using Gram-Schmidt if there is a repeated eigenvalue.
3. Let $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and let Q be the matrix whose columns are the orthonormal basis vectors of step 2, ordered so that the i-th column gives an eigenvector of eigenvalue λ_i. Then,

$$Q^T A Q = D \text{ and } Q^{-1} = Q^T.$$

Example

Find matrices D and Q as above for $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$.

Unitary matrices

We now mention the complex analogues of the orthogonal and real symmetric matrices.

Definition

A complex $n \times n$ matrix U is *unitary* if the columns of U form an *orthonormal* basis of $\mathbb{C}^n$ (with respect to the Hermitian product).

Examples

- Both $\begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$ and $\frac{1}{\sqrt{2}}\begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}$ are unitary.
- Any (real) orthogonal matrix is unitary if viewed as a complex matrix.

Definition

Let A be a complex $m \times n$ matrix. We define its *conjugate transpose* (or *Hermitian transpose* or *adjoint*) $\overline{A}^T$ to be the $n \times m$ matrix whose entries are

$$(\overline{A}^T)_{ij} = \overline{A_{ji}}.$$

Other common notations for the conjugate transpose of A are $A^\dagger$ or A^*.

Theorem

If U is a unitary $n \times n$ matrix and $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$, then

- $U^{-1} = \overline{U}^T$, *i.e.* $\overline{U}^T U = I$.
- $\|U\mathbf{u}\| = \|\mathbf{u}\|$.
- $\langle U\mathbf{u}, U\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$ *(where $\langle \mathbf{u}, \mathbf{v}\rangle$ is the Hermitian product on $\mathbb{C}^n$).*
- $|\det(U)| = 1$.

Hermitian (self-adjoint) matrices

The complex analogue of a real symmetric matrix is as follows.

Definition

A matrix A is *Hermitian* or *self-adjoint* if $\overline{A}^T = A$.

Examples

- Both $\begin{bmatrix} 1 & 2i \\ -2i & 3 \end{bmatrix}$ and $\begin{bmatrix} 0 & -i & 4 \\ i & 1 & 5+2i \\ 4 & 5-2i & 9 \end{bmatrix}$ are Hermitian.
- $\begin{bmatrix} i & 0 \\ 0 & i \end{bmatrix}$ is not Hermitian: its *conjugate* transpose is $\begin{bmatrix} -i & 0 \\ 0 & -i \end{bmatrix}$.
- Any (real) symmetric matrix is Hermitian if viewed as a complex matrix.

Hermitian matrices generalise symmetric ones in a very nice way.

Theorem

Let A be an $n \times n$ Hermitian matrix. Then:

1. *all eigenvalues of A are real;*
2. *eigenvectors from distinct eigenvalues are orthogonal;*
3. *A is diagonalisable;*
   (In fact, there is an orthonormal basis of eigenvectors.)
4. *we can write $A = UDU^{-1}$ where D is diagonal and U is unitary.*

The proofs are identical to the symmetric case.

In fact, the standard proof of 1 in the symmetric case amounts to treating the given real symmetric matrix as a complex Hermitian one and showing that its eigenvalues coincide with their complex conjugates.

Theorem

Let A be an $n \times n$ Hermitian matrix. Then the assignment

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

defines a Hermitian inner product on $\mathbb{C}^n$ if and only if every eigenvalue λ_i of A satisfies $\lambda_i > 0$.

Restricting to real symmetric A gives a real inner product.

Idea of Proof The symmetry and bilinearity properties follow in a straightforward way, regardless of the eigenvalues.
Thus, we must show that $\langle \cdot, \cdot \rangle$ is positive-definite if and only if the eigenvalues are positive. This is clear via orthogonal diagonalisation.

Powers of a matrix

In applications, one often comes across the need to apply a transformation many times. If the transformation can be represented by diagonalisable matrix A, then it is easy to compute A^k and thus the action of the k-th application of the transformation.

The first point to appreciate is that computing powers of a diagonal matrix D is easy.

Example
With $D = \mathrm{diag}(1, -3, 2)$, write down D^2, D^3 and D^k.

In general, we have

Lemma

Suppose A is diagonalisable, so that $D = P^{-1}AP$ is diagonal for some P. Then for all integers $k \geqslant 1$, we have

$$A^k = PD^kP^{-1}.$$

Example

For the matrix of slide 239 we can write

$$A = PDP^{-1} \quad \text{with} \quad A = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}, \quad P = \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix}.$$

Thus,

$$A^k = PD^kP^{-1}, \qquad \text{where} \qquad P^{-1} =$$

Explicitly,

$$A^k =$$

Example *Population Movement between Victoria and Queensland*

The population of Victoria is 6 million and the population of Queensland is 4.8 million. Assume 2% of Victorians move to Queensland each year, 1% of Queenslanders move to Victoria each year and everybody else stays put.
Let x_i be the Victorian population (in millions) after i years and y_i be the Queensland population (in millions) after i years and $p_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$.

