MA1522: Linear Algebra for Computing

Chapter 1: Linear Systems

1.1 Introduction to Linear Systems

Introduction to Linear Systems

Consider the following

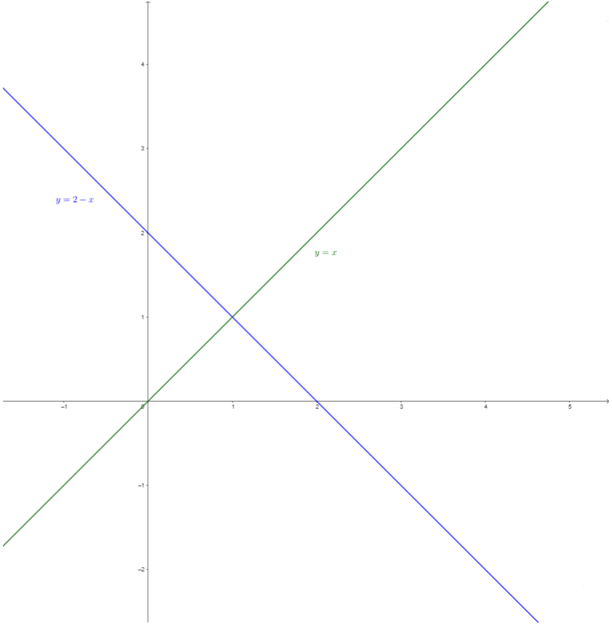
$$\begin{cases} x & + & y & = & 2 \\ x & - & y & = & 0 \end{cases}$$

which is (formerly and fondly) know as *simultaneous equations*.

(Only) Solutions: $x = 1$, $y = 1$.

Geometrically, the solution is the intersection of 2 lines.

The above is an example of a *linear system*. In most applications, a linear system involves a large number of variables and equations. We will learn what they are and how to solve them in this chapter.



Definition of Linear Equations

Definition

A *linear* equation with n variables in *standard form* has the form

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

Here $a_1, a_2, ..., a_n$ are known constants, called the *coefficients*, b is called the *constant*, and $x_1, x_2, ..., x_n$ are *variables*.

The linear equation is *homogeneous* if $b = 0$, i.e.

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

Examples

Which of the following equations are linear in x and y?

1. $2x + y = 3$
2. $x = 2$
3. $y = x\sin(\frac{\pi}{6})$
4. $x = ky$, for some $k \in \mathbb{R}$.
5. $xy = 3$
6. $y = x^2$
7. $y = \sin(x)$
8. $3\cos(x) + 4\sin(y) = 2$

Which of the linear equations are in standard form? Which of the linear equations are homogeneous?

Definition of Linear Systems

Definition

A *system of linear equations*, or a *linear system* consists of a finite number of linear equations. In general, a linear system with n variables and m equations in *standard form* is written as

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

The linear system is *homogeneous* if $b_1 = b_2 = \ldots = b_m = 0$, that is, all the linear equations are homogeneous.

Solutions to a Linear System

Definition

Given a linear system

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}\right.$$

we say that

$$x_1 = c_1,\ x_2 = c_2,\ \ldots,\ x_n = c_n$$

is a *solution* to the linear system if the equations are simultaneously satisfied after making the substitution.

General Solutions

Definition

The *general solution* to a linear system captures all possible solutions to the linear system.

When there are infinitely many solutions to a linear system, we can assign some variables as parameters, and express the other variables in terms of the parameters.

The general solution of the example above will be $x = 5 - 2s,\ y = s,\ s \in \mathbb{R}$. Here the symbol $s \in \mathbb{R}$ means that s can be any real number, which we call a parameter of the general solution. Alternatively, the general solution can be expressed as $y = \frac{5-t}{2},\ x = t,\ t \in \mathbb{R}$.

Caution

Do not use the variables as parameters. That is, do not write $x = 5 - 2x$, as it is begging the question.

Inconsistent Linear System

Consider now another linear system.

$$\begin{cases} x & + & y & = & 2 \\ x & - & y & = & 0 \\ 2x & + & y & = & 1 \end{cases}$$

Are you able to provide any solution to the system?

Observe that $x = 1 = y$ is the only solution to the first 2 equations, but not a solution to the third equations.

Hence, the system has not solution. In this case, we say that the system is <u>inconsistent</u>.

Inconsistent Linear Systems

Definition

A linear system is said to be ***inconsistent*** if it does not have any solutions. It is ***consistent*** otherwise, that is, a linear system is consistent if it has at least one solution.

We have seen examples of linear system that are inconsistent, consistent with a unique solution, or consistent with infinitely many solutions. Are there examples of linear systems that have more than 1 but only finitely many solutions?

Linear Systems with 2 Variables

Consider a linear system with 2 variables

$$\begin{cases} a_1 x & + & b_1 y & = & c_1 \\ a_2 x & + & b_2 y & = & c_2 \end{cases}$$

for some real numbers $a_i, b_i, c_i,\ i = 1, 2$.

Each equation is a line in the 2-dimensional graph . The solution of this system corresponds to a point of intersection of the lines, and there are three possibilities.

(i) The lines are parallel and distinct. System has no solution.

(ii) The lines intersect at only one point. System has a unique solution.

(iii) The lines coincide. System has infinitely many solutions.

Head over to the following link to visualize the linear system.
https://www.geogebra.org/m/ahvsz6j9

Linear Systems with 3 Variables

Consider a linear system with 3 variables

$$\begin{cases} a_1x & + & b_1y & + & c_1z & = & d_1 \\ a_2x & + & b_2y & + & c_2z & = & d_2 \\ a_3x & + & b_3y & + & c_3z & = & d_3 \end{cases}$$

for some real numbers $a_i, b_i, c_i, d_i,\ i = 1, 2, 3$.

Each equation is a plane in the 3d-space. First note that 2 planes intersect along a line.

(i) The system has no solution if either 2 planes are parallel but distinct, or the planes intersects, but the lines of intersections do not intersect.

(ii) The system has a unique solution if all 3 planes intersects at a point.

(iii) The system has infinitely many solutions with 1 parameter in the general solution if the 3 planes intersects along a line.

(iv) The system has infinitely many solutions with 2 parameters in the general solution if the 3 planes coincide.

Head over to the following link to visualize the linear system.
https://www.geogebra.org/m/wxneuj7k

Discussions

1. Give an example of a linear system with 3 variables such that the general solution has 2 parameters.

2. Is it possible to have a linear system with 3 variables, 3 equations, with the general solution having 3 parameters?

3. In the next section we will see that it is true that a linear system will always only have either no solution, a unique solution, or infinitely many solutions.

1.2 Solving a Linear System and Row-Echelon Form

Augmented Matrix

A linear system

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

can be expressed uniquely as an augmented matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right).$$

That is, we first express the linear system in standard form, then put the coefficients in the entries of the left hand side of the bar in the augmented matrix, and the constant on the right hand side of the bar in the augmented matrix.

Remark

The augmented matrix is an example of a matrix, which we will discuss in further details in Chapter 2.

Row-Echelon Form

Definition

In an (augmented) matrix, a *zero row* is a row with all entries 0. A row is called a *nonzero row* otherwise. The first nonzero entry from the left of a nonzero row is called a *leading entry*.

An (augmented) matrix is in *row-echelon form* (REF) if

1. If zero rows exists, they are at the bottom of the matrix.
2. The leading entries are further to the right as we move down the rows.

An augmented matrix in REF has the form

$$\left(\begin{array}{cccccccc|c}
* & & & & & & & & * \\
0 & \cdots & 0 & * & & & & & * \\
0 & \cdots & 0 & 0 & \cdots & 0 & * & & * \\
\vdots & & & & & & & \vdots & \vdots \\
0 & \cdots & & & & & \cdots & 0 & 0
\end{array}\right).$$

Definition

In row-echelon form, a column containing a leading entry is called a *pivot column*. It is called a *non-pivot column* otherwise.

Reduced Row-Echelon Form

The (augmented) matrix is in reduced row-echelon form (RREF) if further

3. The leading entries are 1.
4. In each pivot column, all entries except the leading entry is 0.

An augmented matrix in RREF has the form

$$\left(\begin{array}{cccccccccc|c}
0 & \cdots & 1 & & * & 0 & & * & 0 & * & * \\
0 & \cdots & 0 & \cdots & 0 & 1 & & * & 0 & * & * \\
0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & * & * \\
0 & \cdots & 0 & & & 0 & & & 0 & 0 & 0 \\
\vdots & & & & & \vdots & & & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & \cdots & & 0 & & \cdots & 0 & 0 & 0
\end{array}\right).$$

Remark

Note that a (an augmented) matrix in reduced row-echelon form is also in row-echelon form. The reduced row-echelon form is a special case of row-echelon form.

Solutions from REF and RREF

When the augmented matrix is in row-echelon or reduced row-echelon form, it is easy to extract the solutions of the linear system.

- If the augmented matrix is in row-echelon form, we perform back substitution to obtain the solutions.
- If the augmented matrix is in reduced row-echelon form, we will read off the solutions directly.

Remarks

1. It is easy to obtain the solutions when the augmented matrix is in row-echelon form (by performing back-substitution) or reduced row-echelon form (reading off the solutions directly).

2. The linear system is inconsistent if and only if the RHS (last column) of the augmented matrix in row-echelon form is a pivot column.

3. Assign parameters to the variables corresponding to the non-pivot columns in the LHS of the augmented matrix.

4. The number of parameters needed is equal to the number of non-pivot columns in the LHS of the augmented matrix.

5. We can convert/reduce the augmented matrix of a linear system to a row-echelon form or its reduced row-echelon form to find the solutions (if exists). This is achieved using elementary row operations.

Challenge

Let $\mathbf{R}$ be a $n \times m$ matrix in reduced row-echelon form. Which of the following statements are true?

1. The number of pivot columns of $\mathbf{R}$ is equal to the number of nonzero rows of $\mathbf{R}$.

2. The number of nonpivot columns of $\mathbf{R}$ is equal to the number of zero rows in $\mathbf{R}$.

For each statement that is false, what restrictions can we impose on $\mathbf{R}$ such that the statement is true?

1.3 Elementary Row Operations

Motivation

Consider the following linear system

$$\begin{cases} x + 2y = 3 \\ 4x + 5y = 6 \end{cases}$$

To solve it, we perform the following operations

$$\begin{array}{lcrl} (\text{equation } 2) - 4 \times (\text{equation } 1) & \to & -3y = -6 & (\text{equation } 3) \\ -\frac{1}{3}(\text{equation } 3) & \to & y = 2 & (\text{equation } 4) \end{array}$$

Substitute equation 4 into equation 1, or

$$(\text{equation } 1) - 2 \times (\text{equation } 4) \quad \to \quad x = -1 \quad (\text{equation } 5)$$

we conclude that $x = -1, y = 2$ is the (unique) solution to the system.

Observe that in all the operations above, the solution(s) to the system is preserved. This motivated the introduction of *elementary row operations*, that is, a fundamental set of operations that do not alternative the solution set when performed on the augmented matrix of a linear system.

Elementary Row Operations

There are 3 types of elementary row operations.

1. Exchanging 2 rows, $R_i \leftrightarrow R_j$,
2. Adding a multiple of a row to another, $R_i + cR_j$, $c \in \mathbb{R}$,
3. Multiplying a row by a nonzero constant, aR_j, $a \neq 0$.

Remark

Performing elementary row operations to the augmented matrix of a linear system preserves the solutions.

Row Equivalent Matrices

Definition

Two (augmented) matrices are *row equivalent* if one can be obtained from the other by performing a set of elementary row operations.

Theorem

Two linear systems have the same solutions if their augmented matrices are row equivalent.

The proof of the theorem can be found in the appendix of chapter 2.

Remarks

Recall that when multiplying a row by a constant, aR_i, we require the multiple to be nonzero, $a \neq 0$.

Consider the following linear system.

$$\begin{array}{ccccc} x & + & y & = & 2 \\ x & - & y & = & 0 \end{array}$$

It has a unique solution $x = y = 1$. Now suppose we multiple the row two of the augmented matrix by 0,

$$\left(\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & -1 & 0 \end{array}\right) \xrightarrow{0R_2} \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right)$$

we obtain a system with infinitely many solution; the general solution is $x = 2 - s$, $y = s$, $s \in \mathbb{R}$.

This demonstrates that if we multiply a row in the augmented matrix by 0, we may change the solution of the linear system.

Remarks

The order by which we perform the row operations matters (that is row operations do not commute).

Consider the following augmented matrix

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right)$$

By taking 2 times of row two, then exchanging the row, we obtain

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{2R_1} \xrightarrow{R_2 \leftrightarrow R_1} \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 2 & 0 & 0 \end{array}\right).$$

However, if we first exchange the rows then take 2 times of row two, we have

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{R_2 \leftrightarrow R_1} \xrightarrow{2R_1} \left(\begin{array}{cc|c} 0 & 2 & 0 \\ 1 & 0 & 0 \end{array}\right).$$

Remarks

Hence, when performing a few row operations on an augmented matrix, we have to write the row operations from left to right.

However, if the row operations commute, that is, in some cases where it does not matter which row operations we perform first, we may stack them together as such

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow[2R_1]{2R_2} \left(\begin{array}{cc|c} 2 & 0 & 0 \\ 0 & 2 & 0 \end{array}\right)$$

Remarks

The notation $R_i + cR_j$ acts on R_i, but keeps R_j unchanged.

For example

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{R_1+R_2} \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 0 \end{array}\right),$$

whereas

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{R_2+R_1} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 1 & 1 & 0 \end{array}\right).$$

Remarks

Here is another example. $R_1 + 2R_2$ is an elementary row operation,

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{R_1+2R_2} \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 1 & 0 \end{array}\right),$$

which is different from the row operation $2R_2 + R_1$

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{2R_2+R_1} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 1 & 2 & 0 \end{array}\right).$$

In fact, $2R_2 + R_1$ is not an elementary row operation, but a combination of 2 elementary row operations, first perform $2R_2$, then perform $R_2 + R_1$;

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \xrightarrow{2R_2} \xrightarrow{R_2+R_1} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 1 & 2 & 0 \end{array}\right).$$

Discussion

Consider the following elementary row operations.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right) \xrightarrow{R_2+3R_1} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 3 & 1 & 0 & 5 \\ 0 & 0 & 1 & 3 \end{array}\right) \xrightarrow{??} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

What is the (unique) elementary row operation that when performed, would change the middle augmented matrix back to its original one?

Since the second row in the augmented matrix in the middle had an extra 3 times of row 1, to undo it, we need to remove the extra 3 times of row 1. That is, if we perform $R_2 - 3R_1$, we will get the original augmented matrix. This shows that the elementary row operation $R_i + cR_j$ can be undone, or "reversed" if we perform $R_i - cR_j$. This is true for all elementary row operations.

Reverse of Elementary Row Operations

Every elementary row operation has a *reverse elementary row operation*. The reverse of the row operations are given as such.

1. The reverse of exchanging 2 rows, $R_i \leftrightarrow R_j$, is itself.

2. The reverse of adding a multiple of a row to another, $R_i + cR_j$ is subtracting the multiple of that row, $R_i - cR_j$.

3. The reverse of multiplying a row by a nonzero constant, aR_j is the multiplication of the reciprocal of the constant, $\frac{1}{a}R_j$.

Readers should convinced themselves that the reverse of the elementary row operations are indeed the ones given above. In fact, this is the definitive property of an elementary row operation; all (linear) operations that can be reversed are composed (made up of a series) of elementary row operations.

Examples

1. The reverse of $R_2 \leftrightarrow R_4$ is itself,

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right) \xrightarrow{R_2 \leftrightarrow R_4} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 1 & 0 & 0 & 2 \end{array}\right) \xrightarrow{R_2 \leftrightarrow R_4} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right).$$

2. The reverse of $R_2 + 3R_3$ is $R_2 - 3R_3$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right) \xrightarrow{R_2 + 3R_3} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 0 & 11 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right) \xrightarrow{R_2 - 3R_3} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right).$$

3. The reverse of $2R_3$ is $\frac{1}{2}R_3$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right) \xrightarrow{2R_3} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 & 6 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right) \xrightarrow{\frac{1}{2}R_3} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right).$$

Remark

- Since performing elementary row operations do not change the solution set of a linear system, and
- the solutions of a linear system can be obtained easily from the augmented matrix in row-echelon form,
- our aim is therefore to use elementary row operations to reduce the augmented matrix of a linear system until it is in row-echelon or even better, in reduced row-echelon form.
- This will be the discussion of the next section.

1.4 Row Reduction, Gaussian and Gauss-Jordan Elimination

Introduction

- Aim: to reduce an augmented matrix to a row-echelon, or reduced row-echelon form using elementary row operations.
- Gaussian or Gauss-Jordan elimination are algorithms that do so.
- These algorithm guarantees to reduce the augmented matrix, but might not be the most efficient.
- May have to modify the algorithms when the augmented matrix contains unknown coefficients.

Gaussian Elimination

Step 1: Locate the leftmost column that does not consist entirely of zeros.

Step 2: Interchange the top row with another row, if necessary, to bring a nonzero entry to the top of the column found in Step 1.

Example

(i) $\left(\begin{array}{cccc|c} 0 & 3 & 9 & \cdots & * \\ 1 & 2 & -3 & \cdots & * \\ 4 & 1 & 0 & \cdots & * \end{array}\right) \xrightarrow{R_1 \leftrightarrow R_2} \left(\begin{array}{cccc|c} 1 & 2 & -3 & \cdots & * \\ 0 & 3 & 9 & \cdots & * \\ 4 & 3 & 0 & \cdots & * \end{array}\right)$

(ii) $\left(\begin{array}{cccc|c} 0 & 0 & 5 & \cdots & * \\ 0 & 1 & 2 & \cdots & * \\ 0 & -3 & 2 & \cdots & * \end{array}\right) \xrightarrow{R_1 \leftrightarrow R_2} \left(\begin{array}{cccc|c} 0 & 1 & 2 & \cdots & * \\ 0 & 0 & 5 & \cdots & * \\ 0 & -3 & 2 & \cdots & * \end{array}\right)$

(iii) $\left(\begin{array}{cccc|c} 1 & 0 & 2 & \cdots & * \\ 0 & 5 & -1 & \cdots & * \\ 0 & 4 & 1 & \cdots & * \end{array}\right)$ no interchanging of rows needed.

Gaussian Elimination

Step 3: For each row below the top row, add a suitable multiple of the top row to it so that the entry below the leading entry of the top row becomes zero.

(i) $\left(\begin{array}{cccc|c} 1 & 2 & -3 & \cdots & * \\ 0 & 3 & 9 & \cdots & * \\ 4 & 1 & 0 & \cdots & * \end{array}\right) \xrightarrow{R_3-4R_1} \left(\begin{array}{cccc|c} 1 & 2 & -3 & \cdots & * \\ 0 & 3 & 9 & \cdots & * \\ 0 & -7 & 12 & \cdots & * \end{array}\right)$

(ii) $\left(\begin{array}{cccc|c} 0 & 1 & 2 & \cdots & * \\ 0 & 0 & 5 & \cdots & * \\ 0 & -3 & 2 & \cdots & * \end{array}\right) \xrightarrow{R_3+3R_1} \left(\begin{array}{cccc|c} 0 & 1 & 2 & \cdots & * \\ 0 & 0 & 5 & \cdots & * \\ 0 & 0 & 8 & \cdots & * \end{array}\right)$

(iii) $\left(\begin{array}{cccc|c} 1 & 0 & 2 & \cdots & * \\ 0 & 5 & -1 & \cdots & * \\ 0 & 4 & 1 & \cdots & * \end{array}\right)$ no change required.

Gaussian Elimination

Step 4: Now cover the top row in the augmented matrix and begin again with Step 1 applied to the submatrix that remains. Continue this way until the entire matrix is in row-echelon form.

(i) $\left(\begin{array}{cccc|c} 1 & 2 & -3 & \cdots & * \\ 0 & 3 & 9 & \cdots & * \\ 0 & -7 & 12 & \cdots & * \end{array}\right) \xrightarrow{R_3+\frac{7}{3}R_2} \left(\begin{array}{cccc|c} 1 & 2 & -3 & \cdots & * \\ 0 & 3 & 9 & \cdots & * \\ 0 & 0 & 33 & \cdots & * \end{array}\right)$

(ii) $\left(\begin{array}{cccc|c} 0 & 1 & 2 & \cdots & * \\ 0 & 0 & 5 & \cdots & * \\ 0 & 0 & 8 & \cdots & * \end{array}\right) \xrightarrow{R_3-\frac{8}{5}R_2} \left(\begin{array}{cccc|c} 0 & 1 & 2 & \cdots & * \\ 0 & 0 & 5 & \cdots & * \\ 0 & 0 & 0 & \cdots & * \end{array}\right)$

(iii) $\left(\begin{array}{cccc|c} 1 & 0 & 2 & \cdots & * \\ 0 & 5 & -1 & \cdots & * \\ 0 & 4 & 1 & \cdots & * \end{array}\right) \xrightarrow{R_3-\frac{4}{5}R_2} \left(\begin{array}{cccc|c} 1 & 0 & 2 & \cdots & * \\ 0 & 5 & -1 & \cdots & * \\ 0 & 0 & 9/5 & \cdots & * \end{array}\right)$

The result of step 1 to 4 reduces to (augmented) matrix to a row-echelon form. The process up to step 4 is called *Gaussian Elimination*.

Gauss-Jordan Elimination

Step 5: Multiply a suitable constant to each row so that all the leading entries become 1.

Step 6: Beginning with the last nonzero row and working upward, add suitable multiples of each row to the rows above to introduce zeros above the leading entries.

$$\left(\begin{array}{ccccccccc|c} 1 & & * & 0 & & * & 0 & & & * \\ 0 & \cdots & 0 & 1 & & * & 0 & & & * \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & & & * \\ 0 & & & 0 & & & 0 & & 0 & 0 \\ \vdots & & & \vdots & & & \vdots & & \vdots & \vdots \\ 0 & \cdots & & 0 & & \cdots & 0 & & 0 & 0 \end{array}\right).$$

If we continue from step 1 to 6, then the entire process is known as *Gauss-Jordan elimination*.

Challenge

1. Is it possible to reduce an augmented matrix to 2 different row-echelon forms?

2. Is it possible to reduce an augmented matrix to 2 different reduced row-echelon form?

1.5 More on Linear Systems

Summary

1. Write the linear system in its standard form.

2. Form the augmented matrix of the linear system.

3. Reduce the augmented matrix to either a row-echelon form or reduced row echelon form. May use Gaussian/Gauss-Jordan elimination.

4. Decide if the system is consistent
   - If the last column is a pivot column, the system is inconsistent.
   - Otherwise, the system is consistent, assign the variables corresponding to the nonpivot columns to be parameters, s, t, $s, t \in \mathbb{R}$, etc.

5. If the system is in reduced row-echelon form, read off the solutions directly.

6. If the system is in row-echelon form only, do back substitution, starting from the lowest nonzero row.

7. Write down the (general) solution to the system.

More Examples

Solve the following linear system

$$\begin{array}{rcrcrcrcrcrcr}
x_1 & + & 3x_2 & - & 2x_3 & + & & + & 2x_5 & & & = & 0 \\
2x_1 & + & 6x_2 & - & 5x_3 & - & 2x_4 & + & 4x_5 & - & 3x_6 & = & -1 \\
 & & & & 5x_3 & + & 10x_4 & & & + & 15x_6 & = & 5 \\
2x_1 & + & 6x_2 & & & + & 8x_4 & + & 4x_5 & + & 18x_6 & = & 6
\end{array}$$

$$\left(\begin{array}{cccccc|c}
1 & 3 & -2 & 0 & 2 & 0 & 0 \\
2 & 6 & -5 & -2 & 4 & -3 & -1 \\
0 & 0 & 5 & 10 & 0 & 15 & 5 \\
2 & 6 & 0 & 8 & 4 & 18 & 6
\end{array}\right) \xrightarrow{R_2-2R_1} \xrightarrow{R_4-2R_1} \left(\begin{array}{cccccc|c}
1 & 3 & -2 & 0 & 2 & 0 & 0 \\
0 & 0 & -1 & -2 & 0 & -3 & -1 \\
0 & 0 & 5 & 10 & 0 & 15 & 5 \\
0 & 0 & 4 & 8 & 0 & 18 & 6
\end{array}\right)$$

$$\xrightarrow{R_3+5R_2} \xrightarrow{R_4+4R_2} \left(\begin{array}{cccccc|c}
1 & 3 & -2 & 0 & 2 & 0 & 0 \\
0 & 0 & -1 & -2 & 0 & -3 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 6 & 2
\end{array}\right) \xrightarrow{R_3 \leftrightarrow R_4} \left(\begin{array}{cccccc|c}
1 & 3 & -2 & 0 & 2 & 0 & 0 \\
0 & 0 & -1 & -2 & 0 & -3 & -1 \\
0 & 0 & 0 & 0 & 0 & 6 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right)$$

The augmented matrix is now in row-echelon form. One may choose to use back-substitution to obtain the answer. Which variables should we assign as parameters?

More Examples

Let us continue to reduce the system to its reduced row-echelon form.

$$\xrightarrow{-R_2}\xrightarrow{(1/6)R_3}\left(\begin{array}{cccccc|c} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_2-3R_3} \left(\begin{array}{cccccc|c} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{R_1+2R_2}\left(\begin{array}{cccccc|c} 1 & 3 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Let $x_2 = r, x_4 = s, x_5 = t$. Then the general solution is

$$x_1 = -3r - 4s - 2t,\ x_2 = r,\ x_3 = -2s,\ x_4 = s,\ x_5 = t,\ x_6 = 1/3, \quad r, s, t \in \mathbb{R}.$$

Linear System with Unknowns in the Coefficient and Constants

Consider the following linear system

$$\begin{array}{ccccccc} x_1 & + & ax_2 & + & 2x_3 & = & 0 \\ x_1 & & & + & x_3 & = & 1 \\ x_1 & & & + & ax_3 & = & 2 \end{array}$$

for some fixed real number $a \in \mathbb{R}$. Here we want to determine the value of a such that the system is

(a) inconsistent,

(b) consistent with a unique solution and find the unique solution, or

(c) consistent with infinitely many solutions and write down the general solution.

$$\left(\begin{array}{ccc|c} 1 & a & 2 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & a & 2 \end{array}\right) \xrightarrow{R_2-R_1} \xrightarrow{R_3-R_1} \left(\begin{array}{ccc|c} 1 & a & 2 & 0 \\ 0 & -a & -1 & 1 \\ 0 & -a & a-2 & 2 \end{array}\right) \xrightarrow{R_3-R_2} \left(\begin{array}{ccc|c} 1 & a & 2 & 0 \\ 0 & -a & -1 & 1 \\ 0 & 0 & a-1 & 1 \end{array}\right).$$

Observe that if we want to continue the Gauss-Jordan elimination, we would need to multiple row 2 by $-\frac{1}{a}$ and row 3 by $\frac{1}{a-1}$. But this is not well defined if $a = 0$ and $a = 1$, respectively. Hence, we need to consider if $a = 0, 1$.

Linear System with Unknowns in the Coefficient and Constants

If $a = 1$, the augmented matrix becomes

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

The system is inconsistent.

If $a = 0$, the augmented matrix becomes

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \end{array}\right) \xrightarrow{R_3 - R_2} \xrightarrow{R_1 + 2R_2} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The system is consistent with infinitely many solution. The general solution is

$$x_1 = 2, \ x_2 = s, \ x_3 = -1, \quad s \in \mathbb{R}.$$

Linear System with Unknowns in the Coefficient and Constants

Finally, suppose $a \neq 0, 1$. By back substitution or reducing the augmented matrix to its reduced row-echelon form, we obtain the unique solution

$$x_1 = \frac{a-2}{a-1}, \quad x_2 = \frac{1}{1-a}, \quad x_3 = \frac{1}{a-1}.$$

The details are left as an exercise. Note that there are infinitely many values of a such that the system has a unique solution, and the solution depends on these values of a. Do not pick a particular value of a for the solution.

(a) The system has no solution if $a = 1$.

(b) The system has a unique solution when $a \neq 0, 1$. The unique solution is

$$x_1 = \frac{a-2}{a-1}, \quad x_2 = \frac{1}{1-a}, \quad x_3 = \frac{1}{a-1}.$$

(c) The system has infinitely many solutions when $a = 0$. The general solution is

$$x_1 = 2,\ x_2 = s,\ x_3 = -1, \quad s \in \mathbb{R}.$$

Another Example

Consider the following linear system

$$\left\{\begin{array}{rcrcrcl} ax & + & 2ay & - & z & = & 2a+2 \\ x & + & y & & & = & 1 \\ x & + & y & + & (a-1)z & = & 3 \end{array}\right.$$

for some fixed real number $a \in \mathbb{R}$.

The augmented matrix of the system is $\left(\begin{array}{ccc|c} a & 2a & -1 & 2a+2 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & a-1 & 3 \end{array}\right)$. Here, if we follow Gaussian elimination strictly, one might consider the operation $R_2 - \frac{1}{a}R_1$. However, this is not well-defined if $a = 0$. Hence, instead of considering cases, we might consider doing a row swap.

$$\left(\begin{array}{ccc|c} a & 2a & -1 & 2a+2 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & a-1 & 3 \end{array}\right) \xrightarrow{R_2 \leftrightarrow R_1} \left(\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ a & 2a & -1 & 2a+2 \\ 1 & 1 & a-1 & 3 \end{array}\right) \xrightarrow[R_3 - R_1]{R_2 - aR_1} \left(\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & a & -1 & a+2 \\ 0 & 0 & a-1 & 2 \end{array}\right)$$

Another Example

We should consider whether $a = 0, 1$ or not.

(a) The system is inconsistent if $a = 1$.

(b) The system has a unique solution if $a \neq 0, 1$. The unique solution is

$$x = \frac{2}{1-a}, \quad y = \frac{a+1}{a-1}, \quad z = \frac{2}{a-1}.$$

(c) The system has infinitely many solutions with 1 parameter if $a = 0$. The general solution is

$$x = 1 - s, \quad y = s, \quad z = -2, \quad s \in \mathbb{R}.$$

The details are left as an exercise.

Hard Example

Consider the following linear system

$$\begin{cases} x_1 & & & + & 3x_3 & + & x_4 & = & 2 \\ 3x_1 & + & ax_2 & + & 9x_3 & & & = & 6 \\ 2x_1 & & & + & (a+6)x_3 & + & ax_4 & = & b+2 \\ 2x_1 & & & + & 6x_3 & + & bx_4 & = & b+2 \end{cases}$$

where a and b are some constants.

(a) Find the conditions on a and b such that the system has no solution.

(b) Find the conditions on a and b such that the system has a unique solution, and write down the unique solution.

(c) Find the conditions on a and b such that the system has infinitely many solutions with 1 parameter, and write down a general solution.

(d) Find the conditions on a and b such that the system has infinitely many solutions with 2 parameter, and write down a general solution.

Hard Example

$$\left(\begin{array}{cccc|c} 1 & 0 & 3 & 1 & 2 \\ 3 & a & 9 & 0 & 6 \\ 2 & 0 & a+6 & a & b+2 \\ 2 & 0 & 6 & b & b+2 \end{array}\right) \xrightarrow[\substack{R_3-2R_1 \\ R_4-2R_1}]{R_2-3R_1} \left(\begin{array}{cccc|c} 1 & 0 & 3 & 1 & 2 \\ 0 & a & 0 & -3 & 0 \\ 0 & 0 & a & a-2 & b-2 \\ 0 & 0 & 0 & b-2 & b-2 \end{array}\right).$$

The cases to consider are $a = 0$ or not and $b = 2$ or not; a total of 4 cases.

Suppose $b = 2$. The augmented matrix becomes

$$\left(\begin{array}{cccc|c} 1 & 0 & 3 & 1 & 2 \\ 0 & a & 0 & -3 & 0 \\ 0 & 0 & a & a-2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

If $a = 0$ then columns 2 and 3 are non-pivots. Hence, the system has infinitely many solutions with 2 parameters.The general solution is

$$x_1 = 2 - 3t, \quad x_2 = s, \quad x_3 = t, \quad x_4 = 0, \quad s, t, \in \mathbb{R}.$$

If $a \neq 0$, then column 4 is non-pivot, and hence the system has infinitely many solutions with 1 parameters. The general solution is

$$x_1 = 2 + \frac{2a-6}{a}s, \quad x_2 = \frac{3}{a}s, \quad x_3 = \frac{(2-a)}{a}s, \quad x_4 = s, \quad s \in \mathbb{R}.$$

Hard Example

Suppose $b \neq 2$. We may proceed to reduce the augmented matrix.

$$\left(\begin{array}{cccc|c} 1 & 0 & 3 & 1 & 2 \\ 0 & a & 0 & -3 & 0 \\ 0 & 0 & a & a-2 & b-2 \\ 0 & 0 & 0 & b-2 & b-2 \end{array}\right) \xrightarrow{\frac{1}{b-2}R_4} \left(\begin{array}{cccc|c} 1 & 0 & 3 & 1 & 2 \\ 0 & a & 0 & -3 & 0 \\ 0 & 0 & a & a-2 & b-2 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right) \xrightarrow[R_2+3R_4]{R_3-(a-2)R_4} \left(\begin{array}{cccc|c} 1 & 0 & 3 & 1 & 2 \\ 0 & a & 0 & 0 & 3 \\ 0 & 0 & a & 0 & b-a \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

From row 2, we can see that the system is inconsistent if $a = 0$.

Otherwise, if $a \neq 0$, all 4 columns on the LHS will be pivot columns, and hence, the system has a unique solution

$$x_1 = \frac{4a-3b}{a}, \quad x_2 = \frac{3}{a}, \quad x_3 = \frac{b-a}{a}, \quad x_4 = 1.$$

Hard Example

(a) The system has no solution if $a = 0$ and $b \neq 2$.

(b) The system has a unique solution if $a \neq 0$ and $b \neq 2$. The unique solution is

$$x_1 = \frac{4a - 3b}{a}, \quad x_2 = \frac{3}{a}, \quad x_3 = \frac{b - a}{a}, \quad x_4 = 1.$$

(c) The system has infinitely many solutions with 1 parameter if $a \neq 0$ and $b = 2$. The general solution is

$$x_1 = 2 + \frac{2a - 6}{a}s, \quad x_2 = \frac{3}{a}s, \quad x_3 = \frac{(2 - a)}{a}s, \quad x_4 = s, \quad s \in \mathbb{R}.$$

(d) The system has infinitely many solutions with 2 parameters if $a = 0$ and $b = 2$. The general solution is

$$x_1 = 2 - 3t, \quad x_2 = s, \quad x_3 = t, \quad x_4 = 0, \quad s, t, \in \mathbb{R}.$$

Constructing a Linear System from the General Solution

Find a linear system with 2 equations such that

$$x_1 = 1 - 2s + t, \quad x_2 = s, \quad x_3 = t, \quad x_4 = 0, \quad s, t \in \mathbb{R}$$

is the general solution.