1. Find a matrix A such that $p_{i+1} = Ap_i$.
2. Find p_1, p_2 and p_n in terms of p_0 and A.
3. Check that A has eigenvalues $0.97, 1$ with corresponding eigenvectors $(-1, 1)$ and $(1, 2)$.
4. Find matrices P and D such that $A = PDP^{-1}$.
5. Use P and D to estimate the population p_n in terms of p_0.
6. What happens in the long run (as $n \to \infty$) to the populations of Victoria and Queensland?

Conic Sections [AR 7.3]

Consider equations in x and y of the form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0,$$

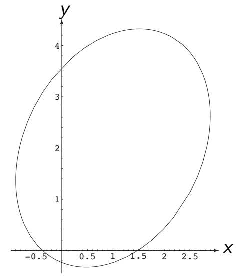
where $a, b, c, d, e, f \in \mathbb{R}$ are constants.

The graphs of such equations are called *conic sections* or *conics*.

Example

Plot the curve given by the equation:

$$9x^2 - 4xy + 6y^2 - 10x - 20y - 5 = 0.$$



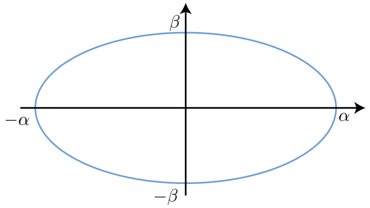
We will see how the shape of this graph can be calculated using diagonalisation.

We shall assume that d and e are zero.
See [AR 7.3] for a discussion of how to reduce to this case.

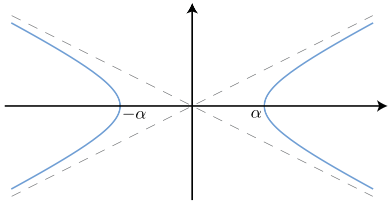
If the equation is simple enough, we can identify the conic by inspection.

Standard central conics:

ellipse $\quad \dfrac{x^2}{\alpha^2} + \dfrac{y^2}{\beta^2} = 1.$



hyperbola $\quad \dfrac{x^2}{\alpha^2} - \dfrac{y^2}{\beta^2} = 1.$



The dashed lines are $y = \pm\frac{\beta}{\alpha}x$.

The equation in matrix form

Consider the curve defined by the equation

$$ax^2 + bxy + cy^2 = k.$$

The equation can be written in matrix form as

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = k.$$

That is,

$$\mathbf{x}^T A\mathbf{x} = k,$$

where $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and A is a real symmetric matrix.

We can diagonalise A in order to simplify the equation so that we can identify the curve.
Let's demonstrate with an example.

Identify and sketch the conic defined by $x^2 + 4xy + y^2 = 1$.

This can be written as $\mathbf{x}^T A\mathbf{x} = 1$, where $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$.

Diagonalising gives $A = QDQ^T$ with $D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$, $Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

Let $\mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix}$ be the co-ordinates of (x, y) relative to the orthonormal basis of eigenvectors $\mathcal{B} = \{(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})\}$

Then $\mathbf{x} = Q\mathbf{x}'$ (Q is precisely the transition matrix $P_{\mathcal{S},\mathcal{B}}$) and the equation of the conic can be rewritten as

$$\begin{aligned} \mathbf{x}^T A\mathbf{x} = 1 &\iff (Q\mathbf{x}')^T\, QDQ^T\, Q\mathbf{x}' = 1 \\ &\iff (\mathbf{x}')^T Q^T\, QDQ^T\, Q\mathbf{x}' = 1 \\ &\iff (\mathbf{x}')^T D\mathbf{x}' = 1 \\ &\iff 3(x')^2 - (y')^2 = 1. \end{aligned}$$

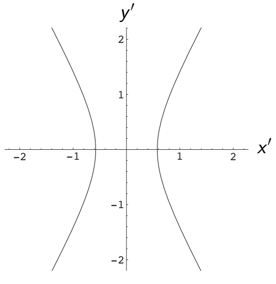
So the curve is a hyperbola.

The x'-axis and the y'-axis are called the *principal axes* of the conic.

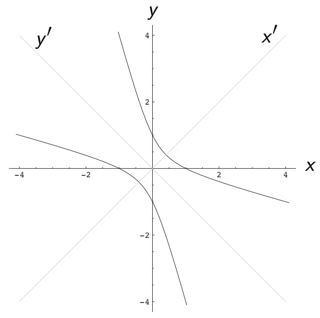
The directions of the principal axes are given by the eigenvectors.

In this example, the directions of the principal axes are

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \quad \text{and} \quad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$



$3(x')^2 - (y')^2 = 1$



$x^2 + 4xy + y^2 = 1$

Summary

- We can represent the equation of a conic (centered at the origin) by a matrix equation $x^T Ax = k$ with A symmetric.
- The eigenvectors of A will be parallel to the principal axes of the conic.
- So if Q represents the change of basis matrix, then Q will be orthogonal and $Q^T AQ = D$ will be diagonal.
- If $\mathbf{x} = Q\mathbf{x}'$, then $\mathbf{x}'$ represents the coordinates with respect to the new basis and the equation of the conic with respect to this basis is $\mathbf{x}'^T D\mathbf{x}' = k$ or $\lambda_1(x')^2 + \lambda_2(y')^2 = k$.
- The conic can now be identified and sketched.