We will substitute $x_2 = s$ and $x_3 = t$ into the first equation and get

$$x_1 = 1 - 2x_2 + x_3.$$

Include the equation $x_4 = 0$, and changing it to a standard form, we obtain the linear system

$$\begin{cases} x_1 + 2x_2 - x_3 = 1 \\ x_4 = 0 \end{cases}.$$

The augmented matrix of the system is

$$\left(\begin{array}{cccc|c} 1 & 2 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right)$$

which is in reduced row-echelon form.

Another Example

Construct an augmented matrix with 3 variables and 4 equations such that it has the following general solution

$$x = 3 - 5s, \quad y = 2 + 2s, \quad z = s, \quad s \in \mathbb{R}.$$

Substituting $z = s$, into the first 2 equations, we get

$$x = 3 - 5z \text{ and } y = 2 + 2z \quad \Rightarrow \left\{ \begin{array}{rcrcrcr} x & & & + & 5z & = & 3 \\ & & y & - & 2z & = & 2 \end{array} \right.$$

However, we need 4 equations. Since we have used all the given information, the other 2 more equations should not contribute anymore information, that is, they should be derived from the first 2. We may consider taking multiples or adding the first 2 equations to obtain more equations. For example.

$$\left\{ \begin{array}{rcrcrcr} x & & & + & 5z & = & 3 \\ & & y & - & 2z & = & 2 \\ x & + & y & + & 3z & = & 5 \\ & & 2y & - & 4z & = & 4 \end{array} \right. \text{ which corresponds to the augmented matrix } \left(\begin{array}{ccc|c} 1 & 0 & 5 & 3 \\ 0 & 1 & -2 & 2 \\ 1 & 1 & 3 & 5 \\ 0 & 2 & -4 & 4 \end{array} \right)$$

MA1522: Linear Algebra for Computing

Chapter 2: Matrix Algebra

2.1 Definition and Special types of Matrices

Definition

Definition

A (real-valued) *matrix* is a rectangular array of (real) numbers

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = (a_{ij})_{m\times n} = (a_{ij})_{i=1}^{m}\,{}_{j=1}^{n}.,$$

where $a_{ij} \in \mathbb{R}$ are real numbers. The *size* of the matrix is said to be $m \times n$ (read as m by n), where m is the number of rows and n is the number of columns.

The numbers in the array are called *entries*. The *(i, j)-entry*, a_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$, is the number in the i-th row j-th column.

Remarks

- The size of a matrix is read as m by n. One should not multiply the numbers. For example, $\mathbf{A} = \begin{pmatrix} 1 & 2 & -1 & 5 & 10 \\ 3 & 5 & 13 & 23 & 41 \\ -7 & 2 & 0 & 0 & 11 \end{pmatrix}$ is a 3 by 5 matrix, not a size 15 matrix.

- There is a need to add a comma when labelling the (i, j)-entry, $a_{i,j}$, if there is ambiguity. For example, a_{123} may mean the $(1, 23)$-entry, in which case we label it is as $a_{1,23}$ instead, or the $(12, 3)$-entry, in which case we label it as $a_{12,3}$ instead.

Special types of Matrices

Vectors

A $n \times 1$ matrix is called a (column) *vector*, and a $1 \times n$ matrix is called a (row) *vector*.

Remark

If it is not specified whether the vector is a column or a row vector, by default we will assume it is a column vector.

Zero matrices

All entries equal 0, denoted as $\mathbf{0}_{m\times n}$. Not necessarily a square matrix.

Example

$$\mathbf{0}_{2\times 3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{0}_{3\times 3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{0}_{4\times 1} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{0}_{1\times 1} = \begin{pmatrix} 0 \end{pmatrix}.$$

Square matrices

Number of rows = number of columns

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

Example

$$\begin{pmatrix} 2 & -3 \\ 7 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 1 \end{pmatrix}.$$

Definition

i. A size $n \times n$ matrix is a square matrix of order n.

ii. The entries a_{ii}, $i = 1, 2, \dots, n$, (explicitly, $a_{11}, a_{22}, \dots, a_{nn}$) are called the diagonal entries of the (square) matrix.

Diagonal, Scalar, Identity matrices

1. Diagonal matrix $\mathbf{D} = (a_{ij})_n$, $a_{ij} = 0$ for $i \neq j$. Denote as $\mathbf{D} = diag(d_1, d_2, ..., d_n) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$.
2. Scalar matrix $\mathbf{C} = (a_{ij})$, $a_{ij} = \begin{cases} c & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$, $\mathbf{C} = diag(c, c, ..., c) = \begin{pmatrix} c & 0 & \cdots & 0 \\ 0 & c & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c \end{pmatrix}$.
3. Identity matrix $\mathbf{I} = (a_{ij})$, $a_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$, $\mathbf{I}_n = diag(1, 1, ..., 1) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$.

A scalar matrix can also be denoted as $\mathbf{C} = c\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. See later for definition of scalar multiplication.

Triangular matrices

Upper triangular $\mathbf{A} = (a_{ij})$, $a_{ij} = 0$ for all $i > j$: $\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & * \end{pmatrix}$

Strictly upper triangular $\mathbf{A} = (a_{ij})$, $a_{ij} = 0$ for all $i \geq j$: $\begin{pmatrix} 0 & * & \cdots & * \\ 0 & 0 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$

Lower triangular $\mathbf{A} = (a_{ij})$, $a_{ij} = 0$ for all $i < j$: $\begin{pmatrix} * & 0 & \cdots & 0 \\ * & * & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{pmatrix}$

Strictly lower triangular $\mathbf{A} = (a_{ij})$, $a_{ij} = 0$ for all $i \leq j$: $\begin{pmatrix} 0 & 0 & \cdots & 0 \\ * & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & 0 \end{pmatrix}$

Symmetric matrices

$\mathbf{A} = (a_{ij})_n, \ a_{ij} = a_{ji}.$

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

Example

$$\begin{pmatrix} 1 & 2 & b \\ a & 0 & 4 \\ 3 & c & d \end{pmatrix} \text{ is symmetric } \Leftrightarrow a = 2, b = 3, c = 4, d \in \mathbb{R}.$$

2.2 Matrix Algebra

Equality

Two matrices are *equal* if they have the same size and their corresponding entries are equal;

$\mathbf{A} = (a_{ij})_{n\times m}$ and $\mathbf{B} = (b_{ij})_{k\times l}$ are equal if and only if $n = k$, $m = l$, and $a_{ij} = b_{ij}$ for all $i = 1, ..., n$, $j = 1, ..., m$.

Example

1. $\begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix} \neq \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$ for any $a, b, c, d, e, f \in \mathbb{R}$ since the matrices do not have the same sizes.

2. $\begin{pmatrix} 1 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Leftrightarrow a = b = 1, c = 3, d = 2.$

Matrix Addition and Scalar Multiplication

Definition

1. Scalar multiplication: $c \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{pmatrix}.$
2. Matrix addition:

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

Example

1. $5 \begin{pmatrix} 6 & 1 & -1 \\ 2 & -4 & 3 \\ 4 & 9 & -11 \end{pmatrix} = \begin{pmatrix} 30 & 5 & -5 \\ 10 & -20 & 15 \\ 20 & 45 & -55 \end{pmatrix}$
2. $\begin{pmatrix} 1 & 2 & -3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 2 & 1 & 5 \\ 1 & -3 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 & 2 \\ 5 & 2 & 8 \end{pmatrix}$

Remark

1. Matrix addition is only defined between matrices of the same size.
2. $-\mathbf{A} = (-1)\mathbf{A}$.
3. Matrix subtraction is defined to be the addition of a negative multiple of another matrix,

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B}.$$

Properties of Matrix Addition and Scalar Multiplication

Theorem

For matrices $\mathbf{A} = (a_{ij})_{m \times n}$, $\mathbf{B} = (b_{ij})_{m \times n}$, $\mathbf{C} = (c_{ij})_{m \times n}$, *and real numbers* $a, b \in \mathbb{R}$,

(i) *(Commutative)* $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$,

(ii) *(Associative)* $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$,

(iii) *(Additive identity)* $\mathbf{0}_{m \times n} + \mathbf{A} = \mathbf{A}$,

(iv) *(Additive inverse)* $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}_{m \times n}$,

(v) *(Distributive law)* $a(\mathbf{A} + \mathbf{B}) = a\mathbf{A} + a\mathbf{B}$,

(vi) *(Scalar addition)* $(a + b)\mathbf{A} = a\mathbf{A} + b\mathbf{A}$,

(vii) *(Associative)* $(ab)\mathbf{A} = a(b\mathbf{A})$,

(viii) *If* $a\mathbf{A} = \mathbf{0}_{m \times n}$, *then either* $a = 0$ *or* $\mathbf{A} = \mathbf{0}$.

Remarks

- The proof of the theorem can be found in the appendix.
- However, intuitively, the properties follow from the properties of addition and multiplication of real numbers, since scalar multiplication and matrix addition is defined entries-wise.
- By the associativity of matrix addition, if $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_k$ are k matrices of the same size, we simply write the sum as
$$\mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_k,$$
without the parenthesis (brackets).

Matrix Multiplication

$$\mathbf{AB} = (a_{ij})_{m\times p}(b_{ij})_{p\times n} = (\textstyle\sum_{k=1}^{p} a_{ik}b_{kj})_{m\times n}$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} & a_{11}b_{14} + a_{12}b_{24} + a_{13}b_{34} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} & a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} & a_{21}b_{14} + a_{22}b_{24} + a_{23}b_{34} \end{pmatrix}$$

That is, the (i, j)-entry of the product **AB** is the sum of the product of the entries in the i-th row of **A** with the j-th column of **B**.

Remark

1. For **AB** to be defined, the number of columns of **A** must agree with the number of rows of **B**. The resultant matrix has the same number of rows as **A**, and the same number of columns as **B**.

$$(m \times p)(p \times n) = (m \times n).$$

2. Matrix multiplication is not commutative,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \neq \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

3. The order of the product matter.
   i. If we multiply **A** to the left of **B**, we are pre-multiplying **A** to **B**.
   ii. If we multiply **A** to the right of **B**, we are post-multiplying **A** to **B**.

Properties of Matrix Multiplication

Theorem

(i) *(Associative)* $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

(ii) *(Left distributive law)* $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

(iii) *(Right distributive law)* $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$.

(iv) *(Commute with scalar multiplication)* $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B})$.

(v) *(Multiplicative identity) For any* $m \times n$ *matrix* $\mathbf{A}$, $\mathbf{I}_m\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{I}_n$.

(vi) *(Nonzero Zero divisor) There exists* $\mathbf{A} \neq \mathbf{0}_{m\times p}$ *and* $\mathbf{B} \neq \mathbf{0}_{p\times n}$ *such that* $\mathbf{AB} = \mathbf{0}_{m\times n}$.

(vii) *(Zero matrix) For any* $m \times n$ *matrix* $\mathbf{A}$, $\mathbf{A}\mathbf{0}_{n\times p} = \mathbf{0}_{m\times p}$ *and* $\mathbf{0}_{p\times m}\mathbf{A} = \mathbf{0}_{p\times n}$

Remarks

- To not overcrowd the slide, we left out the sizes of the matrices in the theorem, assuming that the matrices have the appropriate sizes for the operations to be well-defined. See the appendix for the detail statements.

- The proof of the theorem can be found in the appendix.

- By the associativity of matrix multiplication, if $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_k$ are k matrices of the right sizes such that their product is well-defined, we write it as

$$\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_k,$$

  without the parenthesis (brackets).

Zero Divisors

Consider

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then

$$\mathbf{AB} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Powers of Square Matrices

Definition

Define the power of square matrices inductively as such.

(i) $\mathbf{A}^0 = \mathbf{I}$,

(ii) $\mathbf{A}^n = \mathbf{A}\mathbf{A}^{n-1}$, for $n \geq 1$.

That is, $\mathbf{A}^n$ is $\mathbf{A}$ multiplied to itself n times, for $n \geq 2$. It follows that $\mathbf{A}^n\mathbf{A}^m = \mathbf{A}^{n+m}$ for positive integers m, n.

Example

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^3 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}$$

Note that powers of a matrix can only be defined for square matrices.

Challenge

Show that for diagonal matrices $\mathbf{A}$ and $\mathbf{B}$, $(\mathbf{A}+\mathbf{B})^2 = \mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2$.

Transpose

Let $\mathbf{A} = (a_{ij})$ be a $m \times n$ matrix. The *transpose* of $\mathbf{A}$, denoted as $\mathbf{A}^T$, is the $n \times m$ matrix whose (i,j)-entry is the (j,i)-entry of $\mathbf{A}$, $\mathbf{A}^T = (b_{ij})_{n\times m}$, $b_{ij} = a_{ji}$.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}_{m\times n} \qquad \mathbf{A}^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}_{n\times m}$$

Properties of Transpose

Theorem

(i) $(\mathbf{A}^T)^T = \mathbf{A}$.

(ii) $(c\mathbf{A})^T = c\mathbf{A}^T$.

(iii) $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.

(iv) $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T\mathbf{A}^T$.

The proof is left as an exercise

The transpose provides an alternative definition of symmetric matrix. A square matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}^T$.

Challenge

Is it true that if $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices of the same order, then so is $\mathbf{AB}$?

2.3 Linear System and Matrix Equation

Matrix Equation

A linear system in standard form

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

Can be expressed as a *matrix equation*

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \qquad \mathbf{Ax} = \mathbf{b}$$

Here $\mathbf{A} = (a_{ij})_{m\times n}$ is called the *coefficient matrix*, $\mathbf{x} = (x_i)_{n\times 1}$ the *variable vector*, and $\mathbf{b} = (b_i)_{m\times 1}$ the *constant vector*.

Properties of Homogeneous Linear System

Recall that a linear system is homogeneous if it has the following corresponding matrix equation

$$\mathbf{Ax} = \mathbf{0},$$

for some $m \times n$ matrix $\mathbf{A}$, a variable n-vector $\mathbf{x}$, and the $m \times 1$ zero matrix (or zero m-vector) $\mathbf{0} = \mathbf{0}_{m\times 1}$.

Theorem
A homogeneous linear system is always consistent.

Proof.
Write the homogeneous linear system as $\mathbf{Ax} = \mathbf{0}$ for some $m \times n$ matrix $\mathbf{A}$. Then

$$\mathbf{A0} = \mathbf{0},$$

that is, the $m \times 1$ zero matrix is a solution to the system. $\square$

Properties of Homogeneous Linear System

Definition

The zero solution is called the *trivial solution*. If $\mathbf{x} \neq \mathbf{0}$ is a nonzero solution to the homogeneous system, it is called a *nontrivial solution*.

Theorem

A homogeneous linear system has infinitely many solutions if and only if it has a nontrivial solution.

Proof.

($\Rightarrow$) If the system has infinitely many solutions, it must surely have a nontrivial solution.

($\Leftarrow$) Suppose now $\mathbf{u} \neq \mathbf{0}$ is a nontrivial solution to the homogeneous system $\mathbf{Ax} = \mathbf{0}$, that is, $\mathbf{Au} = \mathbf{0}$. Now for any real number $s \in \mathbb{R}$, since scalar multiplication commutes with matrix multiplication,

$$\mathbf{A}(s\mathbf{u}) = s(\mathbf{Au}) = s\mathbf{0} = \mathbf{0},$$

that is, $s\mathbf{u}$ is a solution to the homogeneous linear system too. Hence, the system has infinitely many solutions. $\square$

Solutions to Homogeneous and Non-homogeneous Linear System

Lemma

Let $\mathbf{v}$ *be a particular solution* $\mathbf{A}\mathbf{x} = \mathbf{b}$, *and* $\mathbf{u}$ *be a particular solution to the homogeneous system* $\mathbf{A}\mathbf{x} = \mathbf{0}$ *with the same coefficient matrix* $\mathbf{A}$. *Then* $\mathbf{v} + \mathbf{u}$ *is also a solution to* $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Proof.

By hypothesis, $\mathbf{A}\mathbf{v} = \mathbf{b}$ and $\mathbf{A}\mathbf{u} = \mathbf{0}$. Hence, by the distribution properties of matrix multiplication,

$$\mathbf{A}(\mathbf{v} + \mathbf{u}) = \mathbf{A}\mathbf{v} + \mathbf{A}\mathbf{u} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

$\square$

Lemma

Suppose $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *are solutions to the linear system* $\mathbf{A}\mathbf{x} = \mathbf{b}$. *Then* $\mathbf{v}_1 - \mathbf{v}_2$ *is a solution to the homogeneous linear system* $\mathbf{A}\mathbf{x} = \mathbf{0}$ *with the same coefficient matrix.*

Proof.

By hypothesis, $\mathbf{A}\mathbf{v}_1 = \mathbf{b}$, $\mathbf{A}\mathbf{v}_2 = \mathbf{b}$. Hence,

$$\mathbf{A}(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{A}\mathbf{v}_1 - \mathbf{A}\mathbf{v}_2 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

$\square$

Challenge

Let $\mathbf{v}$ be a particular solution to a non-homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{b} \neq \mathbf{0}$.

Show that

$$\mathbf{v} + s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k, \quad s_1, s_2, \ldots, s_k \in \mathbb{R}$$

is the general solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ if and only if

$$s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k, \quad s_1, s_2, \ldots, s_k \in \mathbb{R}$$

is the general solution to the homogeneous linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$.

Introduction to Submatrices

Let $\mathbf{A}$ be an $m \times n$ matrix,

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

Rows of $\mathbf{A}$:

$$\begin{gathered} \mathbf{r}_1 = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix} \\ \mathbf{r}_2 = \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix} \\ \vdots \\ \mathbf{r}_m = \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \end{gathered}$$

Columns of $\mathbf{A}$:

$$\mathbf{c}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{m1} \end{pmatrix} \quad \mathbf{c}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \cdots \\ a_{m2} \end{pmatrix} \quad \ldots \quad \mathbf{c}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdots \\ a_{mn} \end{pmatrix}$$

The rows and columns of a matrix are examples of submatrices of a matrix.

Submatrix

Definition

A $p \times q$ *submatrix* of an $m \times n$ matrix $\mathbf{A}$, $p \leq m$, $q \leq n$, is formed by taking a $p \times q$ block of the entries of the matrix $\mathbf{A}$.

Example

Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 5 & 7 & 1 \\ 2 & 4 & 6 & 1 & 2 \\ 1 & 2 & 2 & 1 & 3 \end{pmatrix}.$$

(i) Each row is a 1 by 5 submatrix, each column is a 3 by 1 submatrix.

(ii) $\begin{pmatrix} 2 & 4 & 6 \\ 1 & 2 & 2 \end{pmatrix}$ is a 2 by 3 submatrix obtained by taking rows 2 and 3, and columns 1 to 3.

(iii) $\begin{pmatrix} 5 & 7 \\ 6 & 1 \\ 2 & 1 \end{pmatrix}$ is a 3 by 2 submatrix obtained by taking rows 1 to 3, and columns 3 and 4.

Block Multiplication

Theorem

Let $\mathbf{A}$ *be an* $m \times p$ *matrix and* $\mathbf{B}$ *a* $p \times n$ *matrix. Let* $\mathbf{A}_1$ *be a* $(m_2 - m_1 + 1) \times p$ *submatrix of* $\mathbf{A}$ *obtained by taking rows* m_1 *to* m_2*, and* $\mathbf{b}_1$ *a* $p \times (n_2 - n_1 + 1)$ *submatrix of* $\mathbf{B}$ *obtained by taking columns* n_1 *to* n_2*. Then the product* $\mathbf{A}_1\mathbf{B}_1$ *is a* $(m_2 - m_1 + 1) \times (n_2 - n_1 + 1)$ *submatrix of* $\mathbf{AB}$ *obtained by taking rows* m_1 *to* m_2 *and columns* n_1 *to* n_2*.*

The proof is left as an exercise. We call this *block multiplication*.

Block Multiplication

In particular, let $\mathbf{b}_j$ be the j-th column of $\mathbf{B}$. Then

$$\mathbf{AB} = \mathbf{A}\begin{pmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n\end{pmatrix} = \begin{pmatrix}\mathbf{Ab}_1 & \mathbf{Ab}_2 & \cdots & \mathbf{Ab}_n\end{pmatrix}.$$

That is, the j-th column of the product $\mathbf{AB}$ is the product of $\mathbf{A}$ with the j-th column of $\mathbf{B}$.

Also, if $\mathbf{a}_i$ is the i-th row of $\mathbf{A}$, then

$$\mathbf{AB} = \begin{pmatrix}\mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m\end{pmatrix}\mathbf{B} = \begin{pmatrix}\mathbf{a}_1\mathbf{B} \\ \mathbf{a}_2\mathbf{B} \\ \vdots \\ \mathbf{a}_m\mathbf{B}\end{pmatrix}.$$

That is, the i-th row of the product $\mathbf{AB}$ is the product of the i-th row of $\mathbf{A}$ with $\mathbf{B}$.

Solving Matrix Equations

Let $\mathbf{A} = \begin{pmatrix} 3 & 2 & -1 \\ 5 & -1 & 3 \\ 2 & 1 & -1 \end{pmatrix}$. Find a 3×3 matrix $\mathbf{X}$ such that

$$\mathbf{AX} = \begin{pmatrix} 3 & 2 & -1 \\ 5 & -1 & 3 \\ 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 3 & 1 & 0 \end{pmatrix}.$$

By block multiplication, we are solving for the 3 linear systems

$$\mathbf{A} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{A} \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{A} \begin{pmatrix} x_3 \\ y_3 \\ z_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Combining Augmented Matrices

Solve the following 3 linear systems

$$\begin{array}{rcrcrcl} 3x & + & 2y & - & z & = & a \\ 5x & - & y & + & 3z & = & b \\ 2x & + & y & - & z & = & c \end{array}, \text{ for } \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Matrix equation:

$$\begin{pmatrix} 3 & 2 & -1 \\ 5 & -1 & 3 \\ 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 3 & 1 & 0 \end{pmatrix}$$

Augmented matrix:

$$\left(\begin{array}{ccc|c|c|c} 3 & 2 & -1 & 1 & 2 & 1 \\ 5 & -1 & 3 & 2 & 1 & 1 \\ 2 & 1 & -1 & 3 & 1 & 0 \end{array}\right)$$

Solve the 3 linear system simultaneously.

Combining Augmented Matrices

In general: p linear systems with the same coefficient matrix $\mathbf{A} = (a_{ij})_{m\times n}$, for $k = 1, \dots, p$,

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_{1k} \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_{2k} \\ & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_{mk} \end{array}\right.$$

Combined augmented matrix:

$$\left(\begin{array}{cccc|c|c|c|c} a_{11} & a_{12} & \cdots & a_{1n} & b_{11} & b_{12} & & b_{1p} \\ a_{21} & a_{22} & \cdots & a_{2n} & b_{21} & b_{22} & & b_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_{m1} & b_{m2} & & b_{mp} \end{array}\right)$$

2.4 Inverse of Matrices

Introduction

1. Suppose $ab = ac$ for some $a, b, c \in \mathbb{R}$ with $a \neq 0$, then $b = c$.
2. This is because if $a \neq 0$, $\frac{1}{a}$ exists, and $a\frac{1}{a} = 1$.
3. Hence, multiplying both sides of $ab = ac$ with $\frac{1}{a}$, we have $b = 1 \times b = \frac{1}{a}ab = \frac{1}{a}ac = 1 \times c = c$.
4. This is used to solve $ax = b$ for some $a, b \in \mathbb{R}$. We can conclude that $x = \frac{b}{a}$.
5. Ideally, we want to apply the same idea to solve a linear system,
$$\mathbf{Ax} = \mathbf{b} \quad \Rightarrow \quad \mathbf{x} = \frac{\mathbf{b}}{\mathbf{A}}.$$
6. Idea: Try to define $\frac{\mathbf{1}}{\mathbf{A}}$.

Introduction

1. To define $\frac{\mathbf{1}}{\mathbf{A}}$, we need the matrix equivalent of 1, since $\frac{1}{a}$ is defined such that $\frac{1}{a}a = 1$.

2. 1 is the multiplicative identity of the real numbers, $1 \times a = a$ for any $a \in \mathbb{R}$.

3. The matrix multiplicative identity is the identity matrix, $\mathbf{I}_n$,

$$\mathbf{IA} = \mathbf{A} = \mathbf{AI}$$

   for any matrix $\mathbf{A}$.

4. So, define $\frac{\mathbf{1}}{\mathbf{A}}$ as the matrix such that $\frac{\mathbf{1}}{\mathbf{A}}\mathbf{A} = \mathbf{I}$.

Problem 1

For matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ with the appropriate sizes, if $\mathbf{AB} = \mathbf{AC}$ and $\mathbf{A} \neq \mathbf{0}$ is not the zero matrix, can we conclude that $\mathbf{B} = \mathbf{C}$?

No. For example,

$$\begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 6 & 8 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 2 & 5 \\ 3 & 4 \end{pmatrix},$$

but $\begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix} \neq \begin{pmatrix} 2 & 5 \\ 3 & 4 \end{pmatrix}$.

Here although $\mathbf{A}$ is not the zero matrix, we do not have the cancellation law.

Problem 2

Recall that matrix multiplication is not commutative, $\mathbf{AB} \neq \mathbf{BA}$. Suppose a matrix $\frac{\mathbf{1}}{\mathbf{A}}$ exists such that $\frac{\mathbf{1}}{\mathbf{A}}\mathbf{A} = \mathbf{I}$. Is it true that $\mathbf{A}\frac{\mathbf{1}}{\mathbf{A}} = \mathbf{I}$?

No. For example,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}, \quad \text{but} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq \mathbf{I}$$

Conclusion

- So, it seems like our dream of solving a linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad \mathbf{x} = \frac{\mathbf{b}}{\mathbf{A}}$$

  is IMPOSSIBLE!

- Except, it is possible in some cases.

- Restrict our attention only to square matrices.

Inverse of Square Matrices

Definition

A $n \times n$ square matrix $\mathbf{A}$ is *invertible* if there exists a matrix $\mathbf{B}$ such that

$$\mathbf{AB} = \mathbf{I}_n = \mathbf{BA}.$$

A matrix is said to be *non-invertible* otherwise.

Example

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

Remarks

1. By definition, for a square matrix $\mathbf{A}$ to be invertible, there must be a $\mathbf{B}$ such that $\mathbf{BA} = \mathbf{I}_n$ AND $\mathbf{AB} = \mathbf{I}_n$ simultaneously.
2. This necessitates $\mathbf{B}$ to also be a $n \times n$ square matrix.
3. Only square matrices are invertible. If $\mathbf{A}$ is a $n \times m$ matrix with $n \neq m$, then for $\mathbf{AB} = \mathbf{I}_n$, $\mathbf{B}$ must be of size $m \times n$, and hence $\mathbf{BA} \neq \mathbf{I}_n$.
4. Hence, all non-square matrices are non-invertible. In fact, we will see later that if $\mathbf{A}$ is a non-square matrix such that $\mathbf{BA} = \mathbf{I}$ (or $\mathbf{AB} = \mathbf{I}$, respectively) for some matrix $\mathbf{B}$, then there exists no matrix $\mathbf{C}$ such that $\mathbf{AC} = \mathbf{I}$ (or $\mathbf{CA} = \mathbf{I}$, respectively).

Uniqueness of inverse

Theorem

If $\mathbf{B}$ *and* $\mathbf{C}$ *are both inverses of a square matrix* $\mathbf{A}$, *then* $\mathbf{B} = \mathbf{C}$.

Proof.

Suppose $\mathbf{B}$ and $\mathbf{C}$ are such that $\mathbf{AB} = \mathbf{I} = \mathbf{BA}$ and $\mathbf{AC} = \mathbf{I} = \mathbf{CA}$. Then by associativity of matrix multiplication,

$$\mathbf{B} = \mathbf{BI} = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{IC} = \mathbf{C}.$$

$\square$

So since the inverse is unique, we can denote the inverse of an invertible matrix $\mathbf{A}$ by $\mathbf{A}^{-1}$ and call it the *inverse* of $\mathbf{A}$.

Question: Is the identity matrix $\mathbf{I}$ invertible? If it is, what is it's inverse?

Non-Invertible Square Matrix

Not all matrices are invertible. For example, consider the matrix $\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$.

It is not invertible since for any order 2 square matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ a+c & b+d \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Definition

A non-invertible square matrix is called a *singular* matrix.

Remark

Some textbook use the term singular interchangeably with non-invertible, that is, they do not insist that singular matrices are square matrices.

Inverse of 2 by 2 Square Matrices

Theorem

A 2×2 square matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. In this case, the inverse is given by

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

We will proof the claim that $\mathbf{A}$ is invertible if and only if $ad - bc \neq 0$ later. The verification that it's inverse is $\mathbf{A}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ is left as an exercise for the readers. The formula above is known as the adjoint formula for inverse.

Cancellation law for matrices

Theorem

Let $\mathbf{A}$ *be an invertible matrix of order* n*.*

(i) *(Left cancellation) If* $\mathbf{B}$ *and* $\mathbf{C}$ *are* $n \times m$ *matrices with* $\mathbf{AB} = \mathbf{AC}$*, then* $\mathbf{B} = \mathbf{C}$*.*

(ii) *(Right cancellation) If* $\mathbf{B}$ *and* $\mathbf{C}$ *are* $m \times n$ *matrices with* $\mathbf{BA} = \mathbf{CA}$*, then* $\mathbf{B} = \mathbf{C}$*.*

Proof.

(i)

$$\mathbf{AB} = \mathbf{AC} \quad \Rightarrow \quad \mathbf{B} = \mathbf{A}^{-1}\mathbf{AB} = \mathbf{A}^{-1}\mathbf{AC} = \mathbf{C}.$$

(ii)

$$\mathbf{BA} = \mathbf{CA} \quad \Rightarrow \quad \mathbf{B} = \mathbf{BAA}^{-1} = \mathbf{CAA}^{-1} = \mathbf{C}.$$

$\square$

Caution

If $\mathbf{AB} = \mathbf{CA}$, we cannot conclude that $\mathbf{B} = \mathbf{C}$.

Invertibility and Linear System

Theorem

Suppose $\mathbf{A}$ *is an* $n \times n$ *invertible square matrix. Then for any* $n \times 1$ *vector* $\mathbf{b}$, $\mathbf{Ax} = \mathbf{b}$ *has a unique solution.*

Proof.

There are two claims in the theorem; it claims that the system $\mathbf{Ax} = \mathbf{b}$ is consistent, and that the solution is unique. Firstly, we will check that $\mathbf{u} = \mathbf{A}^{-1}\mathbf{b}$ is a solution. Indeed,

$$\mathbf{Au} = \mathbf{A}(\mathbf{A}^{-1}\mathbf{b}) = \mathbf{b}.$$

Now suppose $\mathbf{v}$ is also a solution, that is, $\mathbf{Av} = \mathbf{b}$. Then

$$\mathbf{Av} = \mathbf{b} = \mathbf{Au}.$$

By the cancellation law, we have $\mathbf{v} = \mathbf{u} = \mathbf{A}^{-1}\mathbf{b}$. Hence, $\mathbf{A}^{-1}\mathbf{b}$ is the unique solution to the system. $\square$

Invertibility and Linear System

Corollary

Suppose $\mathbf{A}$ *is invertible. Then the trivial solution is the only solution to the homogeneous system* $\mathbf{A}\mathbf{x} = \mathbf{0}$.

Proof.

This follows immediately from the previous theorem by letting $\mathbf{b} = \mathbf{0}$. $\square$

Exercise

Verify that the homogeneous system

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

has only the trivial solution.

Remark

1. The theorem and its corollary on invertibility and linear system are actually equivalent statements. That is,
   (i) a square matrix $\mathbf{A}$ is invertible if and only if $\mathbf{Ax} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$, and

   (ii) a square matrix $\mathbf{A}$ is invertible if and only if the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.

2. In fact, one can observe that $\mathbf{Ax} = \mathbf{b}$ has a unique solution if and only if the reduced row-echelon form of the augmented matrix $\left(\mathbf{A} \mid \mathbf{b}\right)$ is $\left(\mathbf{I} \mid \mathbf{A}^{-1}\mathbf{b}\right)$.

3. This also hint towards the fact that $\mathbf{A}$ is invertible if and only if the reduced row-echelon form of $\mathbf{A}$ is the identity matrix.

4. The proofs will be given in the next section.

Algorithm to Computing Inverse

Suppose $\mathbf{A}$ is an invertible $n \times n$ matrix. By uniqueness of the inverse, there must be a unique solution to

$$\mathbf{AX} = \mathbf{I}.$$

By block multiplication, we are solving the augmented matrix

$$\left(\mathbf{A} \mid \mathbf{I}\right) \xrightarrow{RREF} \left(\mathbf{I} \mid \mathbf{A}^{-1}\right).$$

Remarks

1. In the algorithm to finding the inverse, we are solving to $\mathbf{AX} = \mathbf{I}$. Technically, we are solving for a *right inverse*.

2. Can we guarantee that the solution is also a *left inverse*? That is, if $\mathbf{AB} = \mathbf{I}$, can we be sure that $\mathbf{BA} = \mathbf{I}$ too?

Properties of Inverses

Theorem

Let $\mathbf{A}$ be an invertible matrix of order n.

(i) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.

(ii) *For any nonzero real number $a \in \mathbb{R}$, $(a\mathbf{A})$ is invertible with inverse $(a\mathbf{A})^{-1} = \frac{1}{a}\mathbf{A}^{-1}$.*

(iii) $\mathbf{A}^T$ *is invertible with inverse* $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.

(iv) *If $\mathbf{B}$ is an invertible matrix of order n, then $(\mathbf{AB})$ is invertible with inverse $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.*

Partial Proof.

The proof of (i) to (iii) is left as a exercise. For (iv),

$$(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I} = \mathbf{AA}^{-1} = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = (\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}).$$

$\square$

Remark

1. The precise statement for (iii) is that as follows.

   A square matrix $\mathbf{A}$ is invertible if and only if its transpose $\mathbf{A}^T$ is. In this case, the inverse of its transpose is the transpose of its inverse, $(\mathbf{A}^{-1})^T$.

2. **Caution**: $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \neq \mathbf{A}^{-1}\mathbf{B}^{-1}$.

3. If $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_k$ are invertible, the product $\mathbf{A}_1\mathbf{A}_s \cdots \mathbf{A}_k$ is invertible with $(\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_k)^{-1} = \mathbf{A}_k^{-1} \cdots \mathbf{A}_2^{-1}\mathbf{A}_1^{-1}$ if $\mathbf{A}_i$ is an invertible matrix for $i = 1, ..., k$.

4. The *negative power* of an invertible matrix is defined to be

   $$\mathbf{A}^{-n} = (\mathbf{A}^{-1})^n$$

   for any $n > 0$.

2.5 Elementary Matrices

Elementary Matrices

Definition

A square matrix $\mathbf{E}$ of order n is called an *elementary matrix* if it can be obtained from the identity matrix $\mathbf{I}_n$ by performing a single elementary row operation

$$\mathbf{I}_n \xrightarrow{r} \mathbf{E},$$

where r is an elementary row operation. The elementary row operation is said to be the row operation corresponding to the elementary matrix.

Example

1. $\mathbf{I}_4 \xrightarrow{R_2+3R_4} \mathbf{E} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

2. $\mathbf{I}_4 \xrightarrow{R_1 \leftrightarrow R_3} \mathbf{E} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$

3. $\mathbf{I}_4 \xrightarrow{3R_2} \mathbf{E} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$

Elementary Matrices and Elementary Row Operations

Let $\mathbf{A}$ be an $n \times m$ matrix and let $\mathbf{E}$ be the $n \times n$ elementary matrix corresponding to the elementary row operation r. Then the product $\mathbf{EA}$ is the resultant of performing the row operation r on $\mathbf{A}$,

$$\mathbf{A} \xrightarrow{r} \mathbf{EA}.$$

That is, performing elementary row operations is equivalent to premultiplying by the corresponding elementary matrix.

Example

1. $\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & -1 \\ 2 & 1 & 4 & 2 \end{pmatrix} \xrightarrow{R_2+2R_1} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 3 & 4 & 5 & -1 \\ 2 & 1 & 4 & 2 \end{pmatrix}$

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

Row Equivalent Matrices

Recall that matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be row equivalent if $\mathbf{B}$ can be obtained from $\mathbf{A}$ by performing a series of elementary row operations.

Suppose now $\mathbf{B}$ is row equivalent to $\mathbf{A}$,

$$\mathbf{A} \xrightarrow{r_1}\xrightarrow{r_2} \cdots \xrightarrow{r_k} \mathbf{B}.$$

Let $\mathbf{E}_i$ be the elementary matrix corresponding to the row operation r_i, for $i = 1, 2, ..., k$. Then

$$\mathbf{B} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}.$$

This follows from the previous discussion,

$$\mathbf{A} \xrightarrow{r_1} \mathbf{E}_1\mathbf{A} \xrightarrow{r_2} \mathbf{E}_2\mathbf{E}_1\mathbf{A} \xrightarrow{r_3} \cdots \xrightarrow{r_k} \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{B}.$$

Inverse of Elementary Matrices

Recall that every elementary row operations has an inverse, and that performing a row operation is equivalent to premultiplying by the corresponding elementary matrix.

Consider now a row operation r, with its corresponding elementary matrix, $\mathbf{E}$. Let r' be the reverse of the row operation r, and $\mathbf{E}'$ its corresponding elementary matrix. By definition, $\mathbf{I} \xrightarrow{r} \mathbf{E}$. Now if we apply r' to $\mathbf{E}$, since it is the reverse of r, we should get back the identity matrix,

$$\mathbf{I} \xrightarrow{r} \mathbf{E} \xrightarrow{r'} \mathbf{I}.$$

Hence, we have

$$\mathbf{I} = \mathbf{E}'\mathbf{E}\mathbf{I} = \mathbf{E}'\mathbf{E}.$$

Similarly, $\mathbf{I} \xrightarrow{r'} \mathbf{E}' \xrightarrow{r} \mathbf{I}$ tells us that

$$\mathbf{I} = \mathbf{E}\mathbf{E}'\mathbf{I} = \mathbf{E}\mathbf{E}'.$$

This shows that $\mathbf{E}$ is invertible with inverse $\mathbf{E}'$, which is also an elementary matrix.

Inverse of Elementary Matrices

Theorem

Every elementary matrices $\mathbf{E}$ *are invertible. The inverse* $\mathbf{E}^{-1}$ *is the elementary matrix corresponding to the reverse of the row operation corresponding to* $\mathbf{E}$.

(i)

$$\mathbf{I}_n \xrightarrow{R_i+cR_j} \mathbf{E} \xrightarrow{R_i-cR_j} \mathbf{I}_n \quad \Rightarrow \quad \mathbf{E} : R_i + cR_j, \ \mathbf{E}^{-1} : R_i - cR_j.$$

(ii)

$$\mathbf{I}_n \xrightarrow{R_i \leftrightarrow R_j} \mathbf{E} \xrightarrow{R_i \leftrightarrow R_j} \mathbf{I}_n \quad \Rightarrow \quad \mathbf{E} : R_i \leftrightarrow R_j, \ \mathbf{E}^{-1} : R_i \leftrightarrow R_j.$$

(iii)

$$\mathbf{I}_n \xrightarrow{cR_i} \mathbf{E} \xrightarrow{\frac{1}{c}R_i} \mathbf{I}_n \quad \Rightarrow \quad \mathbf{E} : cR_i, \ \mathbf{E}^{-1} : \frac{1}{c}R_i.$$

2.6 Equivalent Statements for Invertibility

Introduction

This section will be mainly proving important equivalent statements of invertibility. That is, as long as any one of these statements holds, we know that all the other statements will hold true too. The equivalent statements of invertibility is like a junction, where knowing that any of the statements opens you up to all the other statements, of which one of them might be useful in solving the problem you have at hand.

Before we get lost in a sea of theorems and proofs, we will be illustrating the statements with some examples.

Elementary Matrices and Inverse

Theorem

If $\mathbf{A} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1$ is a product of elementary matrices, then $\mathbf{A}$ is invertible.

Proof.

This follows from the fact that the product of invertible matrices is invertible. Hence, if a square matrix $\mathbf{A}$ can be written as a product of elementary matrices, then since elementary matrices are invertible, $\mathbf{A}$ invertible. $\square$

Corollary

If the reduce row-echelon form of $\mathbf{A}$ is the identity matrix, then $\mathbf{A}$ is invertible.

Proof.

By the hypothesis, $\mathbf{A}$ is row equivalent to $\mathbf{I}$. This means that there are elementary matrices $\mathbf{E}_1, \mathbf{E}_2, ..., \mathbf{E}_k$ such that

$$\mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{I}.$$

By premultiplying the inverse of the elementary matrices, we have

$$\mathbf{A} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1}\mathbf{I} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1}.$$

However, since the inverse of elementary matrices are elementary matrices, this shows that $\mathbf{A}$ is a product of elementary matrices, and hence invertible. $\square$

Equivalent Statements for Invertibility

Theorem

A square matrix $\mathbf{A}$ is invertible if and only if the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.

Proof.

($\Rightarrow$) We have shown that if $\mathbf{A}$ is invertible, then the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.

($\Leftarrow$) Now suppose the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution. This necessitates the reduced row-echelon form of $\mathbf{A}$ to be the identity matrix. For otherwise, the reduced row-echelon form must have a non-pivot column, which then implies that the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has infinitely many solutions, a contradiction. Hence, by the previous corollary, $\mathbf{A}$ is invertible. $\square$

The proof also shows that if $\mathbf{A}$ is invertible, then its reduced row-echelon form is the identity matrix.

Theorem

A square matrix $\mathbf{A}$ is invertible if and only if it's reduced row-echelon form is the identity matrix.

Theorem

A square matrix $\mathbf{A}$ is invertible if and only if it is a product of elementary matrices.

The proofs of theorems are left as exercises.

Left and Right Inverses

Definition

Let $\mathbf{A}$ be a $n \times m$ matrix.

(i) A $m \times n$ matrix $\mathbf{B}$ is said to be a *left inverse* of $\mathbf{A}$ if $\mathbf{BA} = \mathbf{I}_m$, where $\mathbf{I}_m$ is the $m \times m$ identity matrix.

(ii) A $m \times n$ matrix $\mathbf{B}$ is said to be a *right inverse* of $\mathbf{A}$ if $\mathbf{AB} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

Observe that $\mathbf{B}$ is a left inverse of $\mathbf{A}$ if and only if $\mathbf{A}$ is a right inverse of $\mathbf{B}$.

Example

1. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ is a left inverse of $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$; $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ is a right inverse of $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

2. The inverse of a square matrix $\mathbf{A}$ is both a (the) left and right inverse, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^{-1}$.

Equivalent Statements for Invertibility

Theorem

A square matrix $\mathbf{A}$ is invertible if and only if it has a left inverse.

Proof.

($\Rightarrow$) An inverse is a left inverse.

($\Leftarrow$) Suppose now $\mathbf{B}$ is a left inverse of $\mathbf{A}$, $\mathbf{BA} = \mathbf{I}$. Consider now the homogeneous system $\mathbf{Ax} = \mathbf{0}$. Let $\mathbf{u}$ be a solution, $\mathbf{Au} = \mathbf{0}$. Then premultiplying by the left inverse $\mathbf{B}$,

$$\mathbf{0} = \mathbf{B}(\mathbf{0}) = \mathbf{B}(\mathbf{Au}) = (\mathbf{BA})\mathbf{u} = \mathbf{Iu} = \mathbf{u}$$

tells us that $\mathbf{u} = \mathbf{0}$, that is, the homogeneous system has only the trivial solution. Hence, $\mathbf{A}$ is invertible. $\square$

Equivalent Statements for Invertibility

Theorem

A square matrix $\mathbf{A}$ is invertible if and only if it has a right inverse.

Proof.

Recall that $\mathbf{A}$ is invertible if and only if $\mathbf{A}^T$ is. Let $\mathbf{B}$ be a right inverse of $\mathbf{A}$, $\mathbf{AB} = \mathbf{I}$. Then by taking the transpose, and observing that the identity matrix is symmetric, we have $\mathbf{B}^T\mathbf{A}^T = \mathbf{I}$. This shows that $\mathbf{B}^T$ is a left inverse of $\mathbf{A}^T$. Therefore, by the previous theorem, $\mathbf{A}^T$ is invertible, which thus proves that $\mathbf{A}$ is invertible. $\square$

Equivalent Statement for Invertibility

Theorem

A square matrix $\mathbf{A}$ *is invertible if and only if* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution for all* $\mathbf{b}$.

Proof.

($\Rightarrow$) We have shown that if $\mathbf{A}$ is invertible, then $\mathbf{Ax} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$.

($\Leftarrow$) Now suppose $\mathbf{Ax} = \mathbf{b}$ is consistent for all $\mathbf{b}$. In particular, $\mathbf{Ax} = \mathbf{e}_i$ is consistent, where $\mathbf{e}_i$ is the i-th column of the identity matrix, for $i = 1, ..., n$, where n is the order of $\mathbf{A}$. Let $\mathbf{b}_i$ be a solution to $\mathbf{Ax} = \mathbf{e}_i$. Let $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}$. Then, by block multiplication

$$\mathbf{AB} = \mathbf{A}\begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix} = \begin{pmatrix} \mathbf{Ab}_1 & \mathbf{Ab}_2 & \cdots & \mathbf{Ab}_n \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{pmatrix} = \mathbf{I},$$

which shows that $\mathbf{B}$ is a right inverse of $\mathbf{A}$. Thus, by the previous theorem, $\mathbf{A}$ is invertible. $\square$

Algorithm for Finding Inverse

Thus, we will now formally introduce the algorithm to testing if a matrix is invertible, and finding its inverse it it is invertible.

Let $\mathbf{A}$ be a $n \times n$ matrix.

Step 1: Form the $n \times 2n$ (augmented) matrix $\left(\begin{array}{c|c} \mathbf{A} & \mathbf{I}_n \end{array} \right)$.

Step 2: Reduce the matrix $\left(\begin{array}{c|c} \mathbf{A} & \mathbf{I} \end{array} \right) \longrightarrow \left(\begin{array}{c|c} \mathbf{R} & \mathbf{B} \end{array} \right)$ to its REF or RREF.

Step 3: If RREF $\mathbf{R} \neq \mathbf{I}$ or REF has a zero row, then $\mathbf{A}$ is not invertible. If RREF $\mathbf{R} = \mathbf{I}$ or REF has no zero row, $\mathbf{A}$ is invertible with inverse $\mathbf{A}^{-1} = \mathbf{B}$.

Equivalent Statements of Invertibility

Theorem (Equivalent statements of invertibility)

Let $\mathbf{A}$ *be a square matrix of order* n*. The following statements are equivalent.*

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}^T$ *is invertible.*

(iii) *(left inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{BA} = \mathbf{I}$.

(iv) *(right inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{AB} = \mathbf{I}$.

(v) *The reduced row-echelon form of* $\mathbf{A}$ *is the identity matrix.*

(vi) $\mathbf{A}$ *can be expressed as a product of elementary matrices.*

(vii) *The homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution.*

(viii) *For any* $\mathbf{b}$, *the system* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution.*

2.7 LU Factorization

Example

In the example above, we may write **A** as a product of a lower triangular matrix and a row-echelon of **A**,

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & 5/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & -1 & 0 \\ 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & -3 & 3 & 3 \end{pmatrix}.$$

Observe furthermore that the diagonal entries of the lower triangular matrix are 1. Such matrices are known as *unit lower triangular* matrices. We will write it as $\mathbf{A} = \mathbf{LU}$, where **L** is a unit lower triangular matrix, and **U** is a row-echelon form of **A**.

Example

Consider now the linear system $\mathbf{Ax} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$. Replacing $\mathbf{A}$ with $\mathbf{LU}$, we have $\mathbf{LUx} = \mathbf{b}$. Let $\mathbf{Ux} = \mathbf{y}$, and we first solve for $\mathbf{Ly} = \mathbf{b}$. But since $\mathbf{L}$ is a unit lower triangular matrix, this is easy. From

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 2 \\ 3 & 5/2 & 1 & 1 \end{array}\right)$$

We can observe that $y_1 = 1$, $y_2 = 1$, $y_3 = -9/2$ is the unique solution. Next, we solve for $\mathbf{Ux} = \mathbf{y}$. Now since $\mathbf{U}$ is in row-echelon form, this is easy too,

$$\left(\begin{array}{ccccc|c} 1 & 2 & 1 & -1 & 0 & 1 \\ 0 & -2 & 0 & 0 & 0 & 1 \\ 0 & 0 & -3 & 3 & 3 & -9/2 \end{array}\right) \xrightarrow[-\frac{1}{3}R_3]{-\frac{1}{2}R_2} \left(\begin{array}{ccccc|c} 1 & 2 & 1 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & -1/2 \\ 0 & 0 & 1 & -1 & -1 & 3/2 \end{array}\right)$$

tells us that $x_1 = \frac{1}{2} - t$, $x_2 = -\frac{1}{2}$, $x_3 = \frac{3}{2} + s + t$, $x_4 = s$, $x_5 = t$, $s, t, \in \mathbb{R}$ is the general solution.

LU Factorization

Definition

A square matrix $\mathbf{L}$ is a *unit lower triangular* matrix if $\mathbf{L}$ is a lower triangular matrix with 1 in the diagonal entries.

An *LU factorization* of a $m \times n$ matrix $\mathbf{A}$ is the decomposition

$$\mathbf{A} = \mathbf{LU},$$

where $\mathbf{L}$ is a unit lower triangular matrix, and $\mathbf{U}$ is a row-echelon form of $\mathbf{A}$.

If such LU factorization exits for $\mathbf{A}$, we say that $\mathbf{A}$ is *LU factorizable*.

Unit Lower Triangular Matrices

Lemma

Let **A** *and* **B** *be unit lower triangular matrices of the same size. Then* **AB** *is a unit lower triangular matrix too.*

Proof.

Write $\mathbf{A} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & 1 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ b_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & 1 \end{pmatrix}$. Then

$$\begin{aligned} \mathbf{AB} &= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ b_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21} + b_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + a_{n2} b_{21} + \cdots + b_{n1} & a_{n2} + a_{n3} b_{b2} + \cdots + b_{n2} & \cdots & 1 \end{pmatrix} \end{aligned}$$

$\square$

Unit Lower Triangular Matrix

(i) Observe that the elementary matrix $\mathbf{E}$ corresponding to the operation $R_i + cR_j$ for $i > j$ for some real number c is a lower triangular matrix.

(ii) Also, since the inverse of an elementary matrix is an elementary matrix corresponding to an elementary row operation same type, $\mathbf{E}^{-1}$ is also a unit lower triangular matrix.

(iii) Hence, by the previous lemma, if $\mathbf{E}_1, \mathbf{E}_2, ..., \mathbf{E}_k$ are series of elementary matrices corresponding to row operations of the type $R_i + cR_j$ for $i > j$ for some c, then $\mathbf{E}_1^{-1}\mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1}$ is a unit lower triangular matrix.

Algorithm to LU Factorization

Suppose $\mathbf{A} \xrightarrow{r_1, r_2, \ldots, r_k} \mathbf{U}$, where each row operation r_l is of the form $R_i + cR_j$ for some $i > j$ and real number c, and $\mathbf{U}$ is an row-echelon form of $\mathbf{A}$. Let $\mathbf{E}_i$ be the elementary matrix corresponding for r_i, for $i = 1, 2, ..., k$. Then

$$\mathbf{E}_k \cdots \mathbf{E}_2 \mathbf{E}_1 \mathbf{A} = \mathbf{U} \quad \Rightarrow \quad \mathbf{A} = \mathbf{E}_1^{-1} \mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1} \mathbf{U} = \mathbf{L}\mathbf{U},$$

where $\mathbf{L} = \mathbf{E}_1^{-1} \mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1}$. Then

$$\mathbf{A} = \mathbf{L}\mathbf{U} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ * & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & 1 \end{pmatrix} \begin{pmatrix} * & & & & & \cdots & * \\ 0 & \cdots & 0 & * & & \cdots & * \\ \vdots & & & & & & \vdots \\ 0 & \cdots & & & & \cdots & * \end{pmatrix}$$

is a LU factorization of $\mathbf{A}$.

In this case, we could obtain $\mathbf{L}$ quickly without computing $\mathbf{E}_1^{-1} \mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1}$. For each row operation $r_l = R_i + c_l R_j$ for some $i > j$ and real number c_l, we will put $-c_l$ in the (i, j)-entry of $\mathbf{L}$.

Remarks

1. The question on the characterization of whether a matrix is LU factorizable is beyond the scope of this course.

2. Any matrix admits a LU factorization with pivoting (LUP factorization), that is, any matrix $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{PLU}$, where $\mathbf{L}$ is a unit lower triangular matrix, $\mathbf{U}$ is a row-echelon form of $\mathbf{A}$, and $\mathbf{P}$ is a permutation matrix (see below for definition).

Definition

A $n \times n$ matrix $\mathbf{P}$ is a *permutation matrix* if every rows and columns has a 1 in only one entry, and 0 everywhere else.

Equivalently, $\mathbf{P}$ is a permutation matrix if and only if $\mathbf{P}$ is the product of elementary matrices corresponding to row swaps.

2.8 Determinant by Cofactor Expansion

Order 1 and 2 Square Matrices

We will define the *determinant* of $\mathbf{A}$ of order n, denoted as $\det(\mathbf{A})$, or $|\mathbf{A}|$, by induction.

1. For $n = 1$, $\mathbf{A} = (a)$, $\det(\mathbf{A}) = a$.
2. For $n = 2$, $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\det(\mathbf{A}) = ad - bc$.

Example

(i) $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}$, $\det(\mathbf{A}) = (1)(1) - (2)(3) = -5$.

(ii) $\mathbf{A} = \begin{pmatrix} 3 & -1 \\ 5 & 1 \end{pmatrix}$, $\det(\mathbf{A}) = (3)(1) - (-1)(5) = 8$.

Inductive Step: Matrix Minor

Suppose we have defined the determinant of all square matrices of order $\leq n-1$. Let $\mathbf{A}$ be a square matrix of order n.

Define $\mathbf{M}_{ij}$, called the *(i,j) matrix minor* of $\mathbf{A}$, to be the matrix obtained from $\mathbf{A}$ be deleting the i-th row and j-th column.

Example

$$\mathbf{A} = \begin{pmatrix} 5 & 1 & 2 & -1 \\ -1 & -3 & 1 & 3 \\ 3 & 8 & 2 & 1 \\ 2 & 0 & 1 & 11 \end{pmatrix}$$

$$\mathbf{M}_{11} = \begin{pmatrix} -3 & 1 & 3 \\ 8 & 2 & 1 \\ 0 & 1 & 11 \end{pmatrix}, \quad \mathbf{M}_{12} = \begin{pmatrix} -1 & 1 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 11 \end{pmatrix}, \quad \mathbf{M}_{23} = \begin{pmatrix} 5 & 1 & -1 \\ 3 & 8 & 1 \\ 2 & 0 & 11 \end{pmatrix}, \quad \mathbf{M}_{43} = \begin{pmatrix} 5 & 1 & -1 \\ -1 & -3 & 3 \\ 3 & 8 & 1 \end{pmatrix}$$

Inductive Step: Cofactor

The (i,j)-cofactor of $\mathbf{A}$, denoted as A_{ij}, is the (real) number given by

$$A_{ij} = (-1)^{i+j}\det(\mathbf{M}_{ij}).$$

Take note of the sign of the (i,j)-entry, $(-1)^{i+j}$. Here's a visualization of the sign of the entries of the matrix

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

Determinant by Cofactor Expansion

The *determinant* of $\mathbf{A}$ is defined to be

$$\det(\mathbf{A}) = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = \sum_{k=1}^{n} a_{ik}A_{ik} \tag{1}$$

$$= a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} = \sum_{k=1}^{n} a_{kj}A_{kj} \tag{2}$$

This is called the *cofactor expansion* along $\left\{ \begin{array}{ccc} \text{row} & i & (3) \\ \text{column} & j & (4) \end{array} \right.$.

Remark

The above is both a theorem and a definition. The theorem states that evaluating the cofactor expansion along any row or column produces the same result. Hence, we may define the determinant to be the cofactor expansion along any rows or columns. Readers may refer to the appendix for details.

Property of Determinant

Theorem (Determinant is invariant under tanspose)

Let $\mathbf{A}$ *be a square matrix. Then the determinant of* $\mathbf{A}$ *is equal to the determinant of* $\mathbf{A}^T$*,*

$$\det(\mathbf{A}) = \det(\mathbf{A}^T).$$

Students can try to prove the theorem, the details can be found in the appendix. The idea is that taking the cofactor expansion of $\mathbf{A}$ along column 1 is equal to the cofactor expansion of $\mathbf{A}^T$ along row 1. But the first produces $\det(\mathbf{A})$, while the latter $\det(\mathbf{A}^T)$.

Order 3 Matrices

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} = aei - afh - bdi + bfg + cdh - ceg.$$

$$\left|\begin{array}{ccc|cc} \mathrm{a} & \mathrm{b} & \mathrm{c} & \mathrm{a} & \mathrm{b} \\ \mathrm{d} & \mathrm{e} & \mathrm{f} & \mathrm{d} & \mathrm{e} \\ \mathrm{g} & \mathrm{h} & \mathrm{i} & \mathrm{g} & \mathrm{h} \end{array}\right|$$

Example

$$\begin{vmatrix} 0 & 1 & 1 \\ 2 & 1 & 0 \\ 1 & 2 & 3 \end{vmatrix} = (0)(1)(3) + (1)(0)(1) + (1)(2)(2) - (1)(1)(1) - (1)(2)(3) - (0)(0)(2) = -3$$

Determinant of Triangular Matrices

Theorem (Determinant of a triangular matrix is the product of diagonal entries)

If $\mathbf{A} = (a_{ij})_n$ is a triangular matrix, then

$$\det(\mathbf{A}) = a_{11}a_{22}\cdots a_{nn} = \prod_{k=1}^{n} a_{ii}.$$

Sketch of proof.

Upper triangular matrix , continuously cofactor expand along first column,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} & a_{24} \\ 0 & a_{33} & a_{34} \\ 0 & 0 & a_{44} \end{vmatrix} = a_{11}a_{22}\begin{vmatrix} a_{33} & a_{34} \\ 0 & a_{44} \end{vmatrix} = a_{11}a_{22}a_{33}a_{44}.$$

Lower triangular matrix, continuously cofactor expand along the first row,

$$\begin{vmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & 0 & 0 \\ a_{32} & a_{33} & 0 \\ a_{42} & a_{43} & a_{44} \end{vmatrix} a_{11}a_{22}\begin{vmatrix} a_{33} & 0 \\ a_{43} & a_{44} \end{vmatrix} = a_{11}a_{22}a_{33}a_{44}.$$

$\square$

2.9 Determinant by Reduction

Determinant and Elementary Row Operations

$R_i + aR_j$

$$\det(\mathbf{B}) = \begin{vmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 2$$

$$\mathbf{A} = \begin{pmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 3 & 15 & 4 & 8 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_3 - 3R_1} \begin{pmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \mathbf{B}$$

$$\det(\mathbf{A}) = \begin{vmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 3 & 15 & 4 & 8 \\ 0 & 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 5 & 1 \\ 0 & 2 & 6 \\ 3 & 15 & 4 \end{vmatrix} = \begin{vmatrix} 2 & 6 \\ 15 & 4 \end{vmatrix} + 3 \begin{vmatrix} 5 & 1 \\ 2 & 6 \end{vmatrix} = 2.$$

$$\det(\mathbf{B}) = \det(\mathbf{A}).$$

Determinant and Elementary Row Operations

cR_i

$$\det(\mathbf{B}) = \begin{vmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 2$$

$$\mathbf{A} = \begin{pmatrix} 1 & 5 & 1 & 2 \\ 0 & 4 & 12 & 6 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{\frac{1}{2}R_2} \begin{pmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \mathbf{B}$$

$$\det(\mathbf{A}) = \begin{vmatrix} 1 & 5 & 1 & 2 \\ 0 & 4 & 12 & 6 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 4.$$

$$\det(\mathbf{B}) = \frac{1}{2}\det(\mathbf{A}).$$

Determinant and Elementary Row Operations

$R_i \leftrightarrow R_j$

$$\det(\mathbf{B}) = \begin{vmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 2$$

$$\mathbf{A} = \begin{pmatrix} 1 & 5 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 2 & 6 & 3 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_4} \begin{pmatrix} 1 & 5 & 1 & 2 \\ 0 & 2 & 6 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \mathbf{B}$$

$$\det(\mathbf{A}) = \begin{vmatrix} 1 & 5 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 2 & 6 & 3 \end{vmatrix} = \begin{vmatrix} 0 & 0 & 1 \\ 0 & 1 & 2 \\ 2 & 6 & 3 \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 2 & 6 \end{vmatrix} = -2.$$

$$\det(\mathbf{B}) = -\det(\mathbf{A}).$$

Determinant and Elementary Row Operations

Theorem

Let $\mathbf{A}$ *be a* $n \times n$ *square matrix. Suppose* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *via a single elementary row operation. Then the determinant of* $\mathbf{B}$ *is obtained as such.*

$\mathbf{A} \xrightarrow{R_i + aR_j} \mathbf{B}$	$\det(\mathbf{B}) = \det(\mathbf{A})$
$\mathbf{A} \xrightarrow{cR_i} \mathbf{B}$	$\det(\mathbf{B}) = c\det(\mathbf{A})$
$\mathbf{A} \xrightarrow{R_i \leftrightarrow R_j} \mathbf{B}$	$\det(\mathbf{B}) = -\det(\mathbf{A})$

The proof can be found in the appendix.

Corollary

The determinant of an elementary matrix $\mathbf{E}$ *is given as such.*

(i) *If* $\mathbf{E}$ *corresponds to* $R_i + aR_j$*, then* $\det(\mathbf{E}) = 1$.

(ii) *If* $\mathbf{E}$ *corresponds to* cR_j*, then* $\det(\mathbf{E}) = c$.

(iii) *If* $\mathbf{E}$ *corresponds to* $R_i \leftrightarrow R_j$*, then* $\det(\mathbf{E}) = -1$.

Proof.

This follows immediately from the previous theorem by letting $\mathbf{A} = \mathbf{I}$. □

Determinant of Row Equivalent Matrices

Theorem

Let $\mathbf{A}$ *and* $\mathbf{R}$ *be square matrices such that*

$$\mathbf{R} = \mathbf{E}_k \cdots \mathbf{E}_2 \mathbf{E}_1 \mathbf{A}$$

for some elementary matrices $\mathbf{E}_1, \mathbf{E}_2, ..., \mathbf{E}_k$. *Then*

$$\det(\mathbf{R}) = \det(\mathbf{E}_k) \cdots \det(\mathbf{E}_2) \det(\mathbf{E}_1) \det(\mathbf{A}).$$

Proof.

This follows immediately from the previous theorem and corollary. $\square$

Remark

Note that the determinant of an elementary matrix is nonzero. This means that the determinant of $\mathbf{R}$ could be computed from the determinant of $\mathbf{A}$ and vice versa.

Determinant of Row Equivalent Matrices

Corollary

Let $\mathbf{A}$ *be a* $n \times n$ *square matrix. Suppose* $\mathbf{A} \xrightarrow{r_1} \xrightarrow{r_2} \cdots \xrightarrow{r_k} \mathbf{R} = \begin{pmatrix} d_1 & * & \cdots & * \\ 0 & d_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$, *where* $\mathbf{R}$ *is the reduced row-echelon form of* $\mathbf{A}$*. Let* $\mathbf{E}_i$ *be the elementary matrix corresponding to the elementary row operation* r_i*, for* $i = 1, ..., k$*. Then*

$$\det(\mathbf{A}) = \frac{d_1 d_2 \cdots d_n}{\det(\mathbf{E}_k) \cdots \det(\mathbf{E}_2) \det(\mathbf{E}_1)}.$$

Proof.

Using the previous theorem and the fact that the determinant of a triangular matrix is the product its diagonal entries,

$$d_1 d_2 \cdots d_n = \det(\mathbf{E}_k) \cdots \det(\mathbf{E}_2) \det(\mathbf{E}_1) \det(\mathbf{A}).$$

Hence,

$$\det(\mathbf{A}) = \frac{d_1 d_2 \cdots d_n}{\det(\mathbf{E}_k) \cdots \det(\mathbf{E}_2) \det(\mathbf{E}_1)}.$$

$\square$

2.10 Properties of Determinant

Determinant and Invertiblity

Theorem

A square matrix $\mathbf{A}$ *is invertible if and only if* $\det(\mathbf{A}) \neq 0$.

Proof.

Suppose $\mathbf{A}$ is invertible. Then we can write $\mathbf{A} = \mathbf{E}_1\mathbf{E}_2 \cdots \mathbf{E}_k$ as a product of elementary matrices. Hence,

$$\det(\mathbf{A}) = \det(\mathbf{E}_1)\det(\mathbf{E}_2)\cdots\det(\mathbf{E}_k),$$

which is nonzero since determinant of elementary matrices are nonzero.
Conversely, suppose $\mathbf{A}$ is singular. Then the last row of the reduce row-echelon form $\mathbf{R}$ of $\mathbf{A}$ is a zero row. Thus $\det(\mathbf{R}) = 0$ by cofactor expanding along the last row. Write $\mathbf{R} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$. Then

$$\det(\mathbf{A}) = \frac{\det(\mathbf{R})}{\det(\mathbf{E}_k)\cdots\det(\mathbf{E}_2)\det(\mathbf{E}_1)} = 0.$$

$\square$

We will add this to the equivalent statement of invertibility.

Equivalent Statements for Invertibility

Let $\mathbf{A}$ be a square matrix of order n. The following statements are equivalent.

(i) $\mathbf{A}$ is invertible.

(ii) $\mathbf{A}^T$ is invertible.

(iii) (left inverse) There is a matrix $\mathbf{B}$ such that $\mathbf{BA} = \mathbf{I}$.

(iv) (right inverse) There is a matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{I}$.

(v) The reduced row-echelon form of $\mathbf{A}$ is the identity matrix.

(vi) $\mathbf{A}$ can be expressed as a product of elementary matrices.

(vii) The homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.

(viii) For any $\mathbf{b}$, the system $\mathbf{Ax} = \mathbf{b}$ has a unique solution.

(ix) The determinant of $\mathbf{A}$ is nonzero, $\det(\mathbf{A}) \neq 0$.

Determinant of Product of Matrices

Theorem

Let $\mathbf{A}$ and $\mathbf{B}$ be square matrices of the same size. Then

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}).$$

Students may try to prove the theorem by considering the cases where $\mathbf{A}$ is invertible or not. The proof can be found in the appendix.

By induction, we get

$$\det(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_k) = \det(\mathbf{A}_1)\det(\mathbf{A}_2)\cdots\det(\mathbf{A}_k).$$

Determinant of inverse

Theorem

If $\mathbf{A}$ *is invertible, then*

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

Proof.

Using the fact that $\det(\mathbf{I}) = 1$, where $\mathbf{I}$ is the identity matrix, and that the determinant of product is the product of determinant, we have

$$1 = \det(\mathbf{I}) = \det(\mathbf{A}^{-1}\mathbf{A}) = \det(\mathbf{A}^{-1})\det(\mathbf{A}).$$

Since $\mathbf{A}$ is invertible, its determinant is nonzero. Hence, we can divide both sides of the equation above by $\det(\mathbf{A})$ to obtain the conclusion

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}.$$

$\square$

Determinant of Scalar Multiplication

Theorem

For any square matrix $\mathbf{A}$ *of order* n *and scalar* $c \in \mathbb{R}$,

$$\det(c\mathbf{A}) = c^n \det(\mathbf{A}).$$

Proof.

Observe that scalar multiplication is equivalent to matrix multiplication by scalar matrix,

$$c\mathbf{A} = (c\mathbf{I})\mathbf{A} = \begin{pmatrix} c & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c \end{pmatrix} \mathbf{A}.$$ Hence,

$$\det(c\mathbf{A}) = \det((c\mathbf{I})\mathbf{A}) = \begin{vmatrix} c & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c \end{vmatrix} \det(\mathbf{A}) = c^n \det(\mathbf{A}).$$

$\square$

Adjoint

Definition

Let $\mathbf{A}$ be a $n \times n$ square matrix. The *adjoint* of $\mathbf{A}$, denoted as $\mathbf{adj}(\mathbf{A})$, is the $n \times n$ square matrix whose (i, j) entry is the (j, i)-cofactor of $\mathbf{A}$,

$$\mathbf{adj}(\mathbf{A}) = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}^T = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}.$$

Adjoint Formula

Theorem

Let $\mathbf{A}$ *be a square matrix and* $\mathbf{adj}(\mathbf{A})$ *it adjoint. Then*

$$\mathbf{A}(\mathbf{adj}(\mathbf{A})) = \det(\mathbf{A})\mathbf{I},$$

where $\mathbf{I}$ *is the identity matrix.*

The proof can be found in the appendix.