Example (a quadric surface)

The equation

$$-x^2 + 2y^2 + 2z^2 + 4xy + 4xz - 2yz = 1$$

represents a 'quadric surface' in $\mathbb{R}^3$.

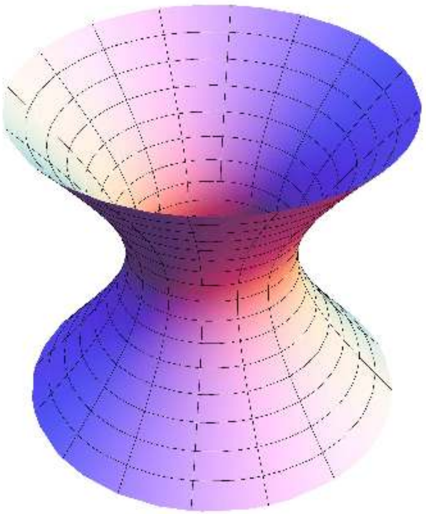
In matrix form, it can be represented as $\mathbf{x}^T A\mathbf{x} = 1$ with

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} -1 & 2 & 2 \\ 2 & 2 & -1 \\ 2 & -1 & 2 \end{bmatrix}.$$

The eigenvalues of A are $3, 3, -3$. So the equation of the surface with respect to an orthonormal basis of eigenvectors is

$$3X^2 + 3Y^2 - 3Z^2 = 1$$

The surface is a 'hyperboloid of one sheet'; see the sketch below.
(You are not expected to identify quadric surfaces in three dimensions.)



Topic 7: Linear Transformations [AR 8.1]

7.1 Linear transformations between general vector spaces

We now turn to thinking about *functions*, also known as *maps*, *mappings* or *transformations*, between vector spaces.

Recall that a function $T : V \to W$ is defined by:

1. a set V (the *domain* of T).
2. a set W (the *codomain* or *target* of T).
3. a rule assigning a value $T(x)$ in W to every x in V.

Some functions $T\colon V \to W$ between vector spaces V and W satisfy

$$T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_1 T(\mathbf{v}_1) + \alpha_2 T(\mathbf{v}_2) + \cdots + \alpha_n T(\mathbf{v}_n).$$

In other words, such T 'preserve linear combinations of vectors'.

Functions like these will be called *linear transformations* and will be important for many of the applications of linear algebra.

Definition (Linear transformation)

Let V and W be vector spaces (over the same field of scalars).

A *linear transformation* from V to W is a map $T\colon V \to W$ such that for each $\mathbf{u}, \mathbf{v} \in V$ and for each scalar α:

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$. ($T$ preserves addition)
2. $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$. ($T$ preserves scalar multiplication)

Loosely speaking, linear transformations are those maps between vector spaces that *'preserve the vector space structure'*.

Examples

Note: Geometric interpretation.

Let $T: V \to W$ be a linear transformation. Then

1. T maps the zero vector to the zero vector.

2. T maps each line through the origin to a line through the origin (or to just the origin).

3. T maps each parallelogram to a parallelogram (or a line segment or point).

7.2 Linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$

We will start by looking at some geometric transformations of $\mathbb{R}^2$.

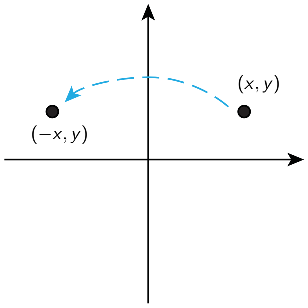
A vector in $\mathbb{R}^2$ is an ordered pair (x, y), with $x, y \in \mathbb{R}$.

To describe the effect of a transformation we will use coordinate matrices. With respect to the standard basis $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$, the vector (x, y) has coordinate matrix $\begin{bmatrix} x \\ y \end{bmatrix}$.

Example

Reflection across y-axis.

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -x \\ y \end{bmatrix}.$$



A common feature of all linear transformations is that they can be represented by a matrix. In the example above,

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ y \end{bmatrix}.$$

The matrix

$$A_T = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

is called the *standard matrix representation of the transformation T*.
Notice that A_T has eigenvalues ± 1.

Why is T a linear transformation?

Examples of (geometric) linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$

1. Reflection across the x-axis has matrix $\begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}$.

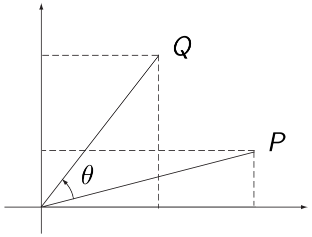
2. Reflection in the line $y = 5x$ has matrix $\begin{bmatrix} -\frac{12}{13} & \frac{5}{13} \\ \frac{5}{13} & \frac{12}{13} \end{bmatrix}$.

Notice that the matrices in 1 and 2 have eigenvalues ± 1.