Corollary (Adjoint Formula for Inverse)

Let $\mathbf{A}$ *be an invertible matrix. Then the inverse of* $\mathbf{A}$ *is given by*

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\mathbf{adj}(\mathbf{A}).$$

The corollary follows immediately from the previous theorem, and the fact that $\det(\mathbf{A}) \neq 0$. From the adjoint formula for inverse, we have the inverse formula for 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

Appendix

Properties of Matrix Addition and Scalar Multiplication

Theorem

For matrices $\mathbf{A} = (a_{ij})_{m \times n}$, $\mathbf{B} = (b_{ij})_{m \times n}$, $\mathbf{C} = (c_{ij})_{m \times n}$, *and real numbers* $a, b \in \mathbb{R}$,

(i) *(Commutative)* $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$,

(ii) *(Associative)* $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$,

(iii) *(Additive identity)* $\mathbf{0}_{m \times n} + \mathbf{A} = \mathbf{A}$,

(iv) *(Additive inverse)* $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}_{m \times n}$,

(v) *(Distributive law)* $a(\mathbf{A} + \mathbf{B}) = a\mathbf{A} + a\mathbf{B}$,

(vi) *(Scalar addition)* $(a + b)\mathbf{A} = a\mathbf{A} + b\mathbf{A}$,

(vii) *(Associative)* $(ab)\mathbf{A} = a(b\mathbf{A})$,

(viii) *If* $a\mathbf{A} = \mathbf{0}_{m \times n}$, *then either* $a = 0$ *or* $\mathbf{A} = \mathbf{0}$.

Proof of the Properties of Scalar Multiplication and Matrix Addition

Proof.

(i) $\mathbf{A} + \mathbf{B} = (a_{ij} + b_{ij})_{m\times n} = (b_{ij} + a_{ij})_{m\times n} = \mathbf{B} + \mathbf{A}$, since addition of real numbers is commutative.

(ii) $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (a_{ij} + (b_{ij} + c_{ij}))_{m\times n} = ((a_{ij} + b_{ij}) + c_{ij})_{m\times n} = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$, since addition of real numbers is associative.

(iii) $\mathbf{A} + \mathbf{0} = (a_{ij} + 0)_{m\times n} = (a_{ij})_{m\times n} = \mathbf{A}$.

(iv) $\mathbf{A} + (-\mathbf{A}) = (a_{ij} + (-a_{ij}))_{m\times m} = (a_{ij} - a_{ij})_{m\times n} = (0)_{m\times n} = \mathbf{0}_{m\times n}$.

(v) $a(\mathbf{A} + \mathbf{B}) = (a(a_{ij} + b_{ij}))_{m\times n} = (aa_{ij} + ab_{ij})_{m\times n} = (aa_{ij})_{m\times n} + (ab_{ij})_{m\times n} = a(a_{ij})_{m\times n} + a(b_{ij})_{m\times n} = a\mathbf{A} + a\mathbf{B}$, where the 2^{nd} equality follows from the distributive property of real number addition and multiplication, and the 4^{th} equality follows from the definition of scalar multiplication.

(vi) $(a + b)\mathbf{A} = ((a + b)a_{ij})_{m\times n} = (aa_{ij} + ba_{ij})_{m\times n} = (aa_{ij})_{m\times n} + (ba_{ij})_{m\times n} = a(a_{ij})_{m\times n} + b(a_{ij})_{m\times n} = a\mathbf{A} + b\mathbf{A}$.

(vii) $(ab)\mathbf{A} = ((ab)a_{ij})_{m\times n} = (aba_{ij})_{m\times n} = (a(ba_{ij}))_{m\times n} = a(ba_{ij})_{m\times n} = a(b\mathbf{A})$.

(viii) Suppose $a\mathbf{A} = (aa_{ij})_{m\times n} = (0)_{m\times n}$. This means that $aa_{ij} = 0$ for all i, j. So, if $a \neq 0$, then necessarily $a_{ij} = 0$ for all i, j, which means that $\mathbf{A} = \mathbf{0}$.

$\square$

Properties of Matrix Multiplication

Theorem

(i) *(Associative) For matrices* $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{p\times q}$, *and* $\mathbf{C} = (c_{ij})_{q\times n}$ $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

(ii) *(Left distributive law) For matrices* $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{p\times n}$, *and* $\mathbf{C} = (c_{ij})_{p\times n}$, $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$.

(iii) *(Right distributive law) For matrices* $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{m\times p}$, *and* $\mathbf{C} = (c_{ij})_{p\times n}$, $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$.

(iv) *(Commute with scalar multiplication) For any real number* $c \in \mathbb{R}$, *and matrices* $\mathbf{A} = (a_{ij})_{m\times p}$, $\mathbf{B} = (b_{ij})_{p\times n}$, $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B})$.

(v) *(Multiplicative identity) For any* $m \times n$ *matrix* $\mathbf{A}$, $\mathbf{I}_m\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{I}_n$.

(vi) *(Zero divisor) There exists* $\mathbf{A} \neq \mathbf{0}_{m\times p}$ *and* $\mathbf{B} \neq \mathbf{0}_{p\times n}$ *such that* $\mathbf{AB} = \mathbf{0}_{m\times n}$.

(vii) *(Zero matrix) For any* $m \times n$ *matrix* $\mathbf{A}$, $\mathbf{A}\mathbf{0}_{n\times p} = \mathbf{0}_{m\times p}$ *and* $\mathbf{0}_{p\times m}\mathbf{A} = \mathbf{0}_{p\times n}$.

Proof of the Properties of Matrix Multiplication

Proof.

We will check that the corresponding entries on each side agrees. The check for the size of matrices agree is trivial and is left to the reader.

(i) The (i,j)-entry of $(\mathbf{AB})\mathbf{C}$ is

$$\sum_{l=1}^{q}(\sum_{k=1}^{p} a_{ik}b_{kl})c_{lj} = \sum_{l=1}^{q}\sum_{k=1}^{p} a_{ik}b_{kl}c_{lj}.$$

The (i,j)-entry of $\mathbf{A}(\mathbf{BC})$ is

$$\sum_{k=1}^{p} a_{ik}(\sum_{l=1}^{q} b_{kl}c_{lj}) = \sum_{k=1}^{p}\sum_{l=1}^{q} a_{ik}b_{kl}c_{lj}.$$

Since both sums has finitely many terms, the sums commute and thus the (i,j)-entry of $(\mathbf{AB})\mathbf{C}$ is equal to the (i,j)-entry of $\mathbf{A}(\mathbf{BC})$.

(ii) The (i,j)-entry of $\mathbf{A}(\mathbf{B}+\mathbf{C})$ is $\sum_{k=1}^{p} a_{ik}(b_{kj}+c_{kj}) = \sum_{k=1}^{p}(a_{ik}b_{kj}+a_{ik}c_{kj}) = \sum_{k=1}^{p} a_{ik}b_{kj} + \sum_{k=1}^{p} a_{ik}c_{kj}$, which is the (i,j)-entry of $\mathbf{AB}+\mathbf{AC}$.

(iii) The proof is analogous to left distributive law.

Proof of the Properties of Matrix Multiplication

Continue.

(iv) The i, j entry of $c\mathbf{AB}$ is $c(\sum_{k=1}^{p} a_{ik} b_{kj}) = \sum_{k=1}^{p} (ca_{ik}) b_{kj} = \sum_{k=1}^{p} a_{ik}(cb_{kj})$.

(v) Note that $\mathbf{I} = (\delta_{ij})$, where $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$. So the (i, j)-entry of $\mathbf{I}_m\mathbf{A}$ is

$$\delta_{i1} a_{1j} + \cdots + \delta_{ii} a_{ij} + \cdots + \delta_{im} a_{mj} = 0a_{1j} + \cdots + 1a_{ij} + \cdots + 0a_{mj} = a_{ij}.$$

The proof for $\mathbf{A} = \mathbf{AI}_n$ is analogous.

(vi) Consider for example $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

(vii) Left to reader, if you have read till this far, surely this proof is trivial to you.

$\square$

Row Equivalent Augmented Matrices have the Same Solutions

Theorem

Let $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{B}\mathbf{x} = \mathbf{c}$ be two linear system such that their augmented matrices $(\mathbf{A} \mid \mathbf{b})$ and $(\mathbf{B} \mid \mathbf{c})$ are row equivalent. Then $\mathbf{u}$ is a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ if and only if it is a solution to $\mathbf{B}\mathbf{x} = \mathbf{c}$. That is, row equivalent augmented matrices have the same set of solutions.

Proof.

By the hypothesis, there exists an invertible matrix $\mathbf{P}$ such that

$$(\mathbf{PA} \mid \mathbf{Pb}) = \mathbf{P}(\mathbf{A} \mid \mathbf{b}).$$

This means that $\mathbf{PA} = \mathbf{B}$ and $\mathbf{Pb} = \mathbf{c}$. Now suppuse $\mathbf{u}$ is a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$, that is, $\mathbf{A}\mathbf{u} = \mathbf{b}$. Then premultiplying the equation by $\mathbf{P}$, we have

$$\mathbf{B}\mathbf{u} = \mathbf{PA}\mathbf{u} = \mathbf{P}\mathbf{b} = \mathbf{c},$$

which shows that $\mathbf{u}$ is a solution to $\mathbf{B}\mathbf{x} = \mathbf{c}$ too. Conversely, suppose $\mathbf{u}$ is a solution to $\mathbf{B}\mathbf{x} = \mathbf{c}$, $\mathbf{B}\mathbf{u} = \mathbf{c}$. Premultiplying both sides of the equation by $\mathbf{P}^{-1}$, we have

$$\mathbf{A}\mathbf{u} = \mathbf{P}^{-1}\mathbf{B}\mathbf{u} = \mathbf{P}^{-1}\mathbf{c} = \mathbf{b},$$

which shows that $\mathbf{u}$ is a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ too. $\square$

Permutation Groups

Definition

The group of permutations of n objects, denoted as S_n, is called the *n-permutation group*. An element in the permutation group is called a *permutation*. An *transposition* is a permutation that only exchanges 2 objects.

Example

$$\begin{aligned} S_1 &= \{(1)\} \\ S_2 &= \{(1,2),(2,1)\} \\ S_3 &= \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\} \end{aligned}$$

We may also interpret a member of the permutation group as a function. For example $\sigma = (2,1,3)$ as the function $\sigma(1) = 2$, $\sigma(2) = 1$, $\sigma(3) = 3$. $(2,1)$, $(1,3,2)$, and $(3,2,1)$ are examples of inversions.

Permutation Groups

Theorem

Every permutation is a composition of transposition.

Example

1. The permutation $(2, 3, 1)$ is obtained by exchanging 1 and 2, then 3 and 2.
2. The permutation $(3, 1, 4, 2)$ is obtained by exchanging 1 and 2, 1 and 3, and 3 and 4.

Definition

Let $\sigma \in S_n$ be a permutation. The *sign* of σ, denoted by $sgn(\sigma)$, is define to be

$$sgn(\sigma) = (-1)^{\text{number of transposition in the decomposition of } \sigma}.$$

Leibniz Formula for Determinant

Definition

Let $\mathbf{A} = (a_{ij})_{n\times n}$ be a $n \times n$ square matrix. Let S_n denote the n-permutation group. The *determinant* of $\mathbf{A}$ is defined to be

$$\det(\mathbf{A}) = \sum_{\sigma\in S_n} sgn(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}.$$

Example

1. For $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, $\det(\mathbf{A}) = sgn((1,2))(a_{11}a_{22}) + sgn((2,1))(a_{12}a_{21}) = a_{11}a_{22} - a_{12}a_{21}$.
2. For $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$,

$$\begin{aligned} \det(\mathbf{A}) &= sgn((1,2,3))a_{11}a_{22}a_{33} + sgn((1,3,2))a_{11}a_{23}a_{32} + sgn((2,1,3))a_{12}a_{21}a_{33} \\ &\quad + sgn((2,3,1))a_{12}a_{23}a_{31} + sgn((3,1,2))a_{13}a_{21}a_{32} + sgn((3,2,1))a_{13}a_{22}a_{31} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \end{aligned}$$

Equivalence in the Definition of Determinant

Theorem

Let $\det(\mathbf{A})$ *denote the determinant of a* $n \times n$ *square matrix* $\mathbf{A}$ *computed using the Leibniz formula. Then*

$$\begin{aligned}\det(\mathbf{A}) &= a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = \sum_{k=1}^{n} a_{ik}A_{ik} \\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} = \sum_{k=1}^{n} a_{kj}A_{kj}\end{aligned}$$

That is, the determinant is equal to the cofactor expansion along any row or column.

This proves the claim that the evaluation of the cofactor expansion along any row or column produces the same result, and that the result is the determinant of a square matrix.

Proof that Determinant is Invariant under Transpose

Theorem (Determinant is invariant under transpose)

Let $\mathbf{A}$ *be a square matrix. Then the determinant of* $\mathbf{A}$ *is equal to the determinant of* $\mathbf{A}^T$*,*

$$\det(\mathbf{A}) = \det(\mathbf{A}^T).$$

Proof.

We will prove by induction on the order of $\mathbf{A}$. The theorem holds trivially for a 1×1 matrix since the transpose of a real number is itself.

Now suppose $\det(\mathbf{A}) = \det(\mathbf{A}^T)$ for all matrices of order k. Let $\mathbf{A}$ be a $k+1 \times k+1$ matrix. Write $\mathbf{A} = (a_{ij})_{k+1 \times k+1}$, for all $i, j = 1, ..., n$. Cofactor expand along the first row of $\mathbf{A}$, we have $\det(\mathbf{A}) = \prod_{i=1}^{k+1} a_{1i}(-1)^{i+1} \det \mathbf{M}_{1i})$. Cofactor expand along the first column of $\mathbf{A}^T$, and noting that the (i, j)-matrix minor of $\mathbf{A}^T$ is the transpose of the (j, i)-matrix minor of $\mathbf{A}$ we have

$$\det(\mathbf{A}^T) = \prod_{i=1}^{k+1} b_{i1}(-1)^{i+1} \det \mathbf{M}_{1i}^T) = \prod_{i=1}^{k+1} a_{1i}(-1)^{i+1} \det \mathbf{M}_{1i}^T) = \prod_{i=1}^{k+1} a_{1i}(-1)^{i+1} \det \mathbf{M}_{1i}) = \det(\mathbf{A}),$$

where for the second equality we use the fact that $b_{ij} = a_{ji}$, and the induction hypothesis in the third equality since the matrix minor $\mathbf{M}_{ij}$ is a $k \times k$ matrix. $\square$

Determinant of Elementary Matrices

Theorem

Let $\mathbf{A}$ *be a* $n \times n$ *square matrix, and* $\mathbf{B}$ *the matrix obtained from* $\mathbf{A}$ *via exchanging the* k*-th row and the* l*-th row,* $\mathbf{A} \xrightarrow{R_k \leftrightarrow R_l} \mathbf{B}$*. Then*

$$\det(\mathbf{B}) = -\det(\mathbf{A}).$$

Proof.

We will proof by induction on the order n of $\mathbf{A}$ for $n \geq 2$. The case for $n = 2$ is clear. Suppose now the statement is true for all matrices of size $n \times n$. Write $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$. Fix $m \neq k, l$. Observe that for each $j = 1, ..., n_1$, the (m, j) matrix minor of $\mathbf{B}$ can be obtained from the (m, j) matrix minor of $\mathbf{A}$ by exchanging the k-th row and the l-th row. Hence, if we let $\mathbf{M}_{mj}$ be the (m, j) matrix minor of $\mathbf{A}$, then by the induction hypothesis, the determinant of (m, j) matrix minor of $\mathbf{B}$ is $-\det(\mathbf{M}_{mj})$. Observe also that $b_{mj} = a_{mj}$ for all $j = 1, ..., n+1$. Hence, cofactor expand along the m-th row, we have

$$\det(\mathbf{B}) = \sum_{i=1}^{n+1} b_{mj}(-1)^{m+j}(-\det(\mathbf{M}_{mj})) = -\sum_{i=1}^{n+1} a_{mj}(-1)^{m+j}\det(\mathbf{M}_{mj}) = -\det(\mathbf{A}).$$

$\square$

Determinant of Elementary Matrices

Theorem

Let $\mathbf{A}$ *be a square matrix, and* $\mathbf{B}$ *the matrix obtained from* $\mathbf{A}$ *via multiplying row* m *by* $c \neq 0$, $\mathbf{A} \xrightarrow{cR_m} \mathbf{B}$. *Then*

$$\det(\mathbf{B}) = c\det(\mathbf{A}).$$

Proof.

Write $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$. Then $b_{ij} = a_{ij}$ for all $i \neq m$ and all $j = 1, ..., n$, and $b_{mj} = ca_{mj}$ for all $j = 1, ..., n$. Also observe that the (m, j) matrix minor of $\mathbf{B}$ is equal to the (m, j) matrix minor of $\mathbf{A}$ for any $j = 1, ..., n$. We will denote the common (m, j) matrix minor as M_{mj}. Now cofactor expand along row m of $\mathbf{B}$,

$$\det(\mathbf{B}) = \sum_{j=1}^{n} b_{mj}\det(M_{mj}) = \sum_{j=1}^{n} ca_{mj}\det(M_{mj}) = c\sum_{j=1}^{n} a_{mj}\det(M_{mj}) = c\det(\mathbf{A}).$$

$\square$

Determinant of Elementary Matrices

Lemma

The determinant of a square matrix with two identical rows is zero.

Proof.

We will proof by induction on the order n of $\mathbf{A}$, for $n \geq 2$. The case for $n = 2$ is clear. Suppose now the statement is true for all matrices of size $n \times n$. Let $\mathbf{A}$ be a $n+1 \times n+1$ matrix such that the k-th row is equal to the l-th row. Now compute the determinant of $\mathbf{A}$ by cofactor expansion along the m-th row, where $m \neq k, l$,

$$\det(\mathbf{A}) = \sum_{j=1}^{n+1} a_{mj}(-1)^{m+j} \det(\mathbf{M}_{mj}).$$

Since $\mathbf{M}_{mj}$ is obtained from $\mathbf{A}$ by deleting the $m \neq k, l$-th row and j-th column, $\mathbf{M}_{mj}$ is a $n \times n$ square matrix with 2 identical rows. Thus, by the induction hypothesis, $\det(\mathbf{M}_{mj}) = 0$ for all $j = 1, ..., n+1$, and hence, $\det(\mathbf{A}) = 0$. □

Lemma

The determinant of a square matrix with two identical columns is zero.

Proof.

Follows immediately from the previous theorem, and the fact that determinant is invariant under transpose. □

Determinant of Elementary Matrices

Theorem

Let $\mathbf{A}$ *be a square matrix, and* $\mathbf{B}$ *the matrix obtained from* $\mathbf{A}$ *by adding a times of row l to row m, for some real number a,* $\mathbf{A} \xrightarrow{R_m + aR_l} \mathbf{B}$*. Then*

$$\det(\mathbf{B}) = \det(\mathbf{A}).$$

Proof.

Write $\mathbf{A} = (a_{ij})$. Then the entries of $\mathbf{B}$ are those of $\mathbf{A}$ except for the m-th row, where the (m, j)-entry of $\mathbf{B}$ is $a_{mj} + aa_{lj}$. Now cofactor expanding along the m-th row of $\mathbf{B}$,

$$\det(\mathbf{B}) = \sum_{j=1}^{n} (a_{mj} + aa_{lj}) A_{mj} = \sum_{j=1}^{n} a_{mj} A_{mj} + \sum_{j=1}^{n} aa_{lj} A_{mj} = \det(\mathbf{A}) + a \left(\sum_{j=1}^{n} a_{lj} A_{mj} \right),$$

where in the last equality, we note that $\sum_{j=1}^{n} a_{mj} A_{mj}$ is the determinant of $\mathbf{A}$ computed by cofactor expanding along the m-th row.

Determinant of Elementary Matrices

Comtinue.
Now consider the matrix $\mathbf{C}$ obtained from $\mathbf{A}$ by replacing the m-th row of $\mathbf{A}$ with the l-th row, that is, the m-th and l-th row of $\mathbf{C}$ are the l-th row of $\mathbf{A}$. Since $\mathbf{C}$ has 2 identical rows, $\det(\mathbf{C}) = 0$. Now since all the other rows of $\mathbf{C}$ are identical to $\mathbf{A}$ except the m-th row, the (m, j)-cofactor of $\mathbf{C}$ is A_{mj}, the (m, j)-cofactor of $\mathbf{A}$. Hence, if we cofactor expand along the m-th row of $\mathbf{C}$, remember that the m-th row is the l-th row of $\mathbf{A}$), we have

$$0 = \det(\mathbf{C}) = \sum_{j=1}^{n} a_{lj} A_{mj}.$$

Thus,

$$\det(\mathbf{B}) = \det(\mathbf{A}) + a\left(\sum_{j=1}^{n} a_{lj} A_{mj}\right) = \det(\mathbf{A}) + a0 = \det(\mathbf{A}).$$

$\square$

Determinant of Product of Matrices

Theorem

Let $\mathbf{A}$ *and* $\mathbf{B}$ *be square matrices of the same size. Then*

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}).$$

Proof.

Suppose $\mathbf{A}$ is singular. Then the product is singular too. Hence,

$$\det(\mathbf{AB}) = 0 = 0\det(\mathbf{B}) = \det(\mathbf{A})\det(\mathbf{B}).$$

Suppose $\mathbf{A}$ is invertible. Write $\mathbf{A} = \mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_k$ as a product of elementary matrices. Then $\mathbf{AB} = \mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_k\mathbf{B}$ and thus

$$\det(\mathbf{AB}) = \det(\mathbf{E}_1)\det(\mathbf{E}_2)\cdots\det(\mathbf{E}_k)\det(\mathbf{B}) = \det(\mathbf{A})\det(\mathbf{B}),$$

where in the last equality, we used the fact that $\det(\mathbf{A}) = \det(\mathbf{E}_1)\det(\mathbf{E}_2)\cdots\det(\mathbf{E}_k)$ $\square$

Proof of the Adjoint Formula

Theorem

Let $\mathbf{A}$ *be a square matrix and* $\mathbf{adj}(\mathbf{A})$ *it adjoint. Then*

$$\mathbf{A}(\mathbf{adj}(\mathbf{A})) = \det(\mathbf{A})\mathbf{I},$$

where $\mathbf{I}$ *is the identity matrix.*

Proof.

Let $(\mathbf{A}(\mathbf{adj}(\mathbf{A})))[i,j]$ denote the (i,j) entry of the product $\mathbf{A}(\mathbf{adj}(\mathbf{A}))$. Suffice to show that

$$\mathbf{A}(\mathbf{adj}(\mathbf{A}))[i,j] = \begin{cases} \det(\mathbf{A}) & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Now write $\mathbf{A} = (a_{ij})$, then

$$\mathbf{A}(\mathbf{adj}(\mathbf{A}))[i,j] = \sum_{k=1}^{n} a_{ik} A_{jk}.$$

If $i = j$, then the cofactor expansion of $\mathbf{A}$ along the i-th row.

Proof of the Adjoint Formula

Comtinue.

For $i \neq j$, consider the matrix $\mathbf{C}$ obtained from $\mathbf{A}$ by replacing the j-th row of $\mathbf{A}$ with the i-th row, that is the i-th row and the j-th row of $\mathbf{C}$ is the i-th row of $\mathbf{A}$. Since $\mathbf{C}$ has 2 identical rows, $\det(\mathbf{C}) = 0$. Also, since all the other rows of $\mathbf{C}$ are identical to $\mathbf{A}$ except the j-th row, the (j,k)-cofactor of $\mathbf{C}$ is A_{jk}, the (j,k)-cofactor of $\mathbf{A}$. Hence, cofactor expanding along the j-th row of $\mathbf{C}$, noting the the j-th row is the i-th row of $\mathbf{A}$,

$$0 = \det(\mathbf{C}) = \sum_{k=1}^{n} a_{ik} A_{jk}.$$

This completes the proof of the theorem. $\square$

Cramer's Rule

Definition

Let $\mathbf{A}$ be a $n \times n$ square matrix and $\mathbf{b}$ a $n \times 1$ vector. Construct a new matrix $\mathbf{A}_i(\mathbf{b})$ be replacing the i-th column of $\mathbf{A}$ with $\mathbf{b}$, for $i = 1, \dots, n$.

Theorem (Cramer's Rule)

Let $\mathbf{A}$ be an invertible $n \times n$ matrix. For any $n \times 1$ vector $\mathbf{b}$, the unique solution to the system $\mathbf{Ax} = \mathbf{b}$ is

$$\mathbf{x} = \frac{1}{\det(\mathbf{A})} \begin{pmatrix} \det(\mathbf{A}_1(\mathbf{b})) \\ \det(\mathbf{A}_2(\mathbf{b})) \\ \vdots \\ \det(\mathbf{A}_n(\mathbf{b})) \end{pmatrix}.$$

Cramer's Rule

Proof.

Since $\mathbf{A}$ is invertible, by the adjoint formula for inverse, the unique solution is

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \frac{1}{\det(\mathbf{A})}\mathbf{adj}(\mathbf{A})\mathbf{b}.$$

Now the i-th entry of $\mathbf{adj}(\mathbf{A})\mathbf{b}$ is

$$\sum_{k=1}^{n}(\mathbf{adj}(\mathbf{A})[i,k])b_k = \sum_{k=1}^{n} b_k A_{ki}$$

On the other hand, by cofactor expansion along the i-th column of $\mathbf{A}_i(\mathbf{b})$, we have

$$\det(\mathbf{A}_i(\mathbf{b})) = \sum_{k=1}^{n} b_k A_{ki}$$

which is exactly the i-th entry of $\mathbf{adj}(\mathbf{A})\mathbf{b}$. Hence, the i-th entry of $\mathbf{A}^{-1}\mathbf{b}$ is $\frac{\det(\mathbf{A}_i(\mathbf{b}))}{\det(\mathbf{A})}$ as required. $\square$

Uniqueness of Reduced Row-Echelon Form

Theorem (Uniqueness of RREF)

Suppose $\mathbf{R}$ *and* $\mathbf{S}$ *are two reduced row-echelon forms of a* $m \times n$ *matrix* $\mathbf{A}$*. Then* $\mathbf{R} = \mathbf{S}$*.*

Proof.

First note that there exists an invertible matrix $\mathbf{P}$ such that

$$\mathbf{PR} = \mathbf{S}. \tag{3}$$

This is because $\mathbf{A}$ is row equivalent to $\mathbf{R}$ and $\mathbf{S}$, and so there are invertible matrices $\mathbf{P}_1, \mathbf{P}_2$ such that $\mathbf{A} = \mathbf{P}_1\mathbf{R}$ and $\mathbf{A} = \mathbf{P}_2\mathbf{S}$. Let $\mathbf{P} = \mathbf{P}_2^{-1}\mathbf{P}_1$. We will prove by induction on the numbers of rows n of $\mathbf{R}$ and $\mathbf{S}$.

Suppose $n = 1$. Then $\mathbf{R}$, $\mathbf{S}$ are row matrices and $\mathbf{P}$ is a nonzero real number. Since the leading entries of $\mathbf{R}$ and $\mathbf{S}$ must be 1, by the equation (3), $\mathbf{P} = 1$. So $\mathbf{R} = \mathbf{S}$.

Now suppose $n > 1$. Write $\mathbf{R} = \begin{pmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n \end{pmatrix}$ and $\mathbf{S} = \begin{pmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \cdots & \mathbf{s}_n \end{pmatrix}$, where $\mathbf{r}_j$ and $\mathbf{s}_j$ is the j-th column of $\mathbf{R}$ and $\mathbf{S}$, respectively. By equation (3), we have

$$\mathbf{P}\mathbf{r}_j = \mathbf{s}_j, \tag{4}$$

for $j = 1, ..., n$. Since $\mathbf{P}$ is invertible, $\mathbf{R}$ and $\mathbf{S}$ must have the same zero columns. By deleting the zero columns and forming a new matrix, we may assume that $\mathbf{R}$ and $\mathbf{S}$ has no zero columns.

Uniqueness of Reduced Row-Echelon Form

Continue.

With this assumption, and the fact that $\mathbf{R}$ and $\mathbf{S}$ are in RREF, necessarily the first column of both $\mathbf{R}$ and $\mathbf{S}$ must have 1 in the first entry and 0 everywhere else. By the equation (3), the first column of $\mathbf{P}$ also have 1 in the first entry and zero everywhere else. So we write $\mathbf{R}$, $\mathbf{S}$, and $\mathbf{P}$ in is submatrices,

$$\mathbf{P} = \begin{pmatrix} 1 & \mathbf{p}' \\ 0 & \\ \vdots & \mathbf{P}' \\ 0 & \end{pmatrix}, \ \mathbf{R} = \begin{pmatrix} 1 & \mathbf{r}' \\ 0 & \\ \vdots & \mathbf{R}' \\ 0 & \end{pmatrix}, \text{ and } \mathbf{S} = \begin{pmatrix} 1 & \mathbf{s}' \\ 0 & \\ \vdots & \mathbf{S}' \\ 0 & \end{pmatrix},$$

where $\mathbf{p}', \mathbf{r}', \mathbf{s}'$ are row matrices. By the equation (3) and block multiplication, we have $\mathbf{P}'\mathbf{R}' = \mathbf{S}'$. Note that $\mathbf{P}'$ is invertible. Since $\mathbf{R}$ and $\mathbf{S}$ are in RREF, $\mathbf{R}'$ and $\mathbf{S}'$ are in RREF too. Hence, by the induction hypothesis, $\mathbf{R}' = \mathbf{S}'$. We are left to show that $\mathbf{r}' = \mathbf{s}'$. Since $\mathbf{R}' = \mathbf{S}'$, and both $\mathbf{R}$ and $\mathbf{S}$ are in RREF, $\mathbf{R}$ and $\mathbf{S}$ must have the same pivot columns, say columns $i_1, i_2, ..., i_r$. In these columns, the entries of $\mathbf{r}'$ and $\mathbf{s}'$ must be zero. For the nonzero entries, by equation (4), and the fact that the entries of the columns agree from second row onward, the entries in the first row of each column agrees too, that is $\mathbf{r}' = \mathbf{s}'$ too. Thus the inductive step in complete, and the statement is proven. $\square$

Row Equivalent Matrices has the same Reduced Row-Echelon Form

Theorem

Two matrices are row equivalent if and only if they have the same reduced row-echelon form.

Proof.

Suppose $\mathbf{A}$ and $\mathbf{B}$ has the same RREF $\mathbf{R}$. Then there are invertible matrices $\mathbf{P}$ and $\mathbf{Q}$ such that $\mathbf{PA} = \mathbf{R}$ and $\mathbf{QB} = \mathbf{R}$. Then

$$\mathbf{Q}^{-1}\mathbf{PA} = \mathbf{Q}^{-1}\mathbf{R} = \mathbf{B}.$$

Since $\mathbf{Q}^{-1}\mathbf{P}$ is invertible, it can be written as a product of elementary matrices, and so $\mathbf{A}$ is row equivalent to $\mathbf{B}$. Suppose now $\mathbf{A}$ is row equivalent to $\mathbf{B}$. Let $\mathbf{P}$ be an invertible matrix such that $\mathbf{PA} = \mathbf{B}$. Let $\mathbf{R}$ be the RREF of $\mathbf{A}$ and $\mathbf{S}$ be the RREF of $\mathbf{B}$. Then $\mathbf{R} = \mathbf{UA}$ and $\mathbf{S} = \mathbf{VB}$ for some invertible matrices $\mathbf{U}$ and $\mathbf{V}$. Then

$$\mathbf{VPU}^{-1}\mathbf{R} = \mathbf{VPA} = \mathbf{VB} = \mathbf{S},$$

which shows that $\mathbf{R}$ is row equivalent to $\mathbf{S}$. By the uniqueness of RREF, $\mathbf{R} = \mathbf{S}$. $\square$

MA1522: Linear Algebra for Computing

Chapter 3: Euclidean Vector Spaces

3.1 Euclidean Vector Spaces

Vectors

Recall that a (real) *n-vector* (or *vector*) is a collection of n ordered real numbers,

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, \text{ where } v_i \in \mathbb{R} \text{ for } i = 1, ..., n.$$

Here the entry v_i is also known as the i-th coordinate.

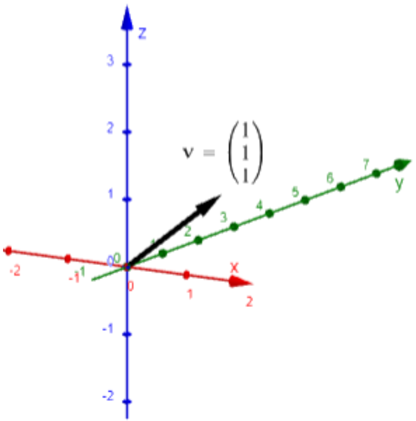
Definition

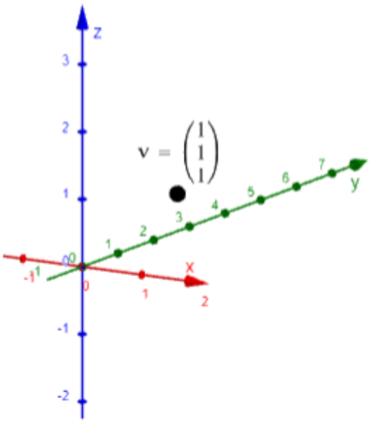
The *Euclidean n-space*, denoted as $\mathbb{R}^n$, is the collection of all n-vectors

$$\mathbb{R}^n = \left\{ \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \middle| \; v_i \in \mathbb{R} \text{ for } i = 1, ..., n. \right\}.$$

Geometric Interpretation of Vectors

Geometrically, a vector **v** can be interpreted as an arrow, with the tail placed at the origin **0**, and the head of the arrow at **v**, or it could represent a position in the Euclidean n-space. For example, the vector $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ in $\mathbb{R}^3$ represents both the point and the arrow.

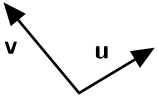


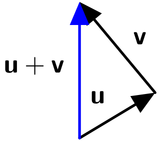


Geometric Interpretation of Vector Algebra

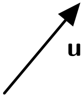
Since vectors are matrices, we are able to apply the matrix algebra on vectors. These operations have geometrical interpretations.

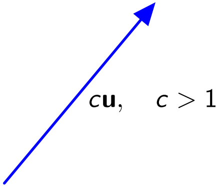
1. Adding $\mathbf{u}$ to $\mathbf{v}$ is visualized as putting the tail of $\mathbf{v}$ at the head of $\mathbf{u}$, and the head of $\mathbf{v}$ is the resultant,





2. Scalar multiple of a vector is scaling the vector,





Vectors Algebra

The following properties follows from properties of matrix algebra. However, try using the geometrical interpretations to prove the following properties.

Theorem

Let $\mathbb{R}^n$ be a Euclidean vector space. Let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be vectors in $\mathbb{R}^n$ and a, b be some real numbers.

(i) *The sum $\mathbf{u} + \mathbf{v}$ is a vector in $\mathbb{R}^n$.*

(ii) *(Commutative) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.*

(iii) *(Associative) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.*

(iv) *(Zero vector) $\mathbf{0} + \mathbf{v} = \mathbf{v}$.*

(v) *The negative $-\mathbf{v}$ is a vector in $\mathbb{R}^n$ such that $\mathbf{v} - \mathbf{v} = \mathbf{0}$.*

(vi) *(Scalar multiple) $a\mathbf{v}$ is a vector in $\mathbb{R}^n$.*

(vii) *(Distribution) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.*

(viii) *(Distribution) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$.*

(ix) *(Associativity of scalar multiplication) $(ab)\mathbf{u} = a(b\mathbf{u})$.*

(x) *If $a\mathbf{u} = \mathbf{0}$, then either $a = 0$ or $\mathbf{u} = \mathbf{0}$.*

Abstract Vector Spaces

Some of these properties of Euclidean vector space tells us that it is a (an abstract) vector space.

Definition

A set V equipped with addition and scalar multiplication is said to be a *vector space* over $\mathbb{R}$ if it satisfies the following axioms.

1. For any vectors $\mathbf{u}, \mathbf{v}$ in V, the sum $\mathbf{u} + \mathbf{v}$ is in V.
2. (Commutative) For any vectors $\mathbf{u}, \mathbf{v}$ in V, $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
3. (Associative) For any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V, $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
4. (Zero vector) There is a vector $\mathbf{0}$ in V such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$ for all vectors $\mathbf{v}$ in V.
5. (Negative) For any vector $\mathbf{u}$ in V, there exists a vector $-\mathbf{u}$ in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
6. For any scalar a in $\mathbb{R}$ and vector $\mathbf{v}$ in V, $a\mathbf{v}$ is a vector in V.
7. (Distribution) For any scalar a in $\mathbb{R}$ and vectors $\mathbf{u}, \mathbf{v}$ in V, $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.
8. (Distribution) For any scalars a, b in $\mathbb{R}$ and vector $\mathbf{u}$ in V, $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$.
9. (Associativity of scalar multiplication) For any scalars a, b in $\mathbb{R}$ and vector $\mathbf{u}$ in V, $a(b\mathbf{u}) = (ab)\mathbf{u}$.
10. For any vector $\mathbf{u}$ in V, $1\mathbf{u} = \mathbf{u}$.

Challenge

1. Show that the set of all degree at most n polynomials with real coefficients is a vector space with the usual addition and scalar multiplication,
   (i) $b(a_n x^n + \cdots + a_1 x + a_0) = b a_n x^n + \cdots + b a_1 x + b a_0$,
   (ii) $(a_n x^n + \cdots + a_1 x + a_0) + (b_n x^n + \cdots + b_1 x + b_0) = (a_n + b_n) x^n + \cdots + (a_1 + b_1) x + (a_0 + b_0)$.

2. Show that the set of all $n \times m$ real-valued matrices is a vector space, with the usual matrix addition and scalar multiplication. The set of all $n \times m$ real-valued matrices is sometimes denoted as $\mathbb{R}^{n \times m}$.

3.2 Dot Product, Norm, Distance

Discussion

Matrix addition and scalar multiplication can be applicable directly to vectors. However, how do we, if it's even possible, define the multiplication of vectors?

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ be two (column) vectors. The multiplication

$$\underset{(n \times 1)}{\mathbf{u}} \quad \underset{(n \times 1)}{\mathbf{v}}$$

is undefined.

Multiplying Vectors

We are able to multiply if we transpose one of the vectors.

1. (**Outer Product**) $\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^T = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \cdots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n v_1 & u_n v_2 & \cdots & u_n v_n \end{pmatrix} = (u_i v_j)_n$ (Not part of syllabus)
2. (**Inner Product**) $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^n u_i v_i$. Also known as *dot product*.

Definition

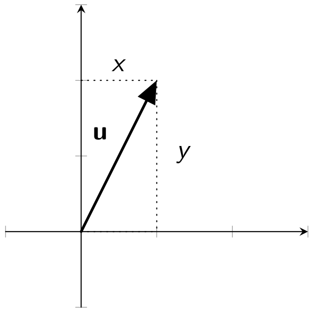
The *inner product* (or *dot product*) of vectors $\mathbf{u} = (u_i)$ and $\mathbf{v} = (v_i)$in $\mathbb{R}^n$ is defined to be

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

Norm in $\mathbb{R}^2$

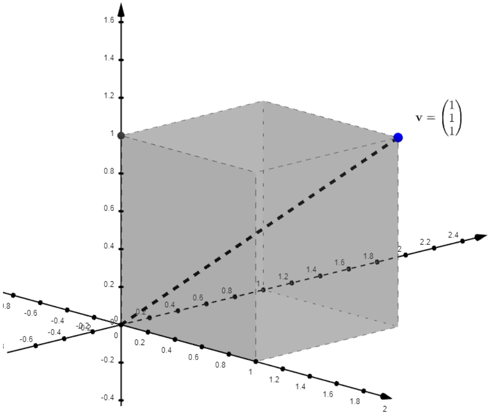
The distance between the point $\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$ and the origin in $\mathbb{R}^2$ is given by

$$\text{distance} = \sqrt{x^2 + y^2}.$$



Question

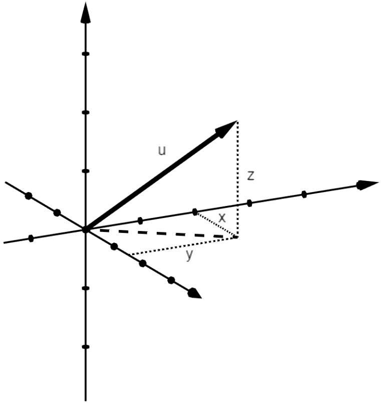
What is the length of the vector $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$?



Norm in $\mathbb{R}^3$

The distance between the point $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ and the origin in $\mathbb{R}^3$ is given by

$$\text{distance} = \sqrt{(\sqrt{x^2 + y^2})^2 + z^2} = \sqrt{x^2 + y^2 + z^2}.$$



Norm in $\mathbb{R}^n$

Definition

The *norm* of a vector $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{u} = (u_i)$, is the square root of the inner product of $\mathbf{u}$ with itself, and is denoted as $\|\mathbf{u}\|$,

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

This is also known as the *length* or *magnitude* of the vector.

Properties of Inner Product and Norm

Theorem

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$, and a, b, c be some scalars.

(i) *(Symmetric)* $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.

(ii) *(Scalar multiplication)* $c\mathbf{u} \cdot \mathbf{v} = (c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v})$.

(iii) *(Distribution)* $\mathbf{u} \cdot (a\mathbf{v} + b\mathbf{w}) = a\mathbf{u} \cdot \mathbf{v} + b\mathbf{u} \cdot \mathbf{w}$.

(iv) *(Positive definite)* $\mathbf{u} \cdot \mathbf{u} \geq 0$ *with equality if and only if* $\mathbf{u} = \mathbf{0}$.

(v) $\|c\mathbf{u}\| = |c|\|\mathbf{u}\|$.

Partial Proof.

Proof for (iv) only. The rest are left as exercise. Let $\mathbf{u} = (u_i)_{n\times 1}$. Since $\mathbf{u}_i \in \mathbb{R}$, $u_i^2 \geq 0$ for all $i = 1, ..., n$. Therefore,

$$\mathbf{u} \cdot \mathbf{u} = u_1^2 + u_2^2 + \cdots + u_n^2 \geq 0.$$

Note also that this is a sum of nonnegative numbers, which is equal to 0 if and only if all the $u_i^2 = 0$, which is equivalent to $u_i = 0$ for all $i = 1, ..., n$. $\square$

Unit Vectors

Definition

A vector $\mathbf{u}$ in $\mathbb{R}^n$ is a *unit vector* if its norm is 1,

$$\|\mathbf{u}\| = 1$$

Example

1. Let $\mathbf{e}_i$ denote the i-th column of the $n \times n$ identity matrix $\mathbf{I}_n$. Then $\mathbf{e}_i$ is a unit vector for all $i = 1, 2, ..., n$.
2. $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ is a unit vector.
3. $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ is not a unit vector; $\frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ is a unit vector pointing in the same direction.

Normalizing a Vector

Let $\mathbf{u}$ be a nonzero vector $\mathbf{u} \neq \mathbf{0}$. By multiplying by the reciprocal of the norm, we get a unit vector,

$$\mathbf{u} \longrightarrow \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

Indeed, $\frac{\mathbf{u}}{\|\mathbf{u}\|}$ is a unit vector,

$$\left(\frac{\mathbf{u}}{\|\mathbf{u}\|}\right) \cdot \left(\frac{\mathbf{u}}{\|\mathbf{u}\|}\right) = \frac{\mathbf{u} \cdot \mathbf{u}}{\|\mathbf{u}\|^2} = 1.$$

This is called *normalizing* $\mathbf{u}$.

Distance Between Vectors

By Pythagorous theorem, the distance between $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ and $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ in $\mathbb{R}^2$ is

$$\text{distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} = \left\| \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\|.$$

Similarly in $\mathbb{R}^3$, the distance between $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$ and $\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$ is

$$\text{distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} = \left\| \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} - \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} \right\|.$$

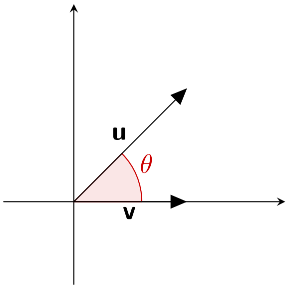
Definition

The *distance* between two vectors $\mathbf{u}$ and $\mathbf{v}$, denoted as $d(\mathbf{u}, \mathbf{v})$, is defined to be

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Angle

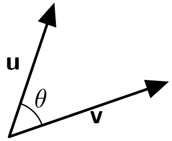
Let $\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}$.



The angle θ between $\mathbf{u}$ and $\mathbf{v}$ is

$$\cos(\theta) = \frac{x}{\sqrt{x^2 + y^2}} = \frac{x x_0}{\|\mathbf{u}\| x_0} = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}.$$

Note that $0 \leq \theta \leq \pi$.



Definition

Define the *angle* θ between two nonzero vectors, $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ to be such that

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}.$$

3.3 Linear Combinations and Linear Spans

Linear Combinations

Definition

Let $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ be vectors in $\mathbb{R}^n$. A *linear combination* of the vectors $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ is

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k,$$

for some $c_1, c_2, ..., c_k \in \mathbb{R}$. The scalars $c_1, c_2, ..., c_k$ are called *coefficients*.

Think of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ as the directions, and $c_1, c_2, ..., c_k$ as the amount of units to walk in the respective directions.

Linear Span

Definition

Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ be vectors in $\mathbb{R}^n$. The *span* of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ is the subset of $\mathbb{R}^n$ containing all the linear combinations of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$,

$$\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\} = \left\{ \; c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k \;\middle|\; c_1, c_2, \dots, c_k \in \mathbb{R} \; \right\}.$$

That is every vector $\mathbf{v}$ in the set $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$,

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k,$$

for some scalars $c_1, c_2, \dots, c_k$.

Algorithm to Check for Linear Combination

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$.

- Form the $n \times k$ matrix $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ whose columns are the vectors in S.
- Then a vector $\mathbf{v}$ in $\mathbb{R}^n$ is in $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ if and only if the system $\mathbf{Ax} = \mathbf{v}$ is consistent.
- If the system is consistent, then the solutions to the system are the possible coefficients of the linear combination. That is, if $\mathbf{u} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ is a solution to $\mathbf{Ax} = \mathbf{v}$, then

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k.$$

Explicitly, $\mathbf{v} \in \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ if and only if $\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v} \end{array} \right)$ is consistent.

When will $\text{span}(S) = \mathbb{R}^n$?

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$. Now instead of checking if a specific vector $\mathbf{v}$ is in $\text{span}(S)$, we may ask if every vector is in the span, that is, whether $\text{span}(S) = \mathbb{R}^n$.

Example

1. $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix} \right\}$. Now we check if every $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is in $\text{span}(S)$.

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 1 & 2 & 3 & y \\ 1 & 1 & 2 & z \end{array} \right) \xrightarrow{RREF} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 2x - y \\ 0 & 1 & 1 & -x + y \\ 0 & 0 & 0 & -x + z \end{array} \right).$$

The system is consistent if and only if $z - x = 0$. This show that not every vector in $\mathbb{R}^3$ is in $\text{span}(S)$, that is, $\text{span}(S) \neq \mathbb{R}^3$. For example, $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ or $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ is not in the span.

When will $\text{span}(S) = \mathbb{R}^n$?

2. Let $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\}$. Is $\text{span}(S) = \mathbb{R}^3$?

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & x \\ 1 & -1 & 2 & y \\ 1 & 0 & 1 & z \end{array} \right) \xrightarrow{RREF} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -x-y+3z \\ 0 & 1 & 0 & x-z \\ 0 & 0 & 1 & x+y-2z \end{array} \right).$$

The system is always consistent regardless of any choice of x, y, z. This show that $\text{span}(S) = \mathbb{R}^3$. In fact, given any $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = (-x-y+3z) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (x-z) \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + (x+y-2z) \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

Discussion

Consider now a vector $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ in $\mathbb{R}^n$. Observe that elementary row operations would not make any entries zero; every entry would still be a linear combination of $x_1, x_2, ..., x_n$.

Example

1. $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} x_1 \\ x_3 \\ x_2 \end{pmatrix}$
2. $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \xrightarrow{R_3 - aR_1} \begin{pmatrix} x_1 \\ x_2 \\ x_3 - ax_1 \end{pmatrix}$
3. $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \xrightarrow{cR_2} \begin{pmatrix} x_1 \\ cx_2 \\ x_3 \end{pmatrix}$, for some $c \neq 0$.

Discussion

This means that in the reduction of $\left(\begin{array}{cccc|c} & & & & x_1 \\ & & & & x_2 \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \vdots \\ & & & & x_n \end{array}\right)$, the entries in the last column will never be 0, but some linear combination of $x_1, x_2, ..., x_n$. In this case, the system is consistent if and only if the reduced row-echelon form of $\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ does not have any zero row.

Algorithm to check if $\text{span}(S) = \mathbb{R}^n$.

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$.

- Form the $n \times k$ matrix $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ whose columns are the vectors in S.
- Then $\text{span}(S) = \mathbb{R}^n$ if and only if the system $\mathbf{Ax} = \mathbf{v}$ is consistent for all $\mathbf{v}$.
- This is equivalent to the reduced row-echelon form of $\mathbf{A}$ having no zero rows.

Explicitly, $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} = \mathbb{R}^n$ if and only if the reduced row-echelon form of $\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ has no zero rows.

Properties of Linear Spans

Theorem (Properties of Linear Spans)

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$.

(i) *The zero vector $\mathbf{0}$ is in $\operatorname{span}(S)$.*

(ii) *The span is closed under scalar multiplication, that is, for any vector $\mathbf{u}$ in $\operatorname{span}(S)$ and scalar α, the vector $\alpha\mathbf{u}$ is a vector in $\operatorname{span}(S)$.*

(iii) *The span is closed under addition, that is, for any vectors $\mathbf{u}, \mathbf{v}$ in $\operatorname{span}(S)$, the sum $\mathbf{u} + \mathbf{v}$ is a vector in $\operatorname{span}(S)$.*

Proof.

We will only provide the main idea of the proof, the details are left to the readers.

(i) $\mathbf{0} = 0\mathbf{u}_1 + 0\mathbf{u}_2 + \cdots + 0\mathbf{u}_k$.

(ii) Write $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$. Then $\alpha\mathbf{v} = (\alpha c_1)\mathbf{u}_1 + (\alpha c_2)\mathbf{u}_2 + \cdots + (\alpha c_k)\mathbf{u}_k$.

(iii) Write $\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$ and $\mathbf{v} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_k\mathbf{u}_k$. Then $\mathbf{u} + \mathbf{v} = (c_1 + d_1)\mathbf{u}_1 + (c_2 + d_2)\mathbf{u}_2 + \cdots + (c_k + d_k)\mathbf{u}_k$.

$\square$

Properties of Linear Spans

Remark

Properties (ii) and (iii) can be combined together into one property (ii'):
The span is closed under linear combinations, that is, if $\mathbf{u}, \mathbf{v}$ are vectors in $\text{span}(S)$ and α, β are any scalars, then the linear combination $\alpha\mathbf{u} + \beta\mathbf{v}$ is a vector in $\text{span}(S)$.

Observe that property (ii') implies that $\text{span}(S)$ is closed under linear combination. That is, suppose $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ are vectors in $\text{span}(S)$, then for any scalars $c_1, c_2, ..., c_m$, the linear combination $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$ is also in the span. For by property (ii'), $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ is in $\text{span}(S)$, and thus by property (ii') again, we have $(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) + c_3\mathbf{v}_3$ is in $\text{span}(S)$ too. Thus, by induction, we can conclude that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$ is in $\text{span}(S)$.
Since this is true for any scalars $c_1, c_2, ..., c_m$, we have arrived at the following corollary.

Corollary (Linear span is closed under linear combinations)

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *be a set of vectors in* $\mathbb{R}^n$. *For any vectors* $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ *in* $\text{span}(S)$, *the span of* $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ *is a subset of* $\text{span}(S)$,

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\} \subseteq \text{span}(S).$$

Algorithm to check for Set Relations between Spans

Now suppose we are given 2 sets of vectors $T = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$.

- By the corollary, if $\mathbf{v}_i \in \text{span}(S)$ for $i = 1, ..., m$, we can conclude that $\text{span}(T) \subseteq \text{span}(S)$.
- Recall that to check if $\mathbf{v}_i \in \text{span}(S)$, we check that the system $\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v}_i \end{array}\right)$ is consistent for all $i = 1, ..., m$.
- There are in total m such linear systems to check. However, since they have the same coefficient matrix, we may combine and check them together, that is, check that

$$\left(\begin{array}{cccc|c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{array}\right)$$

  is consistent.

Algorithm to check for Set Relations between Spans

Theorem

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ be sets of vectors in $\mathbb{R}^n$. Then $\text{span}(T) \subseteq \text{span}(S)$ if and only if $\left(\begin{array}{cccc|c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{array}\right)$ is consistent.

So, to check if $\text{span}(S) = \text{span}(T)$, we check that

- $\text{span}(S) \subseteq \text{span}(T)$, that is,

$$\left(\begin{array}{c|c} \text{“}T\text{”} & \text{“}S\text{”} \end{array}\right) = \left(\begin{array}{cccc|c|c|c|c} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m & \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{array}\right) \text{ is consistent, and}$$

- $\text{span}(T) \subseteq \text{span}(S)$, that is,

$$\left(\begin{array}{c|c} \text{“}S\text{”} & \text{“}T\text{”} \end{array}\right) = \left(\begin{array}{cccc|c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{array}\right) \text{ is consistent.}$$

Challenge

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$. Referring to the properties of a spanning set or otherwise, show that the set $V = \text{span}(S)$ is a (abstract) vector space. That is, it satisfies the 10 axioms of the definition of vector spaces.

3.4 Subspaces

Solution Sets to a Linear system

Recall that the set of solutions to a linear system $\mathbf{Ax} = \mathbf{b}$ is a subset in $\mathbb{R}^n$ (it is the empty set if the system is inconsistent). We may express this set *implicitly* as

$$V = \left\{ \ \mathbf{u} \in \mathbb{R}^n \ \middle| \ \mathbf{Au} = \mathbf{b} \ \right\},$$

or *explicitly* as

$$V = \left\{ \ \mathbf{u} + s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k \ \middle| \ s_1, s_2, ..., s_k \in \mathbb{R} \ \right\},$$

where $\mathbf{u} + s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k,\ s_1, s_2, ..., s_k \in \mathbb{R}$ is the general solution.

Solution Sets to Linear Systems

Write the implicit expression of the following solution set

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} + s \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \;\middle|\; s, t \in \mathbb{R} \right\}.$$

$$\begin{array}{lllllllll} x & = & 1 - 2s + t, & y & = & 2 + s, & z & = & t - 1 \\ x & = & 1 - 2(y - 2) + z + 1, & s & = & y - 2, & t & = & z + 1 \end{array} \Rightarrow x + 2y - z = 6$$

So, implicitly, the set has the expression

$$\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \;\middle|\; x + 2y - z = 6 \right\}.$$

Discussion

Recall that the general solution of a homogeneous system $\mathbf{Ax} = \mathbf{0}$ has the form

$$s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k, \quad s_1, s_2, ..., s_k \in \mathbb{R}.$$

Explicitly, the solution set is

$$V = \left\{ \ s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k \ \middle| \ s_1, s_2, ..., s_k \in \mathbb{R} \ \right\}.$$

Observe however that this is just $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$,

$$V = \left\{ \ s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k \ \middle| \ s_1, s_2, ..., s_k \in \mathbb{R} \ \right\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}.$$

By the properties of a linear span, this would mean that the solution set to a homogeneous system is a vector space that is a subset of the Euclidean vector space. We call a vector space nested inside another vector space a *subspace*.

Subspace

It turns out that for a subset V of the Euclidean space $\mathbb{R}^n$ to satisfy all 10 axioms of being a vector space, suffice for it to satisfies only 3 of them.

Definition

A subset V of $\mathbb{R}^n$ is a *subspace* if it satisfies the following properties.

(i) V contains the zero vector $\mathbf{0} \in V$.

(ii) V is closed under scalar multiplication. For any vector $\mathbf{v}$ in V and scalar α, the vector $\alpha\mathbf{v}$ is in V.

(iii) V is closed under addition. For any vectors $\mathbf{u}, \mathbf{v}$ in V, the sum $\mathbf{u} + \mathbf{v}$ is in V.

Remark

(i) Property (i) can be replaced with property (i'): V is nonempty.

(ii) Properties (ii) and (iii) is equivalent to property (ii'):
V is closed under linear combination. For any $\mathbf{u}, \mathbf{v}$ in V, and scalars α, β, the linear combination $\alpha\mathbf{u} + \beta\mathbf{v}$ is in V.

Solution Space of Homogeneous System

Theorem

The solution set $V = \{ \mathbf{u} \mid \mathbf{A}\mathbf{u} = \mathbf{b} \}$ to a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is a subspace if and only if $\mathbf{b} = \mathbf{0}$, that is, the system is homogeneous.

Proof.

($\Rightarrow$) Suppose $V = \{ \mathbf{u} \mid \mathbf{A}\mathbf{u} = \mathbf{b} \}$ is a subspace. By property (i), it must contain the origin, which means that $\mathbf{0}$ must be a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$. Hence,

$$\mathbf{0} = \mathbf{A}\mathbf{0} = \mathbf{b} \quad \Rightarrow \quad \mathbf{b} = \mathbf{0}.$$

($\Leftarrow$) Suppose $\mathbf{b} = \mathbf{0}$, that is, $V = \{ \mathbf{u} \mid \mathbf{A}\mathbf{u} = \mathbf{0} \}$ the solution set to a homogeneous system.

- Clearly $\mathbf{0} \in V$
- For any $\mathbf{v} \in V$, that $\mathbf{A}\mathbf{v} = \mathbf{0}$, and any $\alpha \in \mathbb{R}$, $\mathbf{A}(\alpha\mathbf{v}) = \alpha\mathbf{A}\mathbf{v} = \alpha\mathbf{0} = \mathbf{0} \Rightarrow \alpha\mathbf{v} \in V$.
- Suppose $\mathbf{u}, \mathbf{v} \in V$, that is $\mathbf{A}\mathbf{u} = \mathbf{0}$ and $\mathbf{A}\mathbf{v} = \mathbf{0}$. Then $\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{A}\mathbf{u} + \mathbf{A}\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0} \Rightarrow \mathbf{u} + \mathbf{v} \in V$.

$\square$

Definition

The solution set to a homogeneous system is call a *solution space*.

Challenge

Prove that if a subset V of $\mathbb{R}^n$ satisfies the 3 criteria of a subspace, then it satisfies all 10 axioms of a vector space.

Equivalent Definition for Subspaces

Theorem

A subset $V \subseteq \mathbb{R}^n$ is a subspace if and only if is a linear span, $V = span(S)$, for some finite set $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$.

Proof.

($\Leftarrow$) This follows from the property of linear span.

($\Rightarrow$) Only present a sketch, details are left as exercise.
Since V is a subspace, it is nonempty. Take a $\mathbf{u}_1 \in V$. If $\text{span}(\mathbf{u}_1) = V$, let $S = \{\mathbf{u}_1\}$. Otherwise, there is a $\mathbf{u}_2 \in V \backslash \text{span}(\mathbf{u}_1)$. If $\text{span}(\mathbf{u}_1, \mathbf{u}_2) = V$, let $S = \{\mathbf{u}_1, \mathbf{u}_2\}$. Otherwise, continue this process to define $\mathbf{u}_i \in V \backslash \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{i-1}\}$. Eventually, the process must stop, that is, there is a $k \in \mathbb{Z}$ such that $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} = V$ (why?). $\square$

Remarks

1. To show that a set V is a subspace, we can either
   (a) find a spanning set, that is find a set S such that $V = \text{span}(S)$, or

   (b) show that V satisfies the 3 conditions of being a subspace.

2. To show that a subset V is not a subspace, we can either
   (i) show that it does not contain the zero vector, $\mathbf{0} \notin V$,

   (ii) find a vector $\mathbf{v} \in V$ and a scalar $\alpha \in \mathbb{R}$ such that $\alpha\mathbf{v} \notin V$, or

   (iii) find vectors $\mathbf{u}, \mathbf{v} \in V$ such that the sum is not in V, $\mathbf{u} + \mathbf{v} \notin V$.

Subspaces of $\mathbb{R}^2$

(i) Zero space: $\left\{\begin{pmatrix}0\\0\end{pmatrix}\right\}$ This is a point.

(ii) Lines, $L = \operatorname{span}\left\{\begin{pmatrix}x_1\\y_1\end{pmatrix}\right\}$ for some fixed $\begin{pmatrix}x_1\\y_1\end{pmatrix} \neq \begin{pmatrix}0\\0\end{pmatrix}$, These are lines, which looks like $\mathbb{R}^1$.

(iii) Whole $\mathbb{R}^2$.

Subspaces of $\mathbb{R}^3$

(i) Zero space: $\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\}$ This is a point.

(ii) Lines: $L = \operatorname{span}\left\{ \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \right\}$ for some fixed $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. These are lines, which looks like $\mathbb{R}^1$.

(iii) Planes, $P = \operatorname{span}\left\{ \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} \right\}$ for some $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$ that are not a scalar multiple of each other, These are planes, which looks like $\mathbb{R}^2$.

(iv) Whole $\mathbb{R}^3$.

Solution Set to Non-homogeneous System

Recall that

$$\mathbf{u} + s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k, s_1, s_2, \ldots, s_k \in \mathbb{R}$$

is a general solution to a consistent non-homogeneous system $\mathbf{Ax} = \mathbf{b}$, $\mathbf{b} \neq \mathbf{0}$ if and only if

$$s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k, s_1, s_2, \ldots, s_k \in \mathbb{R}$$

is a general solution to the homogeneous system $\mathbf{Ax} = \mathbf{0}$, where $\mathbf{u}$ is a particular solution to the non-homogeneous system $\mathbf{Ax} = \mathbf{b}$.

Theorem (Affine Space)

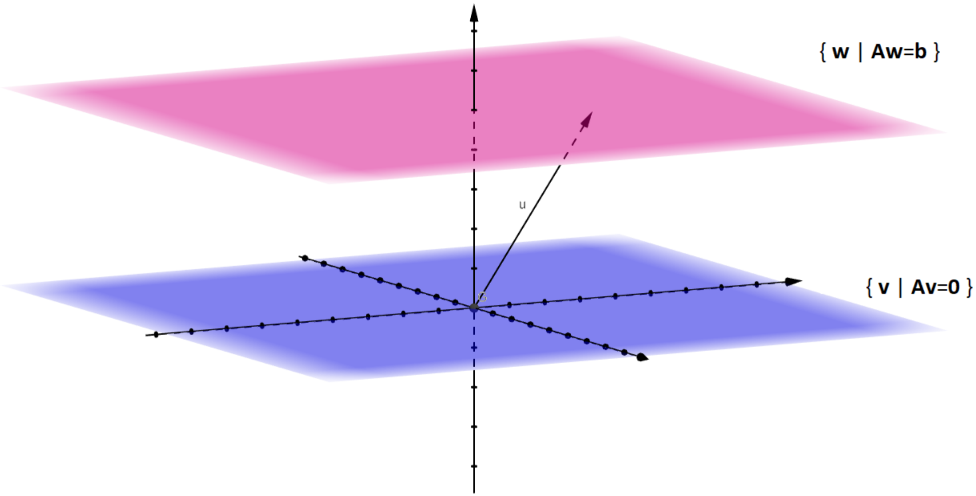
The solution set $W = \left\{ \mathbf{w} \mid \mathbf{Aw} = \mathbf{b} \right\}$ of a non-homogeneous linear system $\mathbf{Ax} = \mathbf{b}$, $\mathbf{b} \neq \mathbf{0}$ is given by

$$\mathbf{u} + V := \left\{ \mathbf{u} + \mathbf{v} \mid \mathbf{v} \in V \right\},$$

where $V = \left\{ \mathbf{v} \mid \mathbf{Av} = \mathbf{0} \right\}$ is the solution space to the associated homogeneous system and $\mathbf{u}$ is a particular solution, $\mathbf{Au} = \mathbf{b}$.

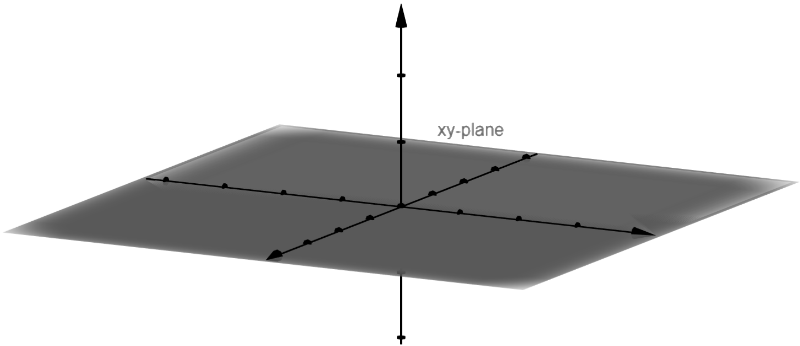
That is, vectors in $\mathbf{u} + V$ are of the form $\mathbf{u} + \mathbf{v}$ for some $\mathbf{v}$ in V.

Solution Set to Linear System



Question

Is $\mathbb{R}^2 \subseteq \mathbb{R}^3$?



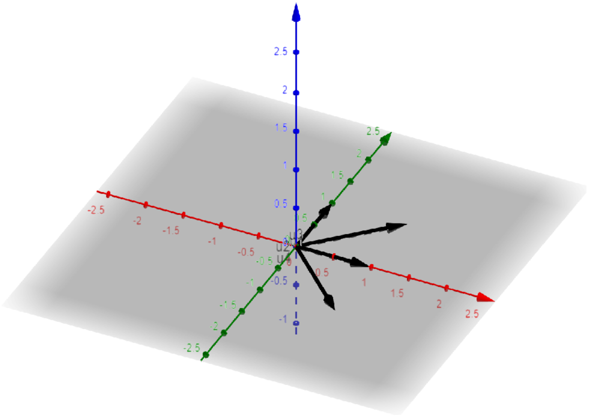
3.5 Linear Independence

Motivation

Consider

$$\mathbf{u}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix}1\\1\\0\end{pmatrix}, \mathbf{u}_4 = \begin{pmatrix}1\\-1\\0\end{pmatrix}.$$

https://www.geogebra.org/m/w2avu5ft



Observe that $V = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2\} \neq \text{span}\{\mathbf{u}_1\}$. This shows that the set $\{\mathbf{u}_1, \mathbf{u}_2\}$ is "optimal"; that is, it is the minimal set to span V. This is because we may use $\mathbf{u}_1 + \mathbf{u}_2$ in place of $\mathbf{u}_3$, and $\mathbf{u}_1 - \mathbf{u}_2$ in place of $\mathbf{u}_4$. Hence, we might say that $\mathbf{u}_3$ and $\mathbf{u}_4$ are "redundant" since they are linear combinations of $\mathbf{u}_1$ and $\mathbf{u}_2$.

Discussion

Now given a set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$.

- A vector $\mathbf{u}_i$ is a redundant vector in the span if it is linearly dependent on the others,
$$\mathbf{u}_i = c_1\mathbf{u}_1 + \cdots + c_{i-1}\mathbf{u}_{i-1} + c_{i+1}\mathbf{u}_{i+1} + \cdots + c_k\mathbf{u}_k.$$
- To check for redundancy, we have to check if the system
$$c_1\mathbf{u}_1 + \cdots + c_{i-1}\mathbf{u}_{i-1} + c_{i+1}\mathbf{u}_{i+1} + \cdots + c_k\mathbf{u}_k = \mathbf{u}_i$$
is consistent for each $i = 1, ..., k$. This is very tedious.
- However, if $\mathbf{u}_i$ is linearly dependent on the other vectors, then we have
$$c_1\mathbf{u}_1 + \cdots + c_{i-1}\mathbf{u}_{i-1} - \mathbf{u}_i + c_{i+1}\mathbf{u}_{i+1} + \cdots + c_k\mathbf{u}_k = \mathbf{0}.$$
- This is a nontrivial solution, and this checks for all $i = 1, .., k$ simultaneously!

Discussion

- For if suppose we are able to find some $c_1, c_2, \ldots, c_k$ not all zero such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0}.$$

- Without lost of generality, say $c_k \neq 0$. Manipulating the equation, we have

$$\frac{c_1}{-c_k}\mathbf{u}_1 + \frac{c_2}{-c_k}\mathbf{u}_2 + \cdots + \frac{c_{k-1}}{-c_k}\mathbf{u}_{k-1} = \mathbf{u}_k,$$

  Then we conclude that $\mathbf{u}_k$ is linearly dependent on $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}\}$.

- If none of the vector is linearly dependent on the others, or that the vectors are linearly independent if we cannot find $c_1, c_2, \ldots, c_k$ not all zero such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0}.$$

Linearly Independent

Definition

A set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is *linearly independent* if the only coefficients $c_1, c_2, ..., c_k$ satisfying the equation

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0},$$

are $c_1 = c_2 = \cdots = c_k = 0$. Otherwise, we say that the set is *linearly dependent*.

Algorithm to Check for Linear Independence

Let $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$.

- $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly independent if and only if the homogeneous system $(\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k)\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- The homogeneous system has only the trivial solution if and only if the reduce row-echelon form of $\begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k\end{pmatrix}$ has no non-pivot column.

Theorem

A subset $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ of $\mathbb{R}^n$ is linearly independent if and only if the reduced row-echelon form of $\mathbf{A} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k\end{pmatrix}$ has no non-pivot columns.

Special Cases

1. $\{\mathbf{0}\}$, where $\mathbf{0} \in \mathbb{R}^n$ is the zero vector is always linearly dependent.

   Take say, $c_1 = 1$, then we have $(1)\mathbf{0} = \mathbf{0}$. Alternatively, the matrix $(\mathbf{0})$ is in RREF and the only column is a non-pivot column.