3. Rotation around the origin anticlockwise by an angle of $\frac{\pi}{2}$ has matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

4. Rotation around the origin anticlockwise by an angle of θ has matrix $\begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}$.

We need to work out the coordinates of the point Q obtained by rotating P.



Examples continued

5. Compression/expansion along the x-axis has matrix $\begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}$.

6. Shear along the x-axis has matrix $\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}$.

   These are best thought of as mappings on a rectangle.

   For example, a shear along the x-axis corresponds to the mapping

.

6'. Shear along the y-axis has matrix $\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$.

Successive Transformations

Example

Find the image of (x, y) after a shear along the x-axis with $c = 1$ followed by a compression along the y-axis with $c = \frac{1}{2}$.

Solution:

Let $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the compression and denote its standard matrix representation by A_R. Similarly let $S\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the shear and denote its standard matrix representation by A_S. Then the coordinate matrix of $R(S(x, y))$ is given by

$$A_R A_S \begin{bmatrix} x \\ y \end{bmatrix}.$$

It remains to recall A_R and A_S, and to compute the matrix products.

Note

1. The composition of two linear transformations $T(\mathbf{v}) = R(S(\mathbf{v}))$ is also written $T(\mathbf{v}) = (R \circ S)(\mathbf{v})$, i.e. $T = RS = R \circ S$.
2. The matrix for the linear transformation S followed by the linear transformation R is the matrix product $A_R A_S$.
   In other words,
   $$A_{R \circ S} = A_R A_S.$$
   (This is why matrix multiplication is defined the way it is!)
3. Notice that (reading left to right) the two matrices are in the *opposite order* to the order in which the transformations are applied.

7.3 Linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

An example of a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$ is

$$T(x_1, x_2, x_3) = (x_2 - 2x_3, 3x_1 + x_3).$$

To prove that this is a linear transformation, we must show that for any

$$\mathbf{u}, \mathbf{v} \in \mathbb{R}^3 \quad \text{and} \quad \alpha \in \mathbb{R},$$

we have

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \text{and} \quad T(\alpha\mathbf{u}) = \alpha T(\mathbf{u}).$$

Proof Let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$. First we note that

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3).$$

Applying T to this gives

$$T(\mathbf{u} + \mathbf{v}) = ((u_2 + v_2) - 2(u_3 + v_3), 3(u_1 + v_1) + (u_3 + v_3)).$$

Rearranging the right-hand side gives

$$\begin{aligned} &((u_2 + v_2) - 2(u_3 + v_3), 3(u_1 + v_1) + (u_3 + v_3)) \\ &\quad = (u_2 - 2u_3, 3u_1 + u_3) + (v_2 - 2v_3, 3v_1 + v_3) = T(\mathbf{u}) + T(\mathbf{v}). \end{aligned}$$

For the second part,

$$\begin{aligned} T(\alpha\mathbf{u}) &= T((\alpha u_1, \alpha u_2, \alpha u_3)) = (\alpha u_2 - 2\alpha u_3, 3\alpha u_1 + \alpha u_3) \\ &= \alpha(u_2 - 2u_3, 3u_1 + u_3) = \alpha T(\mathbf{u}). \end{aligned}$$

$\square$

We can also write T in matrix form.

Using coordinate matrices, we have

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_2 - 2x_3 \\ 3x_1 + x_3 \end{bmatrix} =$$

Thus, $[T(\mathbf{v})] = A_T[\mathbf{v}]$, where $[\mathbf{v}]$ is the coordinate matrix of $\mathbf{v}$, $[T(\mathbf{v})]$ is the coordinate matrix of $T(\mathbf{v})$

and $A_T =$

A_T is called the *standard matrix representation for T*.

In fact, all linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be represented by matrices.

Theorem

Every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ has a standard matrix representation A_T specified by

$$A_T = [\,[T(\mathbf{e}_1)] \quad [T(\mathbf{e}_2)] \quad \cdots \quad [T(\mathbf{e}_n)]\,],$$

where each $[T(\mathbf{e}_i)]$ denotes the coordinate matrix of $T(\mathbf{e}_i)$ and $\mathcal{S} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$. Then

$$[T\mathbf{v}] = A_T[\mathbf{v}]$$

for all $\mathbf{v} \in \mathbb{R}^n$.

Note

- The matrix A_T has size $m \times n$.
- Alternative notations for A_T include: $[T]$ or $[T]_{\mathcal{S}}$ or $[T]_{\mathcal{S},\mathcal{S}}$.

Proof

Let $\mathbf{v} \in \mathbb{R}^n$ and write

$$\mathbf{v} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \cdots + \alpha_n \mathbf{e}_n.$$

The coordinate matrix of $\mathbf{v}$ with respect to the standard basis is then

$$[\mathbf{v}] = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

We seek an $m \times n$ matrix A_T such that

$$[T(\mathbf{v})] = A_T[\mathbf{v}] \qquad \text{(independently of } v\text{).} \qquad (\dagger)$$

In words, this equation says that A_T times the column matrix $[\mathbf{v}]$ is equal to the coordinate matrix of $T(\mathbf{v})$.

Since T is linear, we have

$$T(\mathbf{v}) = \alpha_1 T(\mathbf{e}_1) + \alpha_2 T(\mathbf{e}_2) + \cdots + \alpha_n T(\mathbf{e}_n).$$

It follows from this that

$$[T(\mathbf{v})] = \alpha_1[T(\mathbf{e}_1)] + \alpha_2[T(\mathbf{e}_2)] + \cdots + \alpha_n[T(\mathbf{e}_n)] \quad \text{(property of coordinate matrices)}$$

$$= \big[\,[T(\mathbf{e}_1)] \quad [T(\mathbf{e}_2)] \quad \cdots \quad [T(\mathbf{e}_n)]\,\big] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}. \quad \text{(just matrix multiplication)}$$

Comparing with equation ($\dagger$) above, we see that we can take

$$A_T = \big[\,[T(\mathbf{e}_1)] \quad [T(\mathbf{e}_2)] \quad \cdots \quad [T(\mathbf{e}_n)]\,\big].$$

$\square$

Summarising: Given a linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^m$, there is a corresponding $m \times n$ matrix A_T satisfying

$$[T(\mathbf{v})] = A_T[\mathbf{v}]$$

That is, the image of any vector $\mathbf{v} \in \mathbb{R}^n$ can be calculated using the matrix A_T.

We revisit the example from Slide 291:

$$T : \mathbb{R}^3 \to \mathbb{R}^2, \quad T(x_1, x_2, x_3) = (x_2 - 2x_3, 3x_1 + x_3).$$

Recall that we found the standard matrix of transformation to be

$$A_T = \begin{bmatrix} 0 & 1 & -2 \\ 3 & 0 & 1 \end{bmatrix}.$$

Now, $T(\mathbf{e}_1) =$

$T(\mathbf{e}_2) =$

$T(\mathbf{e}_3) =$

Note that:

- the number of columns in $A_T = \dim(\mathbb{R}^3) = \dim(\text{domain of } T)$.
- the number of rows in $A_T = \dim(\mathbb{R}^2) = \dim(\text{codomain of } T)$.

Examples

1. Define $T : \mathbb{R}^3 \to \mathbb{R}^4$ by $T(x_1, x_2, x_3) = (x_1, x_3, x_2, x_1 + x_3)$. Calculate A_T.

2. Give a reason why the mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$ specified by $T(x_1, x_2) = (x_1 - x_2, x_1 + 1)$ is not a linear transformation.

7.4 Matrix representations in general [AR 8.4]

We've talked about the standard matrix representations for linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. We can now generalise the above theory to study linear transformations from any n-dimensional vector space V to any m-dimensional vector space W (with the same scalars).

To do this we first introduce bases and represent vectors by their coordinate matrices.

Let U and V be finite-dimensional vector spaces.

Suppose that

- $T\colon U \to V$ is a linear transformation.
- $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ is an (ordered) basis for U.
- $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_m\}$ is an (ordered) basis for V.

We want a matrix that can be used to calculate the effect of T. Specifically, if we denote the matrix by $A_{\mathcal{C},\mathcal{B}}$, we want

$$[T(\mathbf{u})]_{\mathcal{C}} = A_{\mathcal{C},\mathcal{B}}[\mathbf{u}]_{\mathcal{B}} \qquad \text{for all } \mathbf{u} \in U. \qquad (*)$$

Note: For this matrix equation to make sense, the size of $A_{\mathcal{C},\mathcal{B}}$ must be $m \times n$.

Theorem

There exists a unique matrix satisfying the above condition $(*)$. It is given by

$$A_{\mathcal{C},\mathcal{B}} = \Big[\ [T(\mathbf{b}_1)]_{\mathcal{C}} \quad [T(\mathbf{b}_2)]_{\mathcal{C}} \quad \cdots \quad [T(\mathbf{b}_n)]_{\mathcal{C}}\ \Big].$$

The proof is the same as for the case of the standard matrix A_T.

The matrix $A_{\mathcal{C},\mathcal{B}}$ is also denoted by $[T]_{\mathcal{C},\mathcal{B}}$ and is called the *matrix of T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$*.

In the special case in which $U = V$ and $\mathcal{B} = \mathcal{C}$, we often write $[T]_{\mathcal{B}}$ in place of $[T]_{\mathcal{B},\mathcal{B}}$.

Example

Find $[T]_{\mathcal{C},\mathcal{B}}$ for the linear transformation $T\colon \mathcal{P}_2 \to \mathcal{P}_1$ given by

$$T(a_0 + a_1x + a_2x^2) = (a_0 + a_2) + a_0x$$

using the bases $\mathcal{B} = \{1, x, x^2\}$ for $\mathcal{P}_2$ and $\mathcal{C} = \{1, x\}$ for $\mathcal{P}_1$.

Solution

$T(1) =$

$T(x) =$

$T(x^2) =$

$[T]_{\mathcal{C},\mathcal{B}} =$

Example

A linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ has matrix $\begin{bmatrix} 5 & 1 & 0 \\ 1 & 5 & -2 \end{bmatrix}$ with respect to the standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$. What is its matrix with respect to the basis $\mathcal{B} = \{(1,1,0), (1,-1,0), (1,-1,-2)\}$ of $\mathbb{R}^3$ and the basis $\mathcal{C} = \{(1,1), (1,-1)\}$ of $\mathbb{R}^2$?