2. If $\mathbf{v} \neq \mathbf{0}$, then $\{\mathbf{v}\} \in \mathbb{R}^n$ is linearly independent.

   The only solution to $c\mathbf{v} = \mathbf{0}$ is $c = 0$. Alternatively, $(\mathbf{v})$ reduces to the matrix with 1 in the first entry and zero otherwise, and the only column is a pivot column.

3. $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if and only if one is a scalar multiple of the other, $\alpha\mathbf{v}_1 = \mathbf{v}_2$ or $\mathbf{v}_1 = \beta\mathbf{v}_2$.

   $\{\mathbf{v}_1, \mathbf{v}_2\}$ linearly dependent if and only if c_1 or $c_2 \neq 0$. Say $c_1 \neq 0$. Then $\mathbf{v}_1 = -\frac{c_2}{c_1}\mathbf{v}_2$. The argument for $c_2 \neq 0$ is analogous.

4. The empty set $\{\} = \varnothing$ is linearly independent.

   Vacuously true since there are no vector to check.

Linear Dependency and Adding or Removing Vectors

Theorem

Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly dependent set of vectors in $\mathbb{R}^n$. Then for any vector $\mathbf{u}$ in $\mathbb{R}^n$,

$$\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$$

is linearly dependent.

Since the set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly dependent, we can find a say $c_i \neq 0$ such that

$$c_1\mathbf{u}_1 + \cdots + c_i\mathbf{u}_i + \cdots + c_k\mathbf{u}_k = \mathbf{0}.$$

Hence, by adding $0\mathbf{u}$, that is, let $c = 0$, we have

$$c_1\mathbf{u}_1 + \cdots + c_i\mathbf{u}_i + \cdots + c_k\mathbf{u}_k + c\mathbf{u} = \mathbf{0},$$

where not all $c, c_1, ..., c_i, ..., c_k$ are zero.

Hence, any set $\{\mathbf{v}_1, ..., \mathbf{v}_k, \mathbf{0}\}$ containing the zero vector is linearly dependent.

Linear Dependency and Adding or Removing Vectors

Theorem

Suppose $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *is linearly independent set of vectors in* $\mathbb{R}^n$ *and* $\mathbf{u}$ *is not a linearly combination of* $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$*. Then the set* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ *is linearly independent.*

i.e. $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *linearly independent and* $\mathbf{u} \notin \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} \Rightarrow \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ *linearly independent.*

Here is a heuristic explanation. Readers may refer to the appendix for the proof.

Since $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly independent, the RREF of $\begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k\end{pmatrix}$ has no non-pivot column. Now since $\mathbf{u} \notin \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$, the last column of the RREF of $\left(\begin{array}{ccc|c}\mathbf{u}_1 & \cdots & \mathbf{u}_k & \mathbf{u}\end{array}\right)$ is a pivot column. But observe that the LHS of the RREF of $\left(\begin{array}{ccc|c}\mathbf{u}_1 & \cdots & \mathbf{u}_k & \mathbf{u}\end{array}\right)$ is the RREF of $\begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k\end{pmatrix}$. Hence, every column in the RREF of $\left(\begin{array}{ccc|c}\mathbf{u}_1 & \cdots & \mathbf{u}_k & \mathbf{u}\end{array}\right) = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{u}\end{pmatrix}$ is a pivot column. This shows that $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ is linearly independent.

Linear Dependency and Adding or Removing Vectors

Theorem

Suppose $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *is linearly independent set of vectors in* $\mathbb{R}^n$*. Then any subset of* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *is linearly independent.*

If $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ has no redundancy, then it is clear that any subset cannot have redundancy. Readers may refer to the appendix for the proof.

3.6 Basis and Coordinates

Motivation

Consider the set $E = \left\{ \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$. It is clear that any vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ in $\mathbb{R}^3$ can be unique written as a linear combination of the vectors in E,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

In fact, we call x, y, z the coordinates of the vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. However, the set E is not the only set that enjoys this property.

Motivation

Consider the set $B = \left\{ \mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$. Now let $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ be a vector in $\mathbb{R}^3$. Then

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 1 & 0 & 1 & y \\ 0 & 1 & 1 & z \end{array} \right) \xrightarrow{RREF} \left(\begin{array}{ccc|c} 1 & 0 & 0 & (x+y-z)/2 \\ 0 & 1 & 0 & (x-y+z)/2 \\ 0 & 0 & 1 & (y-x+z)/2 \end{array} \right)$$

tells us that the linear combination

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{x+y-z}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \frac{x-y+z}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{-x+y+z}{2} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

is unique.

Motivation

On the other hand, consider the set $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$. The vector $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ is not a linear combination of the vectors in S,

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{array} \right) \xrightarrow{RREF} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right).$$

This shows that $\text{span}(S) \neq \mathbb{R}^3$.

Motivation

Consider another set $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$. Check that the span of S is indeed the whole $\mathbb{R}^3$, $\text{span}(S) = \mathbb{R}^3$. However, the linear combination is not unique. For example, consider the vector $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$,

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 2 \\ 1 & 1 & 0 & 1 & 1 \end{array} \right) \xrightarrow{RREF} \left(\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \end{array} \right)$$

tells us that

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = (1-s)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (1+s)\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + s\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - (1+s)\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

for any $s \in \mathbb{R}$. Observe that this is because the set S is not linearly independent.

Motivation

Consider now the solution space $V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \; x + y - 2z = 0 \right\}$. Since it is a subspace of $\mathbb{R}^3$, it is a vector space itself. Explicitly, we have

$$V = \left\{ s \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \middle| \; s, t \in \mathbb{R} \right\} = \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Open GeoGebra: https://geogebra.org/3d.

1. Type in $x + y - 2z = 0$, enter.
2. Type in $u1 = (-1, 1, 0)$ and hit enter, and $u2 = (2, 0, 1)$ and hit enter.
3. It is easy to see that every vector in V can be written uniquely as a linear combination of the $\mathbf{u}_1$ and $\mathbf{u}_2$.

Motivation

Let $V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \, y - z = 0 \right\} = \left\{ s \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \middle| \, s, t, \in \mathbb{R} \right\} = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$. Check that the set $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$ spans V. However, the vector $\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ in V can be written as as linear combination of vectors in S in more than one way,

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

Observe that the set S is linearly dependent.

Basis

Definition

Let V be a subspace of $\mathbb{R}^n$. A set $S = \{\mathbf{u}_1, \cdots \mathbf{u}_k\} \subseteq V$ is a *basis* for V if

(i) S spans V, $\text{span}(S) = V$, and

(ii) S is linearly independent.

Theorem

Suppose S is a basis for V. then every vectors $\mathbf{v} \in V$ can be written as a linear combination of vectors in S uniquely.

Proof.

(i) $\text{span}(S) = V$ tells us that every vector $\mathbf{v} \in V$ can be written as a combination of vectors in S.

(ii) S is linearly independent tells us that if $\mathbf{v}$ is a linear combination of vectors in S, the coefficient is unique.

$$\begin{aligned} & \mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = d_1\mathbf{u}_1 + \cdots + d_k\mathbf{u}_k \\ \Leftrightarrow \quad & (c_1 - d_1)\mathbf{u}_1 + \cdots + (c_k - d_k)\mathbf{u}_k = \mathbf{0} \\ \Leftrightarrow \quad & c_1 = d_1, \quad \cdots \quad c_k = d_k \end{aligned}$$

$\square$

Basis for Solution Set of Homogeneous System

Let $V = \{ \ \mathbf{u} \ \big| \ \mathbf{Au} = \mathbf{0} \ \}$ be the solution space to the homogeneous system $\mathbf{Ax} = \mathbf{0}$. Suppose

$$s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k, \quad s_1, s_2, \ldots, s_k \in \mathbb{R}$$

is the general solution. Then $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is a basis for the subspace $V = \{ \ \mathbf{u} \ \big| \ \mathbf{Au} = \mathbf{0} \ \}$.

Basis for the zero space $\{\mathbf{0}\}$

Recall that the zero space $\{\mathbf{0}\}$ is a subspace. Find a basis for $\{\mathbf{0}\}$

The basis for the zero space $\{\mathbf{0}\}$ is the empty set $\{\}$ or $\emptyset$.

- Firstly, $\text{span}\{\mathbf{0}\} = \{\mathbf{0}\}$ but the set $\{\mathbf{0}\}$ is not linearly independent.
- However, if S is a set that contains any nonzero vector, then $\text{span}(S)$ will be strictly bigger than the zero space, $\{\mathbf{0}\} \subsetneqq \text{span}(S)$.
- The empty set is linearly independent vacuously.
- However, $\text{span}\{\}$ does not make sense.
- The real definition of the span of S is the smallest subspace V such that $S \subseteq V$. That is $V = \text{span}(S)$ if $V \subseteq W$ for all subspaces W containing S.
- Since the zero space is the smallest subspace containing the empty set, span of the empty set is the zero space.

Basis for $\mathbb{R}^n$ and Invertibility

A priori, there is no relationship between linear independence and spanning a subspace. However, in the special case when the subset S of $\mathbb{R}^n$ contains exactly n vectors, then linear independence is equivalent to spanning $\mathbb{R}^n$.

Theorem

A $n \times n$ square matrix $\mathbf{A}$ is invertible if and only if the columns are linearly independent.

Proof.

Write $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ and let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ be the set containing the columns of $\mathbf{A}$. Then $\mathbf{A}$ is invertible if and only if the reduce row-echelon form is the identity matrix. But we have also seen that S is linearly independent if and only if the reduce row-echelon form of $\mathbf{A}$ has no non-pivot columns, which for a square matrix, must mean that the reduce row-echelon form is the identity matrix. $\square$

Theorem

A $n \times n$ square matrix $\mathbf{A}$ is invertible if and only if the columns spans $\mathbb{R}^n$.

Proof.

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ be the set containing the columns of $\mathbf{A}$. Then S spans $\mathbb{R}^n$ if and only if the reduced row-echelon form of $\mathbf{A}$ do not have any nonzero row, which for a square matrix, would mean that the reduce row-echelon form is the identity matrix. This is equivalent to $\mathbf{A}$ being invertible. $\square$

Basis for $\mathbb{R}^n$ and Invertibility

Corollary

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ be a subset of $\mathbb{R}^n$ containing n vectors. Then S is linearly independent if and only if S spans $\mathbb{R}^n$.

Proof.

Let $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ be the matrix whose columns are the vectors in S. Then $\mathbf{A}$ is a square matrix. Then by the two theorems, S is linearly independent if and only if $\mathbf{A}$ is invertible, if and only if S spans $\mathbb{R}^n$. $\square$

Corollary

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a subset of $\mathbb{R}^n$ and $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ be the matrix whose columns are vectors in S. Then S is a basis for $\mathbb{R}^n$ if and only if $k = n$ and $\mathbf{A}$ is an invertible matrix.

Proof.

($\Rightarrow$) If $k < n$, then S cannot span $\mathbb{R}^n$. If $k > n$, then S cannot be linearly independent. Hence, if S is a basis, S must have exactly n vectors, and by the previous theorem, $\mathbf{A}$ must be invertible.
($\Leftarrow$) Conversely, if $k = n$ and $\mathbf{A}$ is invertible, then S is a basis by the previous theorem. $\square$

Basis for $\mathbb{R}^n$ and Invertibility

Theorem

A $n \times n$ square matrix $\mathbf{A}$ invertible if and only if the rows of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.

Theorem

A square matrix $\mathbf{A}$ of order n is invertible if and only if the rows of $\mathbf{A}$ are linearly independent.

The proofs of the 2 theorems follow from the fact that $\mathbf{A}$ is invertible if and only if $\mathbf{A}^T$ is, and the rows of $\mathbf{A}$ are the columns of $\mathbf{A}^T$.

Equivalent Statements for Invertibility

Theorem

Let $\mathbf{A}$ be a square matrix of order n. The following statements are equivalent.

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}^T$ *is invertible.*

(iii) *(left inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{BA} = \mathbf{I}$.

(iv) *(right inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{AB} = \mathbf{I}$.

(v) *The reduced row-echelon form of* $\mathbf{A}$ *is the identity matrix.*

(vi) $\mathbf{A}$ *can be expressed as a product of elementary matrices.*

(vii) *The homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution.*

(viii) *For any* $\mathbf{b}$, *the system* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution.*

(ix) *The determinant of* $\mathbf{A}$ *is nonzero,* $\det(\mathbf{A}) \neq 0$.

(x) *The columns/rows of* $\mathbf{A}$ *are linearly independent.*

(xi) *The columns/rows of* $\mathbf{A}$ *spans* $\mathbb{R}^n$.

Challenge

Recall that the set of 2×2 matrices, $\mathbb{R}^{2\times 2}$, is a vector space. Show that the set

$$\left\{ \mathbf{M}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{M}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \mathbf{M}_4 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \mathbf{M}_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

is a basis for $\mathbb{R}^{2\times 2}$.

Introduction to Coordinates Relative to a Basis

Let $V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \; z = 0 \right\} = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$. Observe that any vector $\begin{pmatrix} x \\ y \end{pmatrix}$ in $\mathbb{R}^2$ identifies with a unique vector $x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$ in V.

Let $T = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$, it is also a basis for V.

- Now a vector $\begin{pmatrix} x \\ y \end{pmatrix}$ in $\mathbb{R}^2$ defines a vector $x \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} x+y \\ x-y \\ 0 \end{pmatrix}$ in V.
- Conversely, a vector $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \frac{x+y}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \frac{x-y}{2} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ in V defines a vector $\begin{pmatrix} (x+y)/2 \\ (x-y)/2 \end{pmatrix}$ in $\mathbb{R}^2$.

Introduction to Coordinates Relative to a Basis

Let $V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \; x + y - z = 0 \right\} = \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$. Then we have the unique correspondence

$$\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \quad \longleftrightarrow \quad x \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} y - x \\ x \\ y \end{pmatrix} \in V.$$

- These examples demonstrate that a subspace V of $\mathbb{R}^n$ can be identified with some $\mathbb{R}^k$.
- That is, instead of giving a vector in V in terms of its coordinates in $\mathbb{R}^n$, we may represent it with a vector in $\mathbb{R}^k$ for some $k \leq n$.
- This identification depends on the choice of basis of V.
- Explicitly, let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a basis for V, a subspace of $\mathbb{R}^n$. Then we have a unique correspondence

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} \in \mathbb{R}^k \longleftrightarrow \mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k \in V.$$

Coordinates Relative to a Basis

Definition

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for V, a subspace of $\mathbb{R}^n$ and

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$$

be the unique expression of a vector $\mathbf{v}$ in V in terms of the basis S. The vector in $\mathbb{R}^k$ defined by the coefficients of the linear combination is called the *coordinates of $\mathbf{v}$ relative to basis S*, and is denoted as

$$[\mathbf{v}]_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}.$$

Remarks

- Even though $\mathbf{v} \in V \subseteq \mathbb{R}^n$ has n coordinates, its coordinates relative to basis S, $[\mathbf{v}]_S$, has k coordinates if the basis S has k vectors.
- Note that the correspondence is unique only if S is a basis. If S is not linearly independent, a few vectors in $\mathbb{R}^k$ can map to the same $\mathbf{v} \in V$.
- The relative coordinates depend on the ordering of the basis. If $S = \left\{ \mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ and

  $T = \left\{ \mathbf{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$, then for $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$,

$$[\mathbf{v}]_S = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \neq \begin{pmatrix} 2 \\ 1 \end{pmatrix} = [\mathbf{v}]_T.$$

Algorithm for Computing Relative Coordinate

Let V be a subspace of $\mathbb{R}^n$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for V.

- Let $\mathbf{v}$ be a vector in V. To find $[\mathbf{v}]_S$, we must find the coefficients $c_1, c_2, ..., c_k$ such that

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k.$$

- Converting it to a matrix equation, we have

$$\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots \mathbf{u}_k \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = \mathbf{v},$$

- which is equivalent to solving the linear system

$$\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v} \end{array} \right).$$

Properties of Coordinates Relative to a Basis

Theorem

Let V be a subspace of $\mathbb{R}^n$ and B a basis for V.

(i) *For any vectors $\mathbf{u}, \mathbf{v} \in V$, $\mathbf{u} = \mathbf{v}$ if and only if $[\mathbf{u}]_B = [\mathbf{v}]_B$.*

(ii) *For any $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m \in V$,*

$$[c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m]_B = c_1[\mathbf{v}_1]_B + c_2[\mathbf{v}_2]_B + \cdots + c_m[\mathbf{v}_m]_B.$$

Proof.

Exercise. $\square$

3.7 Dimensions

Introduction

- Intuitively, we say that $\mathbb{R}^3$ is 3-dimensional, and $\mathbb{R}^2$ is 2-dimensional.
- Let $V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \; z = 0 \right\} = \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$. By the discussion in coordinates relative to a basis, we can identify V with $\mathbb{R}^2$, and hence intuitively say that V is 2-dimensional.
- However, the identification of V with $\mathbb{R}^k$ depends on the choice of the basis of V.
- Recall that bases for any nonzero subspace $V \neq \{\mathbf{0}\}$ is not unique.
- So suppose now $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ are bases for a subspace V. Using S, we identify V with $\mathbb{R}^k$ and using T, we identify V with $\mathbb{R}^m$. Then do we say that V is k-dimensional, or m-dimensional?
- Ideally, we want $m = k$, which is in fact true!

More Properties of Coordinates Relative to a Basis

Theorem

Let V be a subspace of $\mathbb{R}^n$ and B a basis for V. Suppose B contains k vectors, $|B| = k$. Let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ be vectors in V. Then

(i) *$\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ is linearly independent (respectively, dependent) if and only if $[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, ..., [\mathbf{v}_m]_B$ is linearly independent (respectively, dependent) in $\mathbb{R}^k$; and*

(ii) *$\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ spans V if and only if $\{[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, ..., [\mathbf{v}_m]_B\}$ spans $\mathbb{R}^k$.*

The proof is given in the appendix, we will provide a heuristic of the proof. By the properties of coordinates of a vector relative to a basis, we have that the linear system $\left(\begin{array}{cccc|c} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m & \mathbf{u} \end{array}\right)$ has the exact same properties as the linear system $\left(\begin{array}{cccc|c} [\mathbf{v}_1]_S & [\mathbf{v}_2]_S & \cdots & [\mathbf{v}_m]_S & [\mathbf{u}]_S \end{array}\right)$. So, let $\mathbf{u} = \mathbf{0}$ to proof property (i). For property (ii), let $[\mathbf{u}]_S$ be a vector in $\mathbb{R}^k$ to prove ($\Rightarrow$), and let $\mathbf{u}$ be a vector in V to prove ($\Leftarrow$).

Dimension

Corollary

Let V be a subspace of $\mathbb{R}^n$ and B a basis for V. Suppose B contains k vectors, $|B| = k$.

(i) *If $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ is a subset of V with $m > k$, then S is linearly dependent.*

(ii) *If $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ is a subset of V with $m < k$, then S is cannot span V.*

Proof.

Consider the set $T = \{[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, ..., [\mathbf{v}_m]_B\}$ in $\mathbb{R}^k$. If $m > k$, then T is linearly dependent, and hence by the previous theorem, so is S. If $m < k$, then T cannot span $\mathbb{R}^k$, and so by the previous theorem, S cannot span V. □

Dimension

Corollary

Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *and* $T = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ *are bases for a subspace* $V \subseteq \mathbb{R}^n$. *Then* $k = m$.

Proof.

Exercise. □

Definition

Let V be a subspace of $\mathbb{R}^n$. The ***dimension*** of V, denoted by $\dim(V)$, is defined to be the number of vectors in any basis of V.

In other words, V is k-dimensional if and only if V identifies with $\mathbb{R}^k$ using coordinates relative to any basis B of V.

Dimension of the Zero Space $\{\mathbf{0}\}$

We will provide a intuitive reasoning of why the empty set is the basis for the zero space $\{\mathbf{0}\}$ in $\mathbb{R}^n$.

- Intuitively, the dimension is the independent degree of freedom of movement: In a 3 dimensional space, we can travel forwards backwards, side ways, and up and down; in a 2-dimensional space, we can travel forwards backwards, as well as side ways; in a 1-dimensional space, we can only walk forward or backwards.
- So, since we have no freedom of movement in the zero space, the zero space should be 0-dimensional.
- But this would tell us that by definition, the basis for the zero space must have no vectors, that is, it must be the empty set.

Dimension of Solution Space

Recall that the vectors in the general solution of a homogeneous system form a basis for the solution space. This means that the dimension of the solution space is equal to the number of parameters in the general solution. This is in turn equal to the number of non-pivot columns in the reduce row-echelon form of the coefficient matrix.

Theorem

Let $\mathbf{A}$ *be a* $m \times n$ *matrix. The number of non-pivot columns in the reduced row-echelon form of* $\mathbf{A}$ *is the dimension of the solution space*

$$V = \{\ \mathbf{u} \in \mathbb{R}^n \ \big|\ \mathbf{A}\mathbf{u} = \mathbf{0}\ \}.$$

Let $s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$ be the general solution to the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$. Then $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is a basis for V and so by definition, $\dim(V) = k$. But this means that the reduced row-echelon form of $\mathbf{A}$ has k non-pivot columns.

Spanning Set Theorem

Theorem

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *be a subset of vectors in* $\mathbb{R}^n$*, and let* $V = \text{span}(S)$*. Suppose* V *is not the zero space,* $V \neq \{\mathbf{0}\}$*. Then there must be a subset of* S *that is a basis for* V*.*

Proof.

If S is linearly independent, then S is a basis for V. Otherwise, one of the vectors $\mathbf{u}_i$ in S can be written as a linear combination of the other. Without lost of generality (rearranging if necessary), say

$$\mathbf{u}_k = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_{k-1}\mathbf{u}_{k-1}$$

for some coefficients $c_1, c_2, ..., c_{k-1}$. We claim that $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{k-1}\}$ still spans V. For if $\mathbf{v}$ is a vector in V, we have

$$\begin{aligned} \mathbf{v} &= a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_k\mathbf{u}_k \\ &= a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_k(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_{k-1}\mathbf{u}_{k-1}) \\ &= (a_1 + a_kc_1)\mathbf{u}_1 + (a_2 + a_kc_2)\mathbf{u}_2 + \cdots + (a_{k-1} + a_kc_{k-1})\mathbf{u}_{k-1} \end{aligned}$$

which shows that $\mathbf{v}$ is a linear combination of vectors in $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{k-1}\}$. If $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{k-1}\}$ is linearly independent, then it is a basis for V. Otherwise, continue the process of throwing away some redundant vectors, we can conclude that there must be a subset of S that is a basis for V. □

Linear Independence Theorem

Theorem

Let V be a subspace of $\mathbb{R}^n$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ a linearly independent subset of V, $S \subseteq V$. Then there must be a set T containing S, $S \subseteq T$ such that T is a basis for V.

Proof.

If $\text{span}(S) = V$, then S is a basis for V. Otherwise, since $\text{span}(S) \subsetneqq V$, there must be a vector in V that is not contained in $\text{span}(S)$, $\mathbf{u}_{k+1} \in V \setminus \text{span}(S)$. Note that since $\mathbf{u}_{k+1} \notin \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$, the set $S_1 = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}_{k+1}\}$ is linearly independent and $\dim(\text{span}(S_1)) = k + 1$. If $\text{span}(S_1) = V$, we are done. Otherwise, repeating the argument above, we can find $\mathbf{u}_{k+2}$ in V such that $S_2 = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}_{k+1}, \mathbf{u}_{k+2}\}$ is a linearly independent subset of V. Continue inductively, this process must stop when the number of vectors in S_m is equal to the dimension of V, for otherwise, if $|S_m| > \dim(V)$, then S_m cannot be linearly independent. So let $T = S_m$ when $|S_m| = \dim(V)$. $\square$

Challenge

Let V be a k-dimensional subspace of $\mathbb{R}^n$. Using the dimension of V (instead of proving using equivalent statements of invertibility), prove that a subset S in V containing k vectors, $|S| = k$, is linearly independent if and only if it spans V.

Discussion

Recall that for a set S to be a basis for a subspace V in $\mathbb{R}^n$, we must check that

(i) $\text{span}(S) = V$, and

(ii) S is linearly independent.

However, if we know the dimension of V and if the number of vectors in the set S is equal to the dimension of V, $|S| = \dim(V)$, then it suffice to check one of the above criteria.

Dimension and Subspaces

Theorem

Let U and V be subspaces of $\mathbb{R}^n$.

(i) *If U is a subset of V, $U \subseteq V$, then the dimension of U is no greater than the dimension of V, $\dim(U) \leq \dim(V)$.*

(ii) *If U is a strict subset of V, $U \subsetneq V$, then the dimension of U is strictly smaller than V, $\dim(U) < \dim(V)$.*

i.e. $U \subseteq V$, then $\dim(U) \leq \dim(V)$ with equality $\Leftrightarrow U = V$.

Sketch of Proof.

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for U. Then $\dim(U) = k$. Since U is a subset of V, S is a linearly independent subset of V. So necessary $\dim(V) \geq k$. If $U \neq V$, then we can find a set T strictly bigger than S, $S \subsetneq T$ such that T is a basis for V. Hence, $\dim(V) = |T| > |S| = k = \dim(U)$. $\square$

Equivalent ways to check for Basis

Theorem (B1)

Let V be a k-dimensional subspace of $\mathbb{R}^n$, $\dim(V) = k$. Suppose S is a linearly independent subset of V containing k vectors, $|S| = k$. Then S is a basis for V.

Proof.

Let $U = \text{span}(S)$. Since S is linearly independent, S is a basis for U, and hence, $\dim(U) = k$. Since $S \subseteq V$, $U \subseteq V$. Also, $\dim(U) = k = \dim(V)$. Therefore, $U = V$, and so S is a basis for V. $\square$

Theorem (B2)

Let V be a k-dimensional subspace of $\mathbb{R}^n$, $\dim(V) = k$. Suppose S is a set containing k vectors, $|S| = k$, such that $V \subseteq \text{span}(S)$. Then S is a basis for V.

Proof.

Let $U = \text{span}(S)$, then $V \subseteq U$. So, $k = \dim(V) \leq \dim(U) \leq k$ which shows that $k = \dim(U)$ and hence $V = U = \text{span}(S)$. Next, observe that S must be linearly independent. For if S is linearly dependent, then $k = \dim(U) = \dim(\text{span}(S)) < k$, a contradiction. $\square$

Equivalent ways to check for basis

In summary

Definition	**(B1)**	**(B2)**
(1) $\text{span}(S) = V$ (2) S is L.I.	(1) $\lvert S\rvert = \dim(V)$ (2) $S \subseteq V$ (3) S is Linearly independent	(1) $\lvert S\rvert = \dim(V)$ (2) $V \subseteq \text{span}(S)$

- Using (B1), we do not need to check that $\text{span}(S) = V$.
- Using (B2), we do not need to check that S is linearly independent.

3.8 Transition Matrices

Introduction

Let $V \subseteq \mathbb{R}^n$ be a subspace. Suppose $S = \{\mathbf{u}_1, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are bases for V. For a $\mathbf{v} \in V$, how are $[\mathbf{v}]_S$ and $[\mathbf{v}]_T$ related?

$$
\begin{tikzcd}
{[\mathbf{v}]_S \in \mathbb{R}^k} \arrow[r, leftrightarrow] \arrow[rr, leftrightarrow, bend right] & \mathbf{v} \in V \subseteq \mathbb{R}^n \arrow[r, leftrightarrow] & {[\mathbf{v}]_T \in \mathbb{R}^k}
\end{tikzcd}
$$

$$
[\mathbf{v}]_S = \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} \quad \longleftrightarrow \quad c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = \mathbf{v} = d_1\mathbf{v}_1 + \cdots + d_k\mathbf{v}_k \quad \longleftrightarrow \quad [\mathbf{v}]_T = \begin{pmatrix} d_1 \\ \vdots \\ d_k \end{pmatrix}
$$

Transition Matrix

Definition

Let V be a subspace of $\mathbb{R}^n$. Suppose $S = \{\mathbf{u}_1, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are bases for the subspace V. Define the *transition matrix* from T to S to be

$$\mathbf{P} = \begin{pmatrix} [\mathbf{v}_1]_S & [\mathbf{v}_2]_S & \cdots & [\mathbf{v}_k]_S \end{pmatrix},$$

the matrix whose columns are the coordinates of the vectors in T relative to the basis S.

Theorem

Let V be a subspace of $\mathbb{R}^n$. Suppose $S = \{\mathbf{u}_1, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are bases for the subspace V. Let $\mathbf{P}$ be the transition matrix from T to S. Then for any vector $\mathbf{w}$ in V,

$$[\mathbf{w}]_S = \mathbf{P}[\mathbf{w}]_T.$$

Heuristic of the Proof

- Let $\mathbf{e}_i$ denote the i-th column of the $k \times k$ identity matrix $\mathbf{I}_k$.
- Recall that for a $m \times k$ matrix $\mathbf{A}$, the product $\mathbf{Ae}_i$ is the i-th column of $\mathbf{A}$,
$$\mathbf{Ae}_i = \mathbf{a}_i.$$
- Suppose $S = \{\mathbf{u}_1, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are bases for a subspace V of $\mathbb{R}^n$ and let $\mathbf{P}$ be the transition matrix from T to S.
- Note that since $\mathbf{v}_i = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_{i-1} + \mathbf{v}_i + 0\mathbf{v}_{i+1} + \cdots + 0\mathbf{v}_k$, $[\mathbf{v}_i]_T = \mathbf{e}_i$.
- Hence,
$$[\mathbf{v}_i]_S = \mathbf{P}[\mathbf{v}_i]_T = \mathbf{Pe}_i \text{ is the } i\text{-th column of } \mathbf{P}.$$

Algorithm to find Transition Matrix

Let $S = \{\mathbf{u}_1, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be basis for a subspace V in $\mathbb{R}^n$.

- To find $\mathbf{P}$, the transition matrix from T to S, we need to find $[\mathbf{v}_i]_S$ for $i = 1, 2, ..., k$.
- This is equivalent to solving $\left(\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v}_i \end{array}\right)$ for $i = 1, 2, ..., k$.
- Since these linear systems have the same coefficient matrix, we can solve them simultaneously,

$$\left(\begin{array}{cccc|cccc} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{array}\right).$$

- Now since S is a basis, the system must have a unique solution, and the reduced row-echelon form of the augmented matrix above will be of the form

$$\left(\begin{array}{c|cccc} \mathbf{I}_k & [\mathbf{v}_1]_S & [\mathbf{v}_2]_S & \cdots & [\mathbf{v}_k]_S \\ \mathbf{0}_{(n-k)\times k} & \mathbf{0} & \mathbf{0} & & \mathbf{0} \end{array}\right) = \left(\begin{array}{c|c} \mathbf{I}_k & \mathbf{P} \\ \mathbf{0}_{(n-k)\times k} & \mathbf{0}_{(n-k)\times k} \end{array}\right)$$

  where $\mathbf{P}$ is the transition matrix from T to S.

In summary,

$$\left(\begin{array}{c|c} \text{“}S\text{”} & \text{“}T\text{”} \end{array}\right) = \left(\begin{array}{cccc|cccc} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{array}\right) \xrightarrow{rref} \left(\begin{array}{c|c} \mathbf{I}_k & \mathbf{P} \\ \mathbf{0}_{(n-k)\times k} & \mathbf{0}_{(n-k)\times k} \end{array}\right),$$

Question

Let $V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \; x - 2y + z = 0 \right\}$. $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ and $T = \left\{ \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \right\}$ are bases for V (check).

Given that

$$\begin{pmatrix} 0 & 2 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 2 & 0 & -1 & 1 \end{pmatrix} \xrightarrow{rref} \begin{pmatrix} 1 & 0 & -1/2 & 1/2 \\ 0 & 1 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Which statement is true?

(a) The transition matrix from T to S is $\mathbf{P} = \begin{pmatrix} -1/2 & 1/2 \\ 1/2 & 1/2 \\ 0 & 0 \end{pmatrix}$.

(b) The transition matrix from S to T is $\mathbf{P} = \begin{pmatrix} -1/2 & 1/2 \\ 1/2 & 1/2 \\ 0 & 0 \end{pmatrix}$.

(c) None of the other options are true.

Inverse of Transition Matrix

Theorem

Suppose $S = \{\mathbf{u}_1, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are bases for a subspace $V \subseteq \mathbb{R}^n$. Let $\mathbf{P}$ be the transition matrix from T to S. Then $\mathbf{P}^{-1}$ is the transition matrix from S to T.

Proof.

Exercise. Note that you cannot assume that $\mathbf{P}$ is invertible. $\square$

Appendix

Linear Dependency and Adding Vectors

Theorem

Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly independent set of vectors in $\mathbb{R}^n$ and $\mathbf{u}$ is not a linearly combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$. Then the set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ is linearly independent.

Proof.

Let $c_1, c_2, ..., c_k, c$ be coefficients satisfying the equation

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k + c\mathbf{u} = \mathbf{0}.$$

If $c \neq 0$, then manipulating the equation gives

$$-\frac{c_1}{c}\mathbf{u}_1 - \frac{c_2}{c}\mathbf{u}_2 - \cdots - \frac{c_k}{c}\mathbf{u}_k = \mathbf{u},$$

a contradiction to $\mathbf{u}$ not being a linear combination of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$. So, necessarily $c = 0$. Then

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k\mathbf{0}$$

tells us that $c_1 = c_2 = \cdots = c_k = 0$ by the independence of S. Therefore only the trivial coefficients satisfy the equation above, which proves that the set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}\}$ is linearly independent $\square$

Linear Dependency and Adding or Removing Vectors

Theorem

Suppose $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly independent set of vectors in $\mathbb{R}^n$. Then any subset of $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is linearly independent.

Proof.

Let $\{\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, ..., \mathbf{u}_{i_l}\}$ be a subset of $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$. Relabel index, or rearranging the vectors in the set, we may assume that the subset is $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_l\}$ for some $l \leq k$. Suppose $c_1, c_2, ..., c_l$ are coefficients satisfying the equation

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_l\mathbf{u}_l = \mathbf{0}.$$

Pad the equation by 0, we have

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_l\mathbf{u}_l + 0\mathbf{u}_{l+1} + \cdots + 0\mathbf{u}_k = \mathbf{0}.$$

This is a linear combination of the vectors $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$, and since the set is indepenendent, necessary the coefficients are 0. In particular, $c_1 = c_2 = \cdots = c_l = 0$, which proves that the set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_l\}$ is independent. $\square$

More Properties of Coordinates Relative to a Basis

Theorem

Let B be a basis for V containing k vectors, $|B| = k$. Let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ be vectors in V. Then

(i) *$\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$ is linearly independent (respectively, dependent) if and only if $[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, ..., [\mathbf{v}_m]_B$ is linearly independent (respectively, dependent) in $\mathbb{R}^k$; and*

(ii) *$\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ spans V if and only if $\{[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, ..., [\mathbf{v}_m]_B\}$ spans $\mathbb{R}^k$.*

Proof.