Solution

We apply T to the elements of $\mathcal{B}$ to get

Then we write the result in terms of $\mathcal{C}$:

We obtain the matrix $[T]_{\mathcal{C},\mathcal{B}} =$

Example

Consider the linear transformation $T\colon V \to V$, where $V = M_{2\times 2}(\mathbb{R})$ is the vector space of real valued 2×2 matrices and T is defined by

$$T(Q) = Q^T = \text{ the transpose of } Q.$$

Find the matrix representation of T with respect to bases

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$$

and

$$\mathcal{C} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

Find the eigenvalues of this matrix representation.

Solution

The task is to work out the coordinates with respect to $\mathcal{C}$ of the image of each element in $\mathcal{B}$.

Matrix = linear transformation + basis

Eigenvalues, which were defined for matrices, only need the linear transformation, i.e. they are independent of the choice of basis.

Definition

Let $T\colon V \to V$ be a linear transformation. A scalar λ is an *eigenvalue* of T if there is a non-zero vector $\mathbf{v} \in V$ such that

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

The vector $\mathbf{v}$ is called an *eigenvector* of T with eigenvalue λ and the *eigenspace* of λ is the subspace of all vectors $\mathbf{v}$ satisfying $(*)$.

Note that there are many applications, e.g. quantum mechanics, which require V to be infinite-dimensional hence T is not represented by a matrix, but eigenvalues are still well-defined.

7.5 Image, kernel, rank and nullity

Let $T\colon U \to V$ be a linear transformation.

Definition (Kernel and Image)

The *kernel* of T is defined to be

$$\ker(T) = \{\mathbf{u} \in U \mid T(\mathbf{u}) = \mathbf{0}\}.$$

The *image* of T is defined to be

$$\mathrm{Im}(T) = \{\mathbf{v} \in V \mid \mathbf{v} = T(\mathbf{u}) \text{ for some } \mathbf{u} \in U\}.$$

The kernel is also called the *nullspace*. It is a subspace of U. Its dimension is denoted by $\mathrm{nullity}(T)$.

The image is also called the *range*. It is a subspace of V. Its dimension is denoted by $\mathrm{rank}(T)$.

Example

Consider the linear transformation $T\colon \mathcal{P}_2 \to \mathbb{R}^2$ defined by

$$T(a_0 + a_1 x + a_2 x^2) = (a_0 - a_1 + a_2, 2a_0 - 2a_1 + 2a_2).$$

Find bases for $\ker(T)$ and $\operatorname{Im}(T)$.

Note

When we calculate $\ker(T)$, we find that we are solving equations of the form $A_T X = \mathbf{0}$. So $\ker(T)$ *corresponds to the solution space for* A_T.

To calculate $\operatorname{Im}(T)$ we can use the fact that, because $\mathcal{B}$ spans U then $T(\mathcal{B})$ must span $T(U)$, the image of T. But the elements of $T(\mathcal{B})$ are given by the *columns* of A_T. Thus *the column space of* A_T *gives the coordinate vectors of the elements of* $\operatorname{Im}(T)$. It follows that $\operatorname{rank}(T) = \operatorname{rank}(A_T)$.

We can now use the result of slide 190 to prove the following:

Theorem (Rank-Nullity Theorem)

For a linear transformation $T\colon U \to V$ *with* $\dim(U) = n$,

$$\operatorname{nullity}(T) + \operatorname{rank}(T) = n = \dim(\textit{domain of } T).$$

Definition

A function $T: U \to V$ is called *injective* or *one-to-one* if $T(\mathbf{x}) = T(\mathbf{y})$ implies that $\mathbf{x} = \mathbf{y}$. It is called *surjective* or *onto* if $T(U) = V$. It is an *isomorphism* if it is both injective and surjective.

For linear transformations, injectivity simplifies to the following:

Theorem

A linear transformation $T: U \to V$ is injective iff $\ker(T) = \{\mathbf{0}\}$.

Example

Is the linear transformation of the example on Slide 307 injective ?

Is it surjective ?

Definition (Invertibility)

A linear transformation $T: U \to V$ is *invertible* if there is a linear transformation $S: V \to U$ such that

1. $(S \circ T)(\mathbf{u}) = \mathbf{u}$ for all $\mathbf{u} \in U$.
2. $(T \circ S)(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$.

Lemma

1. *If T is invertible, then the linear transformation S is unique. It is called the inverse of T and is denoted by T^{-1}.*
2. *A linear transformation T is invertible iff it is an isomorphism, i.e. T is both injective and surjective.*
3. *Choose any bases $\mathcal{B}$ and $\mathcal{C}$ of U and V, respectively. Then, T is invertible if and only if $[T]_{\mathcal{C},\mathcal{B}}$ is invertible. Moreover,*

$$[T^{-1}]_{\mathcal{B},\mathcal{C}} = ([T]_{\mathcal{C},\mathcal{B}})^{-1}.$$

7.6 Change of basis

Transition Matrices

We have seen that a matrix representation of a linear transformation $T: U \to V$ depends on both the choices of bases for U and V.

In fact different matrix representations are related by a matrix which depends on the bases but not on T. To understand this, we undertake a study of converting coordinates with respect to one basis to coordinates using another basis.

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases for the same vector space V and let $\mathbf{v} \in V$.

Q. How are $[\mathbf{v}]_{\mathcal{B}}$ and $[\mathbf{v}]_{\mathcal{C}}$ related?

A. By multiplication by a matrix!

Theorem

There exists a unique matrix $P_{\mathcal{C},\mathcal{B}}$ such that for any vector $\mathbf{v} \in V$,

$$[\mathbf{v}]_{\mathcal{C}} = P_{\mathcal{C},\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}.$$

The matrix $P_{\mathcal{C},\mathcal{B}}$ is given by

$$P_{\mathcal{C},\mathcal{B}} = [\,[\mathbf{b}_1]_{\mathcal{C}} \cdots [\mathbf{b}_n]_{\mathcal{C}}]$$

and is called the transition matrix from $\mathcal{B}$ to $\mathcal{C}$.

In words, the columns of $P_{\mathcal{C},\mathcal{B}}$ are the coordinate matrices, with respect to $\mathcal{C}$, of the elements of $\mathcal{B}$.

It is also sometimes denoted by $P_{\mathcal{B}\to\mathcal{C}}$.

Proof:
We want to find a matrix P such that for all vectors $\mathbf{v}$ in V,

$$[\mathbf{v}]_{\mathcal{C}} = P[\mathbf{v}]_{\mathcal{B}}.$$

Recall that if $T\colon V \to V$ is a linear transformation, then

$$[T(\mathbf{v})]_{\mathcal{C}} = [T]_{\mathcal{C},\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}, \qquad (*)$$

where

$$[T]_{\mathcal{C},\mathcal{B}} = \big[\, [T(\mathbf{b}_1)]_{\mathcal{C}} \; [T(\mathbf{b}_2)]_{\mathcal{C}} \; \cdots \; [T(\mathbf{b}_n)]_{\mathcal{C}} \,\big]$$

is the matrix representation of T.

Applying this to the special case where $T(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v}$ (i.e., T is the identity linear transformation) gives

$$[T]_{\mathcal{C},\mathcal{B}} = \big[\, [\mathbf{b}_1]_{\mathcal{C}} \, [\mathbf{b}_2]_{\mathcal{C}} \, \ldots \, [\mathbf{b}_n]_{\mathcal{C}} \big]$$

and $(*)$ becomes

$$[\mathbf{v}]_{\mathcal{C}} = [T]_{\mathcal{C},\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}.$$

So we can take $P = [T]_{\mathcal{C},\mathcal{B}}$.

Exercise
Finish the proof by showing that P is unique. That is, if Q is a matrix satisfying $[\mathbf{v}]_{\mathcal{C}} = Q[\mathbf{v}]_{\mathcal{B}}$ for all $\mathbf{v}$, then $Q = P$.

$\square$

A simple case

The transition matrix is easy to calculate when one of $\mathcal{B}$ or $\mathcal{C}$ is the standard basis $\mathcal{S}$.

Example

In $\mathbb{R}^2$, write down the transition matrix from $\mathcal{B}$ to $\mathcal{S}$, where

$$\mathcal{B} = \{(1,1),(1,-1)\} \text{ and } \mathcal{S} = \{(1,0),(0,1)\}.$$

Use this to compute $[\mathbf{v}]_{\mathcal{S}}$, given that $[\mathbf{v}]_B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Solution

$P_{\mathcal{S},\mathcal{B}} = \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{S}} & [\mathbf{b}_2]_{\mathcal{S}} \end{bmatrix} =$

$[\mathbf{v}]_{\mathcal{S}} = P_{\mathcal{S},\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} =$

Going in the other direction

A useful fact is that *transition matrices are always invertible.*

Starting with the equation

$$[\mathbf{v}]_{\mathcal{C}} = P_{\mathcal{C},\mathcal{B}}[\mathbf{v}]_B$$

and rearranging gives

$$[\mathbf{v}]_{\mathcal{B}} = P^{-1}_{\mathcal{C},\mathcal{B}}[\mathbf{v}]_{\mathcal{C}}.$$

But we know that

$$[\mathbf{v}]_{\mathcal{B}} = P_{\mathcal{B},\mathcal{C}}[\mathbf{v}]_{\mathcal{C}}$$

and so, by the uniqueness part of the above theorem, it must be the case that

$$P_{\mathcal{B},\mathcal{C}} = (P_{\mathcal{C},\mathcal{B}})^{-1}.$$

Example

For $\mathcal{B}$ and $\mathcal{S}$ as in the previous example, compute the transition matrix $P_{\mathcal{B},\mathcal{S}}$ from $\mathcal{S}$ to $\mathcal{B}$.