(i) Follows from the properties of coordinates relative to a basis, $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_m = \mathbf{0}_{n\times 1}$ if and only if

$$\mathbf{0}_{k\times 1} = [\mathbf{0}_{n\times 1}]_B = [c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_m]_B = c_1[\mathbf{v}_1]_B + c_2[\mathbf{v}_2]_B + \cdots + c_k[\mathbf{v}_m]_B.$$

More Properties of Coordinates Relative to a Basis

Continue of Proof.

(ii) ($\Leftarrow$) Suppose $\{[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, ..., [\mathbf{v}_m]_B\}$ spans $\mathbb{R}^k$. Given any $\mathbf{v} \in V$, $[\mathbf{v}]_B \in \mathbb{R}^k$ and so can find $c_1, ..., c_m$ such that $[\mathbf{v}]_B = c_1[\mathbf{v}_1]_B + c_2[\mathbf{v}_2]_B + \cdots + c_m[\mathbf{v}_m]_B$ in $\mathbb{R}^k$. Then $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$, which proves that $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ spans V.

($\Rightarrow$) Let $B = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$. Suppose $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$ spans V. Any vector $\mathbf{w} = (w_1, w_2, ..., w_k) \in \mathbb{R}^k$ defines a vector $\mathbf{v} = w_1\mathbf{u}_1 + w_2\mathbf{u}_2 + \cdots + w_k\mathbf{u}_k$ in V, and so can write $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$. Then

$$\mathbf{w} = [\mathbf{v}]_B = c_1[\mathbf{v}_1]_B + c_2[\mathbf{v}_2]_B + \cdots + c_m[\mathbf{v}_m]_B$$

shows that $\{[\mathbf{v}_1]_B, [\mathbf{v}_2]_B, ..., [\mathbf{v}_m]_B\}$ spans $\mathbb{R}^k$.

$\square$

Transition Matrix

Theorem

Let V be a subspace of $\mathbb{R}^n$. Suppose $S = \{\mathbf{u}_1, ..., \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are bases for the subspace V. Let $\mathbf{P}$ be the transition matrix from T to S. Then for any vector $\mathbf{w}$ in V,

$$[\mathbf{w}]_S = \mathbf{P}[\mathbf{w}]_T.$$

Proof.

Let $\mathbf{v}_j = a_{1j}\mathbf{u}_1 + a_{2j}\mathbf{u}_2 + \cdots + a_{kj}\mathbf{u}_k = \sum_{i=1}^{k} a_{ij}\mathbf{u}_i$. Then $[\mathbf{v}_j]_S = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{kj} \end{pmatrix}$. Write $[\mathbf{w}]_S = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_k \end{pmatrix}$ and $[\mathbf{w}]_T = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$.

Then

$$\begin{aligned}
d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_k\mathbf{u}_k = \mathbf{w} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k \\
&= c_1(a_{11}\mathbf{u}_1 + a_{21}\mathbf{u}_2 + \cdots + a_{k1}\mathbf{u}_k) + c_2(a_{12}\mathbf{u}_1 + a_{22}\mathbf{u}_2 + \cdots + a_{k2}\mathbf{u}_k) \\
&\quad + \cdots + c_k(a_{1k}\mathbf{u}_1 + a_{2k}\mathbf{u}_2 + \cdots + a_{kk}\mathbf{u}_k) \\
&= (c_1a_{11} + c_2a_{12} + \cdots + c_ka_{1k})\mathbf{u}_1 + (c_1a_{21} + c_2a_{22} + \cdots + c_ka_{2k})\mathbf{u}_2 \\
&\quad + \cdots + (c_1a_{k1} + c_2a_{k2} + \cdots + c_ka_{kk})\mathbf{u}_k
\end{aligned}$$

Transition Matrix

Continue.

Since S is a basis, the coefficients must be unique. So comparing the coefficients, we get

$$
\begin{aligned}
[\mathbf{w}]_S = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_k \end{pmatrix} &= \begin{pmatrix} c_1 a_{11} + c_2 a_{12} + \cdots + c_k a_{1k} \\ c_1 a_{21} + c_2 a_{22} + \cdots + c_k a_{2k} \\ \vdots \\ c_1 a_{k1} + c_2 a_{k2} + \cdots + c_k a_{kk} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} \\
&= \begin{pmatrix} [\mathbf{v}_1]_S & [\mathbf{v}_2]_S & \cdots & [\mathbf{v}_k]_S \end{pmatrix} = \mathbf{P}[\mathbf{w}]_T
\end{aligned}
$$

$\square$

MA1522: Linear Algebra for Computing

Chapter 4: Subspaces Associated to a Matrix

4.1 Column Space, Row Space, and Nullspace

Column and Row Space

Definition

Let $\mathbf{A}$ be an $m \times n$ matrix,

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

The *row space* of $\mathbf{A}$, is the subspace of $\mathbb{R}^n$ spanned by the rows of $\mathbf{A}$,

$$\mathrm{Row}(\mathbf{A}) = \mathrm{span}\{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix}, \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix}, \dots, \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}\}$$

The *column space* of $\mathbf{A}$, is the subspace of $\mathbb{R}^m$ spanned by the columns of $\mathbf{A}$,

$$\mathrm{Col}(\mathbf{A}) = \mathrm{span}\left\{ \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\}$$

Question

Let $\mathbf{A}$ be an $m \times n$ matrix,

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

The row space is a subspace of

(a) $\mathbb{R}^n$

(b) $\mathbb{R}^m$

The column space is a subspace of

(a) $\mathbb{R}^n$

(b) $\mathbb{R}^m$

Question

Let $\mathbf{A} = \begin{pmatrix} 5 & 2 & -1 & -1 \\ 1 & -1 & 4 & -3 \\ 8 & 3 & -1 & -2 \\ 9 & 3 & 0 & -3 \end{pmatrix}$. Which of the following statements are true?

(i) $\left\{ \begin{pmatrix} 5 \\ 1 \\ 8 \\ 9 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 3 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 4 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -3 \\ -2 \\ -3 \end{pmatrix} \right\}$ is a basis for the column space of $\mathbf{A}$.

(ii) $\left\{ \begin{pmatrix} 5 \\ 2 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 4 \\ -3 \end{pmatrix}, \begin{pmatrix} 8 \\ 3 \\ -1 \\ -2 \end{pmatrix}, \begin{pmatrix} 9 \\ 3 \\ 0 \\ -3 \end{pmatrix} \right\}$ is a basis for the row space of $\mathbf{A}$.

Row Operations Preserves Row Space

Theorem (Row operations preserve row space)

Suppose $\mathbf{A}$ *and* $\mathbf{B}$ *are row equivalent matrices. Then the row space of* $\mathbf{A}$ *is equal to the row space of* $\mathbf{B}$, $Row(\mathbf{A}) = Row(\mathbf{B})$.

Sketch of proof.

(i) Suppose $\mathbf{A} \xrightarrow{R_i \leftrightarrow R_j} \mathbf{B}$. Then the rows of $\mathbf{B}$ are the rows of $\mathbf{A}$ rearranged, which will not change the span.

(ii) Suppose $\mathbf{A} \xrightarrow{cR_i} \mathbf{B}$, $c \neq 0$. Then $\text{Row}(\mathbf{B}) \subseteq \text{Row}(\mathbf{A})$. But $\mathbf{B} \xrightarrow{\frac{1}{c}R_i} \mathbf{A}$ shows that $\text{Row}(\mathbf{A}) \subseteq \text{Row}(\mathbf{B})$ too.

(iii) Suppose $\mathbf{A} \xrightarrow{R_i + aR_j} \mathbf{B}$. Then $\text{Row}(\mathbf{B}) \subseteq \text{Row}(\mathbf{A})$. But $\mathbf{B} \xrightarrow{R_i - aR_j} \mathbf{A}$ shows that $\text{Row}(\mathbf{A}) \subseteq \text{Row}(\mathbf{B})$ too.

$\square$

Finding Basis for Row Space

Theorem

If a matrix $\mathbf{R}$ *is in reduced row-echelon form, then the nonzero rows of* $\mathbf{R}$ *form a basis for its row space.*

Sketch of proof.

Write $\mathbf{R} = \begin{pmatrix} 0 & \cdots & 1 & \cdots & * & 0 & \cdots & * & 0 & \cdots \\ 0 & \cdots & 0 & \cdots & 0 & 1 & \cdots & * & 0 & \cdots \\ 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots \\ 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots \\ \vdots & & & & & \vdots & & & \vdots & \vdots \end{pmatrix}$. By definition, the 1 in the leading entry in each nonzero row is only nonzero entry in that coordinate among the rows. This shows that each nonzero row cannot be a linear combination of the other rows. Hence, the rows of $\mathbf{R}$ are linearly independent. It is clear that the nonzero rows spans the row space of $\mathbf{R}$. □

Theorem

For any matrix $\mathbf{A}$*, the nonzero rows of the reduced row-echelon form of* $\mathbf{A}$ *form a basis for for the row space of* $\mathbf{A}$*.*

Proof.

Follows from the fact that $\mathbf{A}$ is row equivalent to its reduced row-echelon form $\mathbf{R}$, and that the nonzero rows of $\mathbf{R}$ form a basis for Row($\mathbf{R}$). □

Challenge

3. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 & -1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

   So $\left\{ \begin{pmatrix} 1 & 0 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix} \right\}$ is a basis for Row($\mathbf{A}$).

   However, in this case, we could have taken the original rows of $\mathbf{A}$

   $$\left\{ \begin{pmatrix} 1 & 1 & 2 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 & 1 \end{pmatrix} \right\}$$

   as a basis for Row($\mathbf{A}$) too. Why?

Discussion

Consider the matrix $\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 & -1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$. Claim: the 4th column is a linear combination of the first 3 columns. (It should be clear that the columns are linearly dependent since there are 4 columns and these are vectors in $\mathbb{R}^3$.)

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 & -1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \xrightarrow{RREF} \mathbf{R} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

We might add a bar between the 3rd and 4th columns to emphasize that we are solving a linear system, but that would not affect the computation/reduction. From the reduced row-echelon form, we conclude that

$$\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 0\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

But this is exactly the linear relations between the columns of $\mathbf{R}$,

$$\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = 0\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Discussion

In fact, the linear relations between **A** and its reduced row-echelon form are preserved for any columns. Consider

$$\mathbf{A} = \begin{array}{c} \begin{array}{ccccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{a}_5 \end{array} \\ \begin{pmatrix} 2 & 8 & 2 & 4 & 2 \\ 1 & 6 & 0 & 4 & 3 \\ -1 & -4 & -1 & -2 & -1 \end{pmatrix} \end{array} \xrightarrow{RREF} \begin{array}{c} \begin{array}{ccccc} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{r}_4 & \mathbf{r}_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 3 & -2 & -3 \\ 0 & 1 & -1/2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} = \mathbf{R}.$$

Then

$$\begin{array}{rclcrcl} \mathbf{r}_3 = & 3\mathbf{r}_1 - \frac{1}{2}\mathbf{r}_2 & \longleftrightarrow & \mathbf{a}_3 = & 3\mathbf{a}_1 - \frac{1}{2}\mathbf{a}_2 \\ \mathbf{r}_4 = & -2\mathbf{r}_1 + \mathbf{r}_2 & \longleftrightarrow & \mathbf{a}_4 = & -2\mathbf{a}_1 + \mathbf{a}_2 \\ \mathbf{r}_5 = & -3\mathbf{r}_1 + \mathbf{r}_2 & \longleftrightarrow & \mathbf{a}_5 = & -3\mathbf{a}_1 + \mathbf{a}_2 \end{array}$$

Also, $\{\mathbf{r}_1, \mathbf{r}_2\}$ is linearly independent, and so is $\{\mathbf{a}_1, \mathbf{a}_2\}$.

Row Operations Preserves Linear Relations Between Columns

Theorem (Row operations preserve linear relations between columns)

Let $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{pmatrix}$ *and* $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}$ *be row equivalent* $m \times n$ *matrices, where* $\mathbf{a}_i$ *and* $\mathbf{b}_i$ *is the* i*-th column of* $\mathbf{A}$ *and* $\mathbf{B}$*, respectively, for* $i = 1, ..., n$*. Then for any coefficients* $c_1, c_2, ..., c_n$*,*

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n = \mathbf{0}$$

if and only if

$$c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n = \mathbf{0}.$$

Sketch of the proof.

Since $\mathbf{A}$ and $\mathbf{B}$ are row equivalent, $\mathbf{A} = \mathbf{PB}$ for some invertible order m matrix $\mathbf{P}$. By block multiplication,

$$\begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{pmatrix} = \mathbf{A} = \mathbf{PB} = \begin{pmatrix} \mathbf{Pb}_1 & \mathbf{Pb}_2 & \cdots & \mathbf{Pb}_n \end{pmatrix} \quad \Rightarrow \quad \mathbf{a}_i = \mathbf{Pb}_i, \; i = 1, ..., n.$$

So, if $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n = \mathbf{0}$,

$$\mathbf{0} = \mathbf{P0} = \mathbf{P}(c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n) = c_1\mathbf{Pb}_1 + c_2\mathbf{Pb}_2 + \cdots + c_n\mathbf{Pb}_n = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n.$$

Use $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}$ to prove the converse. $\square$

Finding basis for Column space

Theorem

Suppose $\mathbf{R}$ *is the reduced row-echelon form of a matrix* $\mathbf{A}$*. Then the columns of* $\mathbf{A}$ *corresponding to the pivot columns in* $\mathbf{R}$ *form a basis for the column space of* $\mathbf{A}$*.*

Proof.

First observe that the pivot columns in $\mathbf{R}$ are linearly independent since they are just the vectors in the standard basis. Also, the non-pivot columns of $\mathbf{R}$ linearly dependent on the pivot columns. Hence, the columns of $\mathbf{A}$ that corresponds to the pivot columns of $\mathbf{R}$ are linearly independent and are sufficient to span $\mathrm{Col}(\mathbf{A})$, and thus form a basis. $\square$

Challenge

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 4 & 1 & 2 \\ 4 & 2 & 2 & 3 & 2 \\ 2 & 1 & -2 & 2 & 0 \\ 6 & 3 & 6 & 4 & 4 \end{pmatrix} \xrightarrow{RREF} \mathbf{R} = \begin{pmatrix} 1 & 1/2 & 0 & 5/6 & 1/3 \\ 0 & 0 & 1 & -1/6 & 1/3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1/2 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 5/6 \\ -1/6 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{6}\left(5\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}\right), \quad \begin{pmatrix} 1/3 \\ 1/3 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{3}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}\right),$$

$$\begin{pmatrix} 1 \\ 2 \\ 1 \\ 3 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 \\ 4 \\ 2 \\ 6 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 3 \\ 2 \\ 4 \end{pmatrix} = \frac{1}{6}\left(5\begin{pmatrix} 2 \\ 4 \\ 2 \\ 6 \end{pmatrix} - \begin{pmatrix} 4 \\ 2 \\ -2 \\ 6 \end{pmatrix}\right), \quad \begin{pmatrix} 2 \\ 2 \\ 0 \\ 4 \end{pmatrix} = \frac{1}{3}\left(\begin{pmatrix} 2 \\ 4 \\ 2 \\ 6 \end{pmatrix} + \begin{pmatrix} 4 \\ 2 \\ -2 \\ 6 \end{pmatrix}\right).$$

Since the first and third columns of **R** are the pivot columns, the first an third columns of **A** form a basis for Col(**A**). However, in this case, we could take any 2 columns of **A** except columns 1 and 2, to be a basis for the column space of **A**. Why?

Remarks

1. Row operations do not preserve column space. Consider

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \mathbf{B} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}.$$

$$\mathrm{Col}(\mathbf{A}) = \mathrm{span}\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \neq \mathrm{span}\left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

2. Row operations do no preserve linear relations between the rows. Consider

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \xrightarrow{R_2 - 2R_1} \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

$$\text{row 2 of } \mathbf{A} = 2 \times \text{ row 1 of } \mathbf{A}, \quad \text{row 2 of } \mathbf{B} = 0 \times \text{ row 1 of } \mathbf{B}$$

Column Space and Consistency of Linear System

Let $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$.

- Then a vector $\mathbf{v}$ is in the column space of $\mathbf{A}$, $\mathbf{v} \in \text{Col}(\mathbf{A})$, if and only if we can find a $\mathbf{u} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ such that

$$\mathbf{Au} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{v}.$$

- This is equivalent to the system $\mathbf{Ax} = \mathbf{v}$ being consistent. This is also equivalent to $\mathbf{v} = \mathbf{Au}$ for some $\mathbf{u}$ in $\mathbb{R}^k$.
- Hence, the column space can be characterized either by the set of vectors $\mathbf{v}$ such that $\mathbf{Ax} = \mathbf{v}$ is consistent, or the set of vectors $\mathbf{v}$ such that $\mathbf{v} = \mathbf{Au}$ for some $\mathbf{u}$,

$$\text{Col}(\mathbf{A}) = \left\{ \ \mathbf{v} = \mathbf{Au} \ \middle| \ \mathbf{u} \in \mathbb{R}^k \ \right\} = \left\{ \ \mathbf{v} \ \middle| \ \mathbf{Ax} = \mathbf{v} \text{ is consistent } \right\}.$$

Nullspace

Definition

The *nullspace* of a $m \times n$ matrix $\mathbf{A}$ is the solution space to the homogeneous system $\mathbf{Ax} = \mathbf{0}$ with coefficient matrix $\mathbf{A}$. It is denoted as

$$\mathrm{Null}(\mathbf{A}) = \{\ \mathbf{v} \in \mathbb{R}^n \mid \mathbf{Av} = \mathbf{0}\ \}.$$

The *nullity* of $\mathbf{A}$ is the dimension of the nullspace of $\mathbf{A}$, denoted as

$$\mathrm{nullity}(\mathbf{A}) = \dim(\mathrm{Null}(\mathbf{A})).$$

4.2 Rank

Rank

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{R}$ its reduced row-echelon form.

$$\begin{aligned}
\dim(\mathrm{Col}(\mathbf{A})) &= \#\text{ of pivot columns in RREF of } \mathbf{A}, \\
&= \#\text{ of leading entries in RREF of } \mathbf{A}, \\
&= \#\text{ of nonzero rows in RREF of } \mathbf{A} = \dim(\mathrm{Row}(\mathbf{A}))
\end{aligned}$$

Definition

Define the *rank* of $\mathbf{A}$ to be the dimension of its column or row space

$$\mathrm{rank}(\mathbf{A}) = \dim(\mathrm{Col}(\mathbf{A})) = \dim(\mathrm{Row}(\mathbf{A})).$$

Exercise

Prove that the rank is invariant under transpose,

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^T).$$

Challenge: Rank and Consistency of Linear Systems

Prove the following theorem.

Theorem

The linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ *is consistent if and only if the rank of* $\mathbf{A}$ *is equal to the rank of the augmented matrix* $(\mathbf{A} \mid \mathbf{b})$,

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}((\mathbf{A} \mid \mathbf{b})).$$

Properties Rank

Let $\mathbf{A} = \begin{pmatrix} 4 & 3 & 5 \\ 5 & -1 & 5 \\ -1 & 0 & 0 \\ 5 & 2 & 5 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 3 & -1 \\ 1 & 1 \\ 3 & 3 \end{pmatrix}$. Then $\text{rank}(\mathbf{A}) = 3$ and $\text{rank}(\mathbf{B}) = 2$. Now,

$$\mathbf{AB} = \begin{pmatrix} 30 & 14 \\ 29 & 9 \\ -3 & 1 \\ 32 & 12 \end{pmatrix},$$

and $\text{rank}(\mathbf{AB}) = 2$.

Here $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$ and $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{B})$.

Properties of Rank

Let $\mathbf{A} = \begin{pmatrix} 1 & -1 & 3 \\ -1 & 0 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} -1 & 0 & -1 \\ 0 & 1 & 2 \\ 3 & 0 & 3 \end{pmatrix}$. Check that $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B}) = 2$. Next,

$$\mathbf{AB} = \begin{pmatrix} 8 & -1 & 6 \\ 1 & 0 & 1 \end{pmatrix}$$

and $\mathrm{rank}(\mathbf{AB}) \leq \mathrm{rank}(\mathbf{A})$ and $\mathrm{rank}(\mathbf{AB}) \leq \mathrm{rank}(\mathbf{B})$.

Properties of Rank

Let $\mathbf{A} = \begin{pmatrix} 1 & 0 & 3 & -1 \\ 3 & -2 & 1 & -1 \\ 3 & -2 & 1 & -1 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} -2 & -3 & -1 \\ -3 & -4 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$. Check that $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B}) = 2$. Now

$$\mathbf{AB} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \Rightarrow \quad \text{rank}(\mathbf{AB}) = 0.$$

Here $\text{rank}(\mathbf{AB}) < \text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})$.

Properties of Rank

Lemma

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{B}$ a $n \times p$ matrix. The column space of the product $\mathbf{AB}$ is a subspace of the column space of $\mathbf{A}$,

$$Col(\mathbf{AB}) \subseteq Col(\mathbf{A}).$$

Sketch of proof.

Write $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots \mathbf{b}_p \end{pmatrix}$. Then

$$\mathbf{AB} = \mathbf{A} \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots \mathbf{b}_p \end{pmatrix} = \begin{pmatrix} \mathbf{Ab}_1 & \mathbf{Ab}_2 & \cdots \mathbf{Ab}_p \end{pmatrix}.$$

Recall that $\mathbf{Au} \in \mathrm{Col}(\mathbf{A})$ for all $\mathbf{u}$, and hence, $\mathbf{Ab}_i \in \mathrm{Col}(\mathbf{A})$ for all $i = 1, .., p$. Therefore

$$\mathrm{Col}(\mathbf{AB}) = \mathrm{span}\{\mathbf{Ab}_1, \mathbf{Ab}_2, ..., \mathbf{Ab}_p\} \subseteq \mathrm{Col}(\mathbf{A}).$$

$\square$

Properties of Rank

Theorem

Let $\mathbf{A}$ *be a* $m \times n$ *matrix and* $\mathbf{B}$ *a* $n \times p$ *matrix. Then*

$$\mathrm{rank}(\mathbf{AB}) \leq \min\{\mathrm{rank}(\mathbf{A}), \mathrm{rank}(\mathbf{B})\}.$$

Proof.

By the previous lemma,

$$\mathrm{rank}(\mathbf{AB}) = \dim(\mathrm{Col}(\mathbf{AB})) \leq \dim(\mathrm{Col}(\mathbf{A})) = \mathrm{rank}(\mathbf{A}).$$

Next, using the previous lemma and the above derivation on $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$, we have

$$\mathrm{rank}(\mathbf{AB}) = \mathrm{rank}((\mathbf{AB})^T) = \mathrm{rank}(\mathbf{B}^T\mathbf{A}^T) \leq \mathrm{rank}(\mathbf{B}^T) = \mathrm{rank}(\mathbf{B}).$$

Hence,

$$\mathrm{rank}(\mathbf{AB}) \leq \min\{\mathrm{rank}(\mathbf{A}), \mathrm{rank}(\mathbf{B})\}.$$

□

Rank-Nullity Theorem

Theorem (Rank-Nullity Theorem)

Let $\mathbf{A}$ *be a* $m \times n$ *matrix. The sum of its rank and nullity is equal to the number of columns,*

$$\operatorname{rank}(\mathbf{A}) + \operatorname{nullity}(\mathbf{A}) = n.$$

Sketch of Proof.

This follows from the fact that the nullity of $\mathbf{A}$ is equal to the number of non-pivot columns in its reduced row-echelon form, and that the rank of $\mathbf{A}$ is equal to the number of pivot columns of its reduced row-echelon form. $\square$

Summary of the Subspaces Associated to a Matrix

Let **A** be a $m \times n$ matrix.

Subspace	**Subspace of**	**Basis**	**Dimension**
Col($\mathbf{A}$)	$\mathbb{R}^m$	Columns of **A** corresponding to pivot columns in RREF	rank($\mathbf{A}$) $=$ no. of pivot columns in RREF
Row($\mathbf{A}$)	$\mathbb{R}^n$	Nonzero rows of RREF	rank($\mathbf{A}$) $=$ no. of nonzero rows in RREF
Null($\mathbf{A}$)	$\mathbb{R}^n$	Vectors in general solution to $\mathbf{Ax} = \mathbf{0}$	nullity($\mathbf{A}$) $=$ no. of nonpivot columns in RREF

Challenge

Let $\mathbf{A}$ and $\mathbf{B}$ be matrices of the same size. Prove that

$$\operatorname{rank}(\mathbf{A}+\mathbf{B}) \leq \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B}).$$

Full rank

Let $\mathbf{A}$ be a $m \times n$ matrix.

$$
\begin{aligned}
\text{rank}(\mathbf{A}) &= \#\text{ of pivot columns in RREF} \leq \text{ no. of columns} = n \\
\text{rank}(\mathbf{A}) &= \#\text{ of nonzero rows in RREF} \leq \text{ no. of rows} = m
\end{aligned}
$$

So, the rank of $\mathbf{A}$ is no greater than the number of rows or columns, whichever is smaller,

$$\text{rank}(\mathbf{A}) \leq \min\{m, n\}.$$

The maximum rank a matrix can attain is when it is equal to either the number of rows or columns, whichever is smaller.

Definition

A $m \times n$ matrix $\mathbf{A}$ is said to be of *full rank* if its rank is equal to either the number of rows or columns,

$$\text{rank}(\mathbf{A}) = \min\{m, n\}.$$

Discussion

What happens when a $m \times n$ matrix $\mathbf{A}$ is full rank? Let us consider cases.

Consider the case when $\mathbf{A}$ is a square matrix $m = n$. Then $\mathbf{A}$ is full rank if and only if its is invertible. The proof is left as an exercise. We will include this in the list of equivalent statements for invertibility.

Equivalent Statements for Invertibility

Theorem

Let $\mathbf{A}$ be a square matrix of order n. The following statements are equivalent.

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}^T$ *is invertible.*

(iii) *(left inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{BA} = \mathbf{I}$.

(iv) *(right inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{AB} = \mathbf{I}$.

(v) *The reduced row-echelon form of* $\mathbf{A}$ *is the identity matrix.*

(vi) $\mathbf{A}$ *can be expressed as a product of elementary matrices.*

(vii) *The homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution.*

(viii) *For any* $\mathbf{b}$*, the system* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution.*

(ix) *The determinant of* $\mathbf{A}$ *is nonzero,* $\det(\mathbf{A}) \neq 0$.

(x) *The columns/rows of* $\mathbf{A}$ *are linearly independent.*

(xi) *The columns/rows of* $\mathbf{A}$ *spans* $\mathbb{R}^n$.

(xii) $\text{rank}(\mathbf{A}) = n$ *(*$\mathbf{A}$ *has full rank).*

(xiii) $\text{nullity}(\mathbf{A}) = 0$.

Discussion

Let $\mathbf{A}$ be a $m \times n$ matrix. Suppose $\mathbf{A}$ is not a square matrix $m \neq n$, and $\mathbf{A}$ is full rank.
Then either $\text{rank}(\mathbf{A}) = n < m$, $\text{rank}(\mathbf{A}) = m < n$. In either cases, some of the equivalent statements of invertibility will still be true of $\mathbf{A}$.

Lemma

Let $\mathbf{A}$ be a $m \times n$ matrix. Then the nullspace of $\mathbf{A}$ is equal to the nullspace of $\mathbf{A}^T\mathbf{A}$,

$$\text{Null}(\mathbf{A}) = \text{Null}(\mathbf{A}^T\mathbf{A}).$$

Proof.

Suppose $\mathbf{u}$ is in the nullspace of $\mathbf{A}$, $\mathbf{Au} = \mathbf{0}$. Then premultiplying by $\mathbf{A}^T$, $\mathbf{A}^T\mathbf{Au} = \mathbf{0}$ too. This shows that $\mathbf{u}$ is in the nullspace of $\mathbf{A}^T\mathbf{A}$. This proves $\text{Null}(\mathbf{A}) \subseteq \text{Null}(\mathbf{A}^T\mathbf{A})$.

Conversely, suppose $\mathbf{u}$ is in the nullspace of $\mathbf{A}^T\mathbf{A}$, $\mathbf{A}^T\mathbf{Au} = \mathbf{0}$. Premultiplying both sides by $\mathbf{u}^T$, and noting that $\mathbf{u}^T\mathbf{A}^T\mathbf{Au} = (\mathbf{Au}) \cdot (\mathbf{Au})$,

$$(\mathbf{Au}) \cdot (\mathbf{Au}) = \mathbf{u}^T\mathbf{A}^T\mathbf{Au} = \mathbf{u}^T(\mathbf{0}) = 0.$$

Hence, $\mathbf{Au} = \mathbf{0}$ too, that is, $\mathbf{u}$ is in the nullspace of $\mathbf{A}$. This proves that $\text{Null}(\mathbf{A}^T\mathbf{A}) \subseteq \text{Null}(\mathbf{A})$ too. $\square$

Full Rank Equals Number of Columns

Theorem

Suppose $\mathbf{A}$ *is a* $m \times n$ *matrix. The following statements are equivalent.*

(i) $\mathbf{A}$ *is full rank, where the rank is equal to the number of columns,* $\text{rank}(\mathbf{A}) = n$.

(ii) *The rows of* $\mathbf{A}$ *spans* $\mathbb{R}^n$, $Row(\mathbf{A}) = \mathbb{R}^n$.

(iii) *The columns of* $\mathbf{A}$ *are linearly independent.*

(iv) *The homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution, that is,* $\text{Null}(\mathbf{A}) = \{\mathbf{0}\}$.

(v) $\mathbf{A}^T\mathbf{A}$ *is an invertible matrix of order* n.

(vi) $\mathbf{A}$ *has a left inverse.*

Proof.

We will only prove the equivalence of the last 3 statements, the rest are left as an exercise. Hint: One might try to prove (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv). Observe that if $\text{rank}(\mathbf{A}) = n$, then the reduced row-echelon form is of the form

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{0}_{(m-n)\times n} \end{pmatrix}.$$

Full Rank Equals Number of Columns

Continue of Proof.
(iv) $\Rightarrow$ (v): Suppose the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution. By the lemma, the homogeneous system $\mathbf{A}^T\mathbf{Ax} = \mathbf{0}$ has only the trivial solution too. But since $\mathbf{A}^T\mathbf{A}$ is a square matrix, by the equivalent statements of invertibility, $\mathbf{A}^T\mathbf{A}$ is invertible.

(v) $\Rightarrow$ (vi): Suppose $\mathbf{A}^T\mathbf{A}$ is invertible. Then

$$\mathbf{I} = (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{A}) = ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{A},$$

which shows that $((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)$ is a left inverse of $\mathbf{A}$.

(vi) $\Rightarrow$ (i): Let $\mathbf{B}$ be a left inverse of $\mathbf{A}$, $\mathbf{BA} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Then by the properties of rank,

$$n = \text{rank}(\mathbf{I}) = \text{rank}(\mathbf{BA}) \leq \text{rank}(\mathbf{A}).$$

But since $\text{rank}(\mathbf{A}) \leq n$, equality holds. $\square$

Full Rank Equals Number of Rows

Theorem

Suppose $\mathbf{A}$ *is a* $m \times n$ *matrix. The following statements are equivalent.*

(i) $\mathbf{A}$ *is full rank, where the rank is equal to the number of rows,* $\text{rank}(\mathbf{A}) = m$.

(ii) *The columns of* $\mathbf{A}$ *spans* $\mathbb{R}^m$, $Col(\mathbf{A}) = \mathbb{R}^m$.

(iii) *The rows of* $\mathbf{A}$ *are linearly independent.*

(iv) *The linear system* $\mathbf{Ax} = \mathbf{b}$ *is consistent for every* $\mathbf{b} \in \mathbb{R}^m$.

(v) $\mathbf{AA}^T$ *is an invertible matrix of order* m.

(vi) $\mathbf{A}$ *has a right inverse.*

Full Rank Equals Number of Rows

If $\mathrm{rank}(\mathbf{A}) = m$, then the reduced row-echelon form is of the form

$$\begin{pmatrix} 0 & \cdots & 1 & \ldots & 0 & \ldots & 0 & \ldots & 0 & \ldots \\ 0 & \cdots & 0 & \ldots & 1 & \ldots & 0 & \ldots & 0 & \ldots \\ 0 & \cdots & 0 & \ldots & 0 & \ldots & 1 & \ldots & 0 & \ldots \\ \vdots & & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \cdots \end{pmatrix}.$$

The proof follows from the previous theorem by replacing $\mathbf{A}$ with $\mathbf{A}^T$. For statement (iv), use rank-nullity theorem. The details are left to the readers.

Challenge

Let $\mathbf{A}$ be a $m \times n$ matrix such that $\text{rank}(\mathbf{A}) = m$. Suppose $m > n$. By the equivalent statements of full rank equals number of columns, $(\mathbf{A}^T\mathbf{A})$ invertible and $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. is a left inverse of $\mathbf{A}$.

Now consider the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ for some vector $\mathbf{b}$ in $\mathbb{R}^m$. Premultiplying the left inverse above on both sides of the equation, we get

$$\mathbf{x} = ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{A}\mathbf{x} =((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{b},$$

that is, $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{b}$ is a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$. But this is true for every $\mathbf{b}$, which by the equivalent statements of full rank equals number or rows, means that the rank of $\mathbf{A}$ is equal to m, the number of row. This is a contradiction to $m > n$.

What is the mistake in the argument above?

MA1522 Linear Algebra for Computing

Chapter 5: Orthogonality, Projection, and Least Square Solution

5.1 Orthogonality

Discussion

Let $\mathbf{u} = (u_i)$ and $\mathbf{v} = (v_i)$ be vectors in $\mathbb{R}^n$. Recall that

- The inner product of vectors $\mathbf{u} = (u_i)$ and $\mathbf{v} = (v_i)$ is defined to be

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

- The norm of a vector $\mathbf{u}$ is

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

- The angle between nonzero vectors $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ is

$$\theta = \cos^{-1}\left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}\right), \quad 0 \leq \theta \leq \pi.$$

So suppose $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ and $\mathbf{u} \cdot \mathbf{v} = 0$, what is the angle between $\mathbf{u}$ and $\mathbf{v}$?

Orthogonal

Definition

Two vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ are orthogonal if

$$\mathbf{u} \cdot \mathbf{v} = 0.$$

Suppose $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal.

- Case 1: Either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$.
- Case 2: Otherwise,

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = 0$$

tells us that $\theta = \frac{\pi}{2}$, that is, $\mathbf{u}$ and $\mathbf{v}$ are perpendicular.

That is, $\mathbf{u}, \mathbf{v}$ are orthogonal if and only if either one of them is the zero vector or they are perpendicular to each other.