Then, use it to compute $[\mathbf{v}]_{\mathcal{B}}$, given $[\mathbf{v}]_{\mathcal{S}} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$.

Solution

We saw that in this case

$$P_{\mathcal{S},\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

It follows that

$P_{\mathcal{B},\mathcal{S}} =$

$[\mathbf{v}]_{\mathcal{B}} =$

Calculating a general transition matrix

Keeping notation as before, we have

$$[\mathbf{v}]_{\mathcal{S}} = P_{\mathcal{S},\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} \quad \text{and} \quad [\mathbf{v}]_{\mathcal{C}} = P_{\mathcal{C},\mathcal{S}}[\mathbf{v}]_{\mathcal{S}}.$$

Combining these, we get

$$[\mathbf{v}]_{\mathcal{C}} = P_{\mathcal{C},\mathcal{S}}[\mathbf{v}]_{\mathcal{S}} = P_{\mathcal{C},\mathcal{S}}P_{\mathcal{S},\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}.$$

Using the uniqueness of the transition matrix, we arrive at

$$P_{\mathcal{C},\mathcal{B}} = P_{\mathcal{C},\mathcal{S}}P_{\mathcal{S},\mathcal{B}} = P_{\mathcal{S},\mathcal{C}}^{-1}P_{\mathcal{S},\mathcal{B}}.$$

Since it is usually easy to calculate the transition matrix from a non-standard basis to the standard basis ($P_{\mathcal{S},\mathcal{B}}$), this makes it straightforward to calculate any transition matrix.

Example

With $U = V = \mathbb{R}^2$ and $\mathcal{B} = \{(1,2),(1,1)\}$ and $\mathcal{C} = \{(-3,4),(1,-1)\}$, find $P_{\mathcal{C},\mathcal{B}}$.

We compute

$$P_{\mathcal{S},\mathcal{B}} = \qquad\qquad P_{\mathcal{S},\mathcal{C}} =$$

so

$$P_{\mathcal{C},\mathcal{S}} = \qquad\qquad \text{and} \quad P_{\mathcal{C},\mathcal{B}} =$$

Relationship Between Different Matrix Representations

Example

Calculate the standard matrix representation of $T : \mathbb{R}^2 \to \mathbb{R}^2$, where

$$T(x, y) = (3x - y, -x + 3y).$$

Solution

Example continued

Now find the matrix of T with respect to the basis
$\mathcal{B} = \{(1,1),(1,-1)\}$.

Solution

Notice that $[T]_{\mathcal{B}}$ is *diagonal*. This makes it very convenient to use the basis $\mathcal{B}$ in order to understand the effect of T.

We will later see a systematic way to find a basis $\mathcal{B}$ making $[T]_{\mathcal{B}}$ diagonal, by using "eigenvectors and eigenvalues" of a linear transformation T.

How are $[T]_{\mathcal{C}}$ and $[T]_{\mathcal{B}}$ related?

Theorem

The matrix representations of $T : V \to V$ with respect to two bases $\mathcal{C}$ and $\mathcal{B}$ are related by the following equation:

$$[T]_{\mathcal{B}} = P_{\mathcal{B},\mathcal{C}}[T]_{\mathcal{C}}P_{\mathcal{C},\mathcal{B}}.$$

Proof:

We need to show that for all $\mathbf{v} \in V$,

$$[T(\mathbf{v})]_{\mathcal{B}} = P_{\mathcal{B},\mathcal{C}}[T]_{\mathcal{C}}P_{\mathcal{C},\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}.$$

Starting with the right-hand side we obtain

$$\begin{aligned} P_{\mathcal{B},\mathcal{C}}[T]_{\mathcal{C}}P_{\mathcal{C},\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} &= P_{\mathcal{B},\mathcal{C}}[T]_{\mathcal{C}}[\mathbf{v}]_{\mathcal{C}} && (\text{property of } P_{\mathcal{C},\mathcal{B}}) \\ &= P_{\mathcal{B},\mathcal{C}}[T(\mathbf{v})]_{\mathcal{C}} && (\text{property of } [T]_{\mathcal{C}}) \\ &= [T(\mathbf{v})]_{\mathcal{B}}. && (\text{property of } P_{\mathcal{B},\mathcal{C}}) \end{aligned}$$

$\square$

Example

For the above linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (3x - y, -x + 3y)$, we saw that

$[T]_{\mathcal{C}} =$

$[T]_{\mathcal{B}} =$

for $\mathcal{C} = \{(1,0),(0,1)\}$ the standard basis and $\mathcal{B} = \{(1,1),(1,-1)\}$.

Since $\mathcal{C}$ is the standard basis, it is easy to write down $P_{\mathcal{C},\mathcal{B}} =$

From which we calculate $P_{\mathcal{B},\mathcal{C}} =$

Calculation verifies that in this case we do indeed have

$$[T]_{\mathcal{B}} = P_{\mathcal{B},\mathcal{C}}[T]_{\mathcal{C}}P_{\mathcal{C},\mathcal{B}}.$$

Notice that the eigenvalues of $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{C}}$ are the same.