Question

Suppose $\mathbf{u}$ and $\mathbf{v}$ are orthogonal. Is it true that for any real numbers $s, t \in \mathbb{R}$, $s\mathbf{u}$ and $t\mathbf{v}$ are orthogonal.

Orthogonal and Orthonormal Sets

Definition

A set $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ of vectors in $\mathbb{R}^n$ of vectors is *orthogonal* if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for every $i \neq j$, that is, vectors in S are pairwise orthogonal.

The set $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ is *orthonormal* if for all $i, j = 1, ..., k$,

$$\mathbf{v}_i \cdot \mathbf{v}_j = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

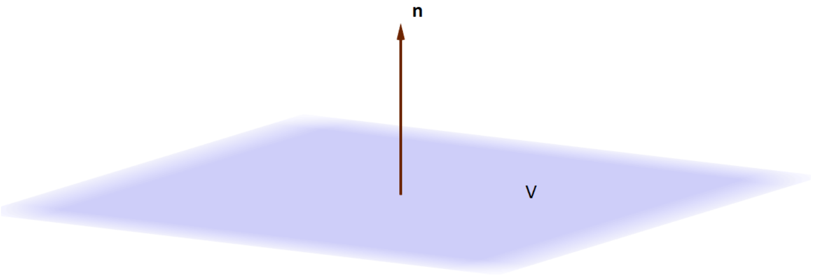
That is, S is orthogonal, and all the vectors are unit vectors.

Orthogonal to a Subspace

Definition

Let V be a subspace of $\mathbb{R}^n$. A vector $\mathbf{n}$ is *orthogonal* to V if for every $\mathbf{v}$ in V, $\mathbf{n} \cdot \mathbf{v} = 0$, that is, $\mathbf{n}$ is orthogonal to every vector in V. We will denote it as $\mathbf{n} \perp V$.

$$\mathbf{n} \perp V \quad \Leftrightarrow \quad \mathbf{n} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{v} \in V$$



If $\mathbf{n} \neq \mathbf{0}$.

Remark

The zero vector $\mathbf{0}$ is orthogonal to every subspace, $\mathbf{0} \perp V$.

Orthogonal to a Subspace

Theorem

Let V be a subspace of $\mathbb{R}^n$ be a subspace and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a spanning set for V, span$(S) = V$. Then a vector $\mathbf{w}$ is orthogonal to V if and only if $\mathbf{w} \cdot \mathbf{u}_i = 0$ for all $i = 1, ..., k$.

Proof.

($\Rightarrow$) Suppose $\mathbf{w}$ is orthogonal to V, $\mathbf{w} \cdot \mathbf{v} = 0$ for all $\mathbf{v} \in V$. Then since $\mathbf{u}_i \in V$ for all i, then $\mathbf{w} \cdot \mathbf{u}_i = 0$ for all i.

($\Leftarrow$) Suppose $\mathbf{w} \cdot \mathbf{u}_i = 0$ for all $i = 1, ..., k$. Now given any $\mathbf{v} \in V$, since S spans V, we can write

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$$

for some $c_1, c_2, ..., c_k \in \mathbb{R}$. Then

$$\begin{aligned} \mathbf{w} \cdot \mathbf{v} &= \mathbf{w} \cdot (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) = c_1\mathbf{w} \cdot \mathbf{u}_1 + c_2\mathbf{w} \cdot \mathbf{u}_2 + \cdots + c_k\mathbf{w} \cdot \mathbf{u}_k \\ &= c_1(0) + c_2(0) + \cdots + c_k(0) = 0 \end{aligned}$$

which proves that $\mathbf{w} \cdot \mathbf{v} = 0$ for all $\mathbf{v} \in V$.

$\square$

That is, to check that $\mathbf{w}$ is orthogonal to V, suffice to check that it is orthogonal to every vector in a spanning set.

Algorithm to check for Orthogonal to a Subspace

Theorem

Let V be a subspace of $\mathbb{R}^n$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a spanning set for V. Then $\mathbf{w}$ is orthogonal to V if and only if $\mathbf{w}$ is in the nullspace of $\mathbf{A}^T$, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$;

$$\mathbf{w} \perp V \quad \Leftrightarrow \quad \mathbf{w} \in \mathrm{Null}(\mathbf{A}^T)$$

Sketch of Proof.

By previous theorem, $\mathbf{w} \perp V$ if and only if $\mathbf{u}_i^T\mathbf{w} = \mathbf{u}_i \cdot \mathbf{w} = 0$ for all $i = 1, 2, ..., k$. By block multiplication, this is equivalent to

$$\mathbf{A}^T\mathbf{w} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}^T \mathbf{w} = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_k^T \end{pmatrix} \mathbf{w} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{w} \\ \mathbf{u}_2^T\mathbf{w} \\ \vdots \\ \mathbf{u}_k^T\mathbf{w} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$\square$

Orthogonal Complement

Observe that in all the examples above, the set of vectors that are orthogonal to a subspace V is a subspace. If fact, it is the nullspace of the matrix whose rows are vectors in a spanning set of V.

Definition

Let V be a subspace of $\mathbb{R}^n$. The *orthogonal complement* of V is the set of all vectors that are orthogonal to V, and is denoted as

$$V^{\perp} = \{\ \mathbf{w} \in \mathbb{R}^n \ \big|\ \mathbf{w} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{v} \text{ in } V\ \}.$$

Theorem

Let V be a subspace of $\mathbb{R}^n$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a spanning set for V. Let $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$. Then the orthogonal complement of V is the nullspace of $\mathbf{A}^T$,

$$V^{\perp} = \text{Null}(\mathbf{A}^T).$$

Challenge

Let $\mathbf{A}$ be a $m \times n$ matrix. Show that the nullspace of $\mathbf{A}$ is the orthogonal complement of the row space of $\mathbf{A}$,

$$\mathrm{Row}(\mathbf{A})^{\perp} = \mathrm{Null}(\mathbf{A}).$$

5.2 Orthogonal and Orthonormal Bases

Orthogonal and Orthonormal Basis

Theorem

Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is an orthogonal set of nonzero vectors. Then S is linearly independent.

Proof.

Let $c_1, c_2, ..., c_k$ be coefficients such that $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$. For each $i = 1, ..., k$,

$$\begin{aligned}
\mathbf{0} &= \mathbf{u}_i \cdot \mathbf{0} = \mathbf{u}_i \cdot (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) = c_1\mathbf{u}_i \cdot \mathbf{u}_1 + \cdots + c_i\mathbf{u}_i \cdot \mathbf{u}_i + \cdots + c_k\mathbf{u}_i \cdot \mathbf{u}_k \\
&= c_1(0) + \cdots + c_i\|\mathbf{u}_i\|^2 + \cdots + c_k(0) \\
&= c_i\|\mathbf{u}_i\|^2,
\end{aligned}$$

where the 4th equality follows from the fact that S is orthogonal. Since $\mathbf{u}_i \neq \mathbf{0}$ is nonzero, this means that necessary $c_i = 0$ for all $i = 1, ..., n$. Hence, S is linearly independent. $\square$

Corollary

Every orthonormal set is linearly independent.

Orthogonal and Orthonormal basis

Definition

Let V be a subspace of $\mathbb{R}^n$. A set S is an *orthogonal basis* (resp, *orthonormal basis*) for V if S is a basis of V and S is an orthogonal (resp, orthonormal) set.

Coordinates Relative to an Orthogonal Basis

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be an orthogonal basis for a subspace V, and let $\mathbf{v} \in V$. Express $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$ as a linear combination of vectors in the basis S.

Then since S is orthogonal,

$$\begin{aligned} \mathbf{u}_i \cdot \mathbf{v} &= \mathbf{u}_i \cdot (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) = c_1\mathbf{u}_i \cdot \mathbf{u}_1 + \cdots + c_i\mathbf{u}_i \cdot \mathbf{u}_i + \cdots + c_k\mathbf{u}_i \cdot \mathbf{u}_k \\ &= c_1(0) + \cdots + c_i\|\mathbf{u}_i\|^2 + \cdots + c_k(0) \\ &= c_i\|\mathbf{u}_i\|^2. \end{aligned}$$

This means that $c_i = \frac{\mathbf{u}_i \cdot \mathbf{v}}{\|\mathbf{u}_i\|^2}$, for all $i = 1, ..., k$. Equivalently, $[\mathbf{v}]_S = \begin{pmatrix} \mathbf{u}_1 \cdot \mathbf{v}/\|\mathbf{u}_1\|^2 \\ \mathbf{u}_2 \cdot \mathbf{v}/\|\mathbf{u}_2\|^2 \\ \vdots \\ \mathbf{u}_k \cdot \mathbf{v}/\|\mathbf{u}_k\|^2 \end{pmatrix}$

Coordinates Relative to an Orthogonal Basis

Theorem

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be an orthogonal basis for a subspace V of $\mathbb{R}^n$. Then for any $\mathbf{v} \in V$,

$$\mathbf{v} = \left(\frac{\mathbf{v}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2}\right)\mathbf{u}_1 + \left(\frac{\mathbf{v}\cdot\mathbf{u}_2}{\|\mathbf{u}_2\|^2}\right)\mathbf{u}_2 + \cdots + \left(\frac{\mathbf{v}\cdot\mathbf{u}_k}{\|\mathbf{u}_k\|^2}\right)\mathbf{u}_k$$

If further S is an orthonormal basis, then

$$\mathbf{v} = (\mathbf{v}\cdot\mathbf{u}_1)\,\mathbf{u}_1 + (\mathbf{v}\cdot\mathbf{u}_2)\,\mathbf{u}_2 + \cdots + (\mathbf{v}\cdot\mathbf{u}_k)\,\mathbf{u}_k.$$

$$\text{that is } S \text{ orthogonal, } [\mathbf{v}]_S = \begin{pmatrix} \frac{\mathbf{v}\cdot\mathbf{u}_1}{\|\mathbf{u}_1\|^2} \\ \frac{\mathbf{v}\cdot\mathbf{u}_2}{\|\mathbf{u}_2\|^2} \\ \vdots \\ \frac{\mathbf{v}\cdot\mathbf{u}_k}{\|\mathbf{u}_k\|^2} \end{pmatrix}, \; S \text{ orthonormal, } [\mathbf{v}]_S = \begin{pmatrix} \mathbf{v}\cdot\mathbf{u}_1 \\ \mathbf{v}\cdot\mathbf{u}_2 \\ \vdots \\ \mathbf{v}\cdot\mathbf{u}_k \end{pmatrix}.$$

Challenge

Let V be a subspace of $\mathbb{R}^n$ and S an orthonormal basis of V. Show that for any $\mathbf{u}, \mathbf{v} \in V$,

1. $\mathbf{u} \cdot \mathbf{v} = [\mathbf{u}]_S \cdot [\mathbf{v}]_S$.

2. $\|\mathbf{u} - \mathbf{v}\| = \|[\mathbf{u}]_S - [\mathbf{v}]_S\|$.

Discussion

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$. Construct the $n \times k$ matrix $\mathbf{Q} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$ whose columns are the vectors in S. Consider the product

$$\mathbf{Q}^T\mathbf{Q} = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_k^T \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \cdots & \mathbf{u}_1^T\mathbf{u}_k \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \cdots & \mathbf{u}_2^T\mathbf{u}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_k^T\mathbf{u}_1 & \mathbf{u}_k^T\mathbf{u}_2 & \cdots & \mathbf{u}_k^T\mathbf{u}_k \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1\cdot\mathbf{u}_1 & \mathbf{u}_1\cdot\mathbf{u}_2 & \cdots & \mathbf{u}_1\cdot\mathbf{u}_k \\ \mathbf{u}_2\cdot\mathbf{u}_1 & \mathbf{u}_2\cdot\mathbf{u}_2 & \cdots & \mathbf{u}_2\cdot\mathbf{u}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_k\cdot\mathbf{u}_1 & \mathbf{u}_k\cdot\mathbf{u}_2 & \cdots & \mathbf{u}_k\cdot\mathbf{u}_k \end{pmatrix},$$

that is, the (i,j)-entry of the product $\mathbf{Q}^T\mathbf{Q}$ is the inner product $\mathbf{u}_i \cdot \mathbf{u}_j$. Hence,

$$S \text{ is } \begin{cases} \text{orthogonal} \\ \text{orthonormal} \end{cases} \Leftrightarrow \mathbf{Q}^T\mathbf{Q} \text{ is } \begin{cases} \text{a diagonal matrix} \\ \text{the identity matrix} \end{cases}.$$

- In particular, if $k = n$, then $\mathbf{Q}$ is a square matrix and $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_n$ implies that $\mathbf{Q}^T = \mathbf{Q}^{-1}$.
- That is, S is orthonormal if and only if $\mathbf{Q}^T = \mathbf{Q}^{-1}$.

Orthogonal matrices

Definition

A $n \times n$ square matrix $\mathbf{A}$ is *orthogonal* if $\mathbf{A}^T = \mathbf{A}^{-1}$, equivalently, $\mathbf{A}^T\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^T$.

Theorem

Let $\mathbf{A}$ *be a square matrix of order* n*. The following statements are equivalent.*

(i) $\mathbf{A}$ *is an orthogonal matrix.*

(ii) *The columns of* $\mathbf{A}$ *form an orthonormal basis for* $\mathbb{R}^n$*.*

(iii) *The rows of* $\mathbf{A}$ *form an orthonormal basis for* $\mathbb{R}^n$*.*

Remarks

- Note that $\mathbf{A}$ is an orthogonal matrix if and only if the columns/rows form an orthonormal basis for $\mathbb{R}^n$.
- We do not have a name given to matrices whose the rows or columns form an orthogonal basis.
- While some textbooks may define a square matrix whose inverse is the transpose as an orthonormal matrix, this course will follow the majority of the literature and call it an orthogonal matrix instead.

5.3 Orthogonal Projection

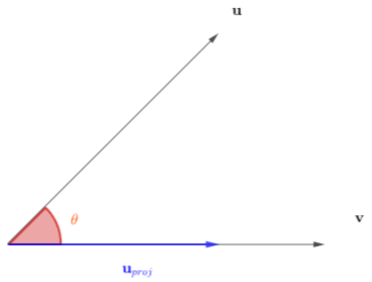
Discussion

Recall that the angle between 2 nonzero vectors **u** and **v** is given by

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}.$$

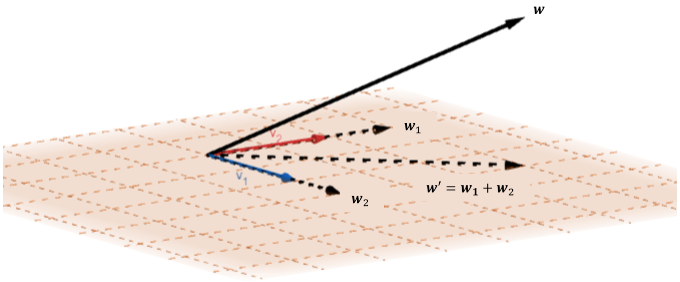
Recall also that the projection of **u** onto the direction of **v** is

$$\|\mathbf{u}\| \cos(\theta) \frac{\mathbf{v}}{\|\mathbf{v}\|} = \|\mathbf{u}\| \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v}.$$



Discussion

- Now suppose $S = \{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for a subspace V. Then
- Then $\mathbf{w}' = \frac{\mathbf{w}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{w}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2$ is the projection of $\mathbf{w}$ onto $\text{span}(S) = V$.
- Moreover, the vector $\mathbf{w}'$ is independent of the choice of orthogonal basis for V. See `https://www.geogebra.org/m/mndafgcp`.



Orthogonal Projection

Theorem (Orthogonal projection theorem)

Let V be a subspace of $\mathbb{R}^n$. Every vector $\mathbf{w}$ in $\mathbb{R}^n$ can be decomposed uniquely as a sum

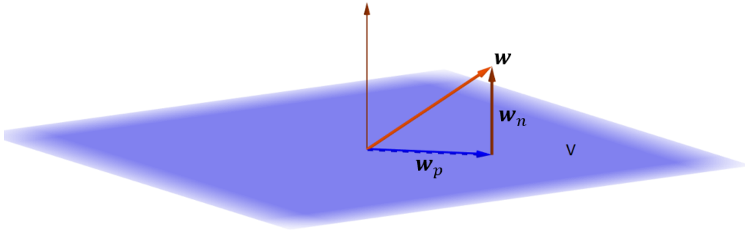
$$\mathbf{w} = \mathbf{w}_p + \mathbf{w}_n,$$

where $\mathbf{w}_n$ is orthogonal to V and $\mathbf{w}_p$ is a vector in V, $\mathbf{w}_n \perp V$, $\mathbf{w}_p \in V$. Moreover, if $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ is an orthogonal basis for V, then

$$\mathbf{w}_p = \frac{\mathbf{w} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{w} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 + \cdots + \frac{\mathbf{w} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k}\mathbf{u}_k.$$

Definition

Define the vector $\mathbf{w}_p$ in the theorem above as the *orthogonal projection* (or just *projection*) of $\mathbf{w}$ onto the subspace V.



Best Approximation Theorem

Theorem (Best approximation theorem)

Let V be a subspace of $\mathbb{R}^n$ and $\mathbf{w}$ a vector in $\mathbb{R}^n$. Let $\mathbf{w}_p$ be the projection of $\mathbf{w}$ onto V. Then $\mathbf{w}_p$ is vector in V closest to $\mathbf{w}$; that is,

$$\|\mathbf{w} - \mathbf{w}_p\| \leq \|\mathbf{w} - \mathbf{v}\|$$

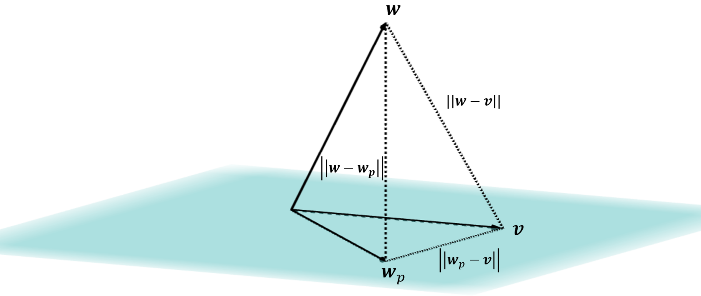
for all $\mathbf{v}$ in V.

Proof.

For any $\mathbf{v}$ in V, $\mathbf{v} - \mathbf{w}_p$ is in V. Thus, $\mathbf{v} - \mathbf{w}_p$ is orthogonal to $\mathbf{w} - \mathbf{w}_p$. Hence, by Pythagorean theorem,

$$\|\mathbf{w} - \mathbf{v}\|^2 = \|\mathbf{w} - \mathbf{w}_p\|^2 + \|\mathbf{v} - \mathbf{w}_p\|^2.$$

Since $\|\mathbf{v} - \mathbf{w}_p\|^2 \geq 0$, the inequality $\|\mathbf{w} - \mathbf{w}_p\| \leq \|\mathbf{w} - \mathbf{v}\|$ is established. $\square$



Challenge

Prove the orthogonal projection theorem.

Theorem (Orthogonal projection theorem)

Let V be a subspace of $\mathbb{R}^n$. Every vector $\mathbf{w}$ in $\mathbb{R}^n$ can be decomposed uniquely as a sum

$$\mathbf{w} = \mathbf{w}_p + \mathbf{w}_n,$$

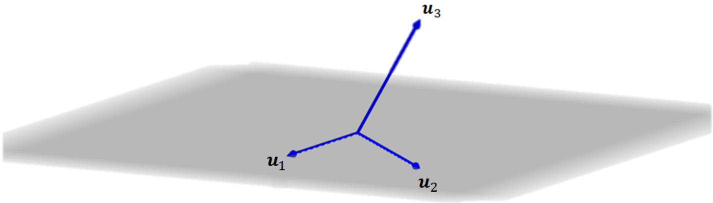
where $\mathbf{w}_n$ is orthogonal to V and $\mathbf{w}_p$ is a vector in V, $\mathbf{w}_n \perp V$, $\mathbf{w}_p \in V$.

Introduction to Gram-Schmidt Process

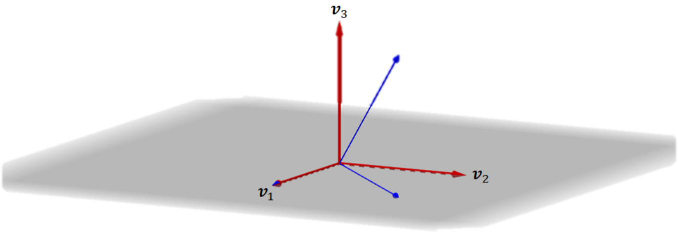
- To compute the projection of a vector $\mathbf{w}$ onto a subspace V, we need to find an orthogonal or orthonormal basis.
- Suppose now S is a basis for a subspace V. The Gram-Schmidt process converts S to an orthonormal basis.
- Given a linear independent set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$, we want to find an orthogonal set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ such that $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$.
- Let $\mathbf{v}_1 = \mathbf{u}_1$, then $\text{span}\{\mathbf{u}_1\} = \text{span}\{\mathbf{v}_1\}$.
- If we let $\mathbf{v}_2 = \mathbf{u}_2 - \mathbf{u}_2'$, where $\mathbf{u}_2'$ is the projection of $\mathbf{u}_2$ onto $\text{span}\{\mathbf{v}_1\}$, then $\mathbf{v}_2$ is orthogonal to $\mathbf{v}_1$. Moreover, it should be clear that $\text{span}\{\mathbf{u}_1, \mathbf{u}_2\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$.
- Continue this process, where we define $\mathbf{v}_{i+1} = \mathbf{u}_{i+1} - \mathbf{u}_{i+1}'$, where $\mathbf{u}_{i+1}'$ is the projection of $\mathbf{u}_{i+1}$ onto $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_i\}$.
- Then $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ will be an orthogonal set such that $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$.
- If we normalize $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$, we will have an orthonormal basis.

Gram-Schmidt Process

We will give a visualization of the Gram-Schmidt process in $\mathbb{R}^3$.



$$\Downarrow$$



Gram-Schmidt Process

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a linearly independent set. Let

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1 \\
\mathbf{v}_2 &= \mathbf{u}_2 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_2}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{u}_3 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_3}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_3}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_k &= \mathbf{u}_k - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_k}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_k}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 - \cdots - \left(\frac{\mathbf{v}_{k-1} \cdot \mathbf{u}_k}{\|\mathbf{v}_{k-1}\|^2}\right)\mathbf{v}_{k-1}.
\end{aligned}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal set (of nonzero vectors), and hence,

$$\left\{\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, \ldots, \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\right\}$$

is an orthonormal set such that $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \operatorname{span}\{\mathbf{u}_k, \ldots, \mathbf{u}_k\}$.

Discussion

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ be a set in $\mathbb{R}^4$. After performing the Gram-Schmidt process on S, $\mathbf{v}_4 = \mathbf{0}$, but $\mathbf{v}_3 \neq \mathbf{0}$. What can you conclude?

5.4 QR Factorization

QR Factorization

Suppose now $\mathbf{A}$ is a $m \times n$ matrix with linearly independent columns, i.e. $\text{rank}(\mathbf{A}) = n$. Write

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{pmatrix}.$$

Since the set $S = \{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n\}$ is linearly independent we may apply the Gram-Schmidt process on S to obtain an orthonormal set $\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n\}$. Set

$$\mathbf{Q} = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{pmatrix}.$$

Recall that for any $j = 1, 2, ..., n$, $\text{span}\{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_j\} = \text{span}\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_j\}$. In particular, $\mathbf{a}_j$ is in $\text{span}\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_j\}$. Thus we may write

$$\mathbf{a}_j = r_{1j}\mathbf{q}_1 + r_{2j}\mathbf{q}_2 + \cdots + r_{jj}\mathbf{q}_j + 0\mathbf{q}_{j+1} + \cdots + 0\mathbf{q}_n = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_j & \cdots & \mathbf{q}_n \end{pmatrix} \begin{pmatrix} r_{1j} \\ \vdots \\ r_{jj} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

QR Factorization

Explicitly,

$$
\begin{aligned}
\mathbf{a}_1 &= r_{11}\mathbf{q}_1 = \begin{pmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_n\end{pmatrix}\begin{pmatrix} r_{11} \\ 0 \\ \vdots \\ 0\end{pmatrix} \\
\mathbf{a}_2 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 = \begin{pmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_n\end{pmatrix}\begin{pmatrix} r_{12} \\ r_{22} \\ \vdots \\ 0\end{pmatrix} \\
&\vdots \\
\mathbf{a}_n &= r_{1n}\mathbf{q}_1 + r_{2n}\mathbf{q}_2 + \cdots + r_{nn}\mathbf{q}_n = \begin{pmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_n\end{pmatrix}\begin{pmatrix} r_{1n} \\ r_{2n} \\ \vdots \\ r_{nn}\end{pmatrix}
\end{aligned}
$$

Thus, we may write

$$
\begin{aligned}
\mathbf{A} &= \begin{pmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{pmatrix} \\
&= \begin{pmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n\end{pmatrix}\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn}\end{pmatrix} \\
&= \mathbf{QR}
\end{aligned}
$$

for some $m \times n$ matrix $\mathbf{Q}$ with orthonormal columns, and a upper triangular $n \times n$ matrix $\mathbf{R}$.

QR Factorization

Theorem (QR Factorization)

Suppose $\mathbf{A}$ *is a* $m \times n$ *matrix with linearly independent columns. Then* $\mathbf{A}$ *can be written as*

$$\mathbf{A} = \mathbf{Q}\mathbf{R}$$

for some $m \times n$ *matrix* $\mathbf{Q}$ *such that* $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_n$ *and invertible upper triangular matrix* $\mathbf{R}$ *with positive diagonal entries.*

Definition

The decomposition given in the theorem above is called a *QR factorization* of $\mathbf{A}$.

Algorithm to QR Factorization

Let $\mathbf{A}$ be a $m \times n$ matrix with linearly independent columns.

1. Perform Gram-Schmidt on the columns of $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{pmatrix}$ to obtain an orthonormal set $\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n\}$.

2. Set $\mathbf{Q} = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{pmatrix}$.

3. Compute $\mathbf{R} = \mathbf{Q}^T\mathbf{A}$.

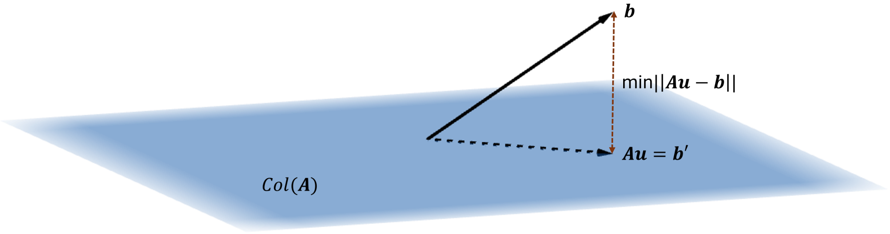
5.5 Least Square Approximation

Introduction

- Recall that the column space of a $m \times n$ matrix $\mathbf{A}$ is the set of all vectors $\mathbf{b}$ such that $\mathbf{Ax} = \mathbf{b}$ is consistent,

$$\mathrm{Col}(\mathbf{A}) = \left\{ \ \mathbf{b} \in \mathbb{R}^m \ \middle| \ \mathbf{Ax} = \mathbf{b} \text{ is consistent } \right\} = \left\{ \ \mathbf{Au} \ \middle| \ \mathbf{u} \in \mathbb{R}^n \ \right\}.$$

- Now suppose $\mathbf{Ax} = \mathbf{b}$ is inconsistent, that is, $\mathbf{b}$ is not in the column space of $\mathbf{A}$.
- We may ask for a vector $\mathbf{b}'$ in the column of $\mathbf{A}$ that is the closest to $\mathbf{b}$, that is, find a $\mathbf{u}$ such that $\mathbf{Au} = \mathbf{b}'$ is the closest to $\mathbf{b}$.
- This is equivalent to finding a $\mathbf{u}$ in $\mathbb{R}^n$ such that $\|\mathbf{Au} - \mathbf{b}\|$ is minimized.



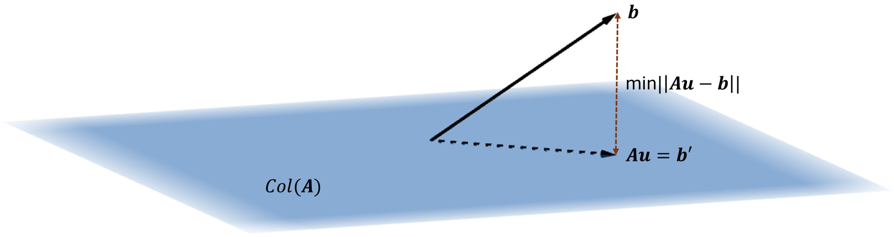
Least Square Approximation

Definition

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b}$ a vector in $\in \mathbb{R}^m$. A vector $\mathbf{u}$ in $\mathbb{R}^n$ is a *least square solution* of $\mathbf{Ax} = \mathbf{b}$ if for every vector $\mathbf{v} \in \mathbb{R}^n$,

$$\|\mathbf{Au} - \mathbf{b}\| \leq \|\mathbf{Av} - \mathbf{b}\|.$$

Geometrically, by the best approximation theorem, the vector $\mathbf{b}' = \mathbf{Au}$ in $\mathrm{Col}(\mathbf{A})$ closest to $\mathbf{b}$ is the projection of $\mathbf{b}$ onto $\mathrm{Col}(\mathbf{A})$.



Least Square Approximation

Theorem

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b}$ a vector in $\mathbb{R}^m$. A vector $\mathbf{u}$ in $\mathbb{R}^n$ is a least square solution to $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{Au}$ is the projection of $\mathbf{b}$ onto the column space of $\mathrm{Col}(\mathbf{A})$.

Theorem

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b}$ a vector in $\mathbb{R}^m$. A vector $\mathbf{u}$ in $\mathbb{R}^n$ is a least square solution to $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{u}$ is a solution to $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$.

Proof.

$\mathbf{Au}$ is the projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$ if and only if $\mathbf{Au} - \mathbf{b}$ is orthogonal to the column space of $\mathbf{A}$. By the orthogonal to a subspace theorem, since the columns of $\mathbf{A}$ spans the column space of $\mathbf{A}$, $\mathbf{Au} - \mathbf{b}$ is orthogonal to the column space of $\mathbf{A}$ if and only if $\mathbf{Au} - \mathbf{b}$ is in the nullspace of $\mathbf{A}^T$,

$$\mathbf{A}^T(\mathbf{Au} - \mathbf{b}) = \mathbf{0}$$

Rearranging the equation, we have

$$\mathbf{A}^T\mathbf{Au} = \mathbf{A}^T\mathbf{b}.$$

□

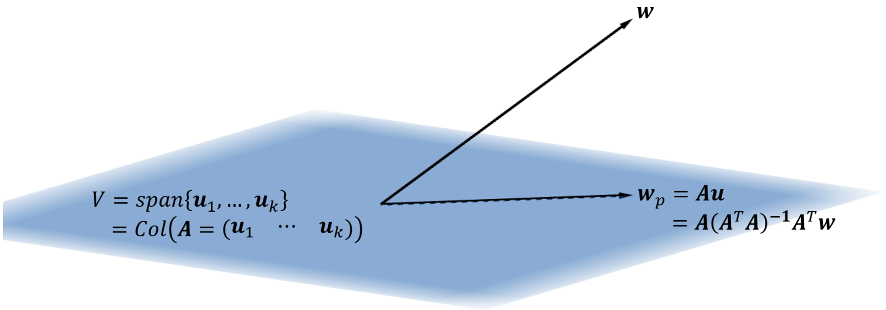
Challenge

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b}$ a vector in $\mathbb{R}^m$. Prove that for any choice of least square solution $\mathbf{u}$, that is, for any solution $\mathbf{u}$ of $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$, the projection $\mathbf{A}\mathbf{u}$ is unique.

Orthogonal Projection (Revisit)

We may use least square solutions to find the projection of a vector onto a subspace.

- Let V be subspace of $\mathbb{R}^n$. Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a basis for V.
- Define $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k \end{pmatrix}$, by construction, the column space of $\mathbf{A}$ is V, $V = \mathrm{Col}(\mathbf{A})$.
- Let $\mathbf{w}$ be a vector in $\mathbb{R}^n$, and $\mathbf{u}$ a least square solution to $\mathbf{A}\mathbf{x} = \mathbf{w}$. Then $\mathbf{w}_p = \mathbf{A}\mathbf{u}$ is the projection of $\mathbf{w}$ onto $\mathrm{Col}(\mathbf{A}) = V$.



Orthogonal Projection (Revisit)

- Now $\mathbf{u}$ is a least square solution to $\mathbf{Ax} = \mathbf{w}$ if and only if it is a solution to $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{w}$. But since the columns of $\mathbf{A}$ are linearly independent, $\mathbf{A}^T\mathbf{A}$ is invertible.
- This means that $\mathbf{u} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w}$.
- Hence, the projection is

$$\mathbf{w}_p = \mathbf{Au} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w}.$$

Theorem

Let V be a subspace of $\mathbb{R}^n$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for V. Then the orthogonal projection of a vector $\mathbf{w}$ onto V is

$$\mathbf{w}_p = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w},$$

where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$.

Exercise

Suppose $\mathbf{A}$ is a $m \times n$ matrix with linearly independent columns, i.e. $\text{rank}(\mathbf{A}) = n$. QR factorize $\mathbf{A}$,

$$\mathbf{A} = \mathbf{QR}.$$

Show that the unique least square solution of $\mathbf{Ax} = \mathbf{b}$ is

$$\mathbf{u} = \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{b}.$$

That is, suffice to solve for

$$\mathbf{Rx} = \mathbf{Q}^T\mathbf{b}.$$

This is easy to solve by hand since $\mathbf{R}$ is a upper triangular matrix (i.e. a REF).

Challenge

Let $V \subseteq \mathbb{R}^n$ be a subspace and suppose $\{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_k\}$ is an orthonormal basis of V. Write

$$\mathbf{Q} = \begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_k \end{pmatrix}.$$

Then for any $\mathbf{w} \in \mathbb{R}^n$, the projection of $\mathbf{w}$ onto V is

$$\mathbf{Q}\mathbf{Q}^T\mathbf{w}.$$

MA1522: Linear Algebra for Computing

Chapter 6: Eigenanalysis

6.1 Eigenvalues and Eigenvectors

Eigenvalues and Eigenvectors

Let $\mathbf{A}$ be a square matrix of order n.

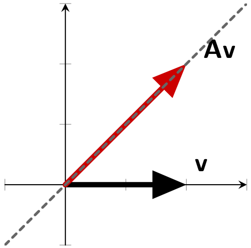
- For any vector $\mathbf{u}$ in $\mathbb{R}^n$, $\mathbf{Au}$ is also a vector in $\mathbb{R}^n$.
- So, we may think of $\mathbf{A}$ as a map $\mathbb{R}^n \to \mathbb{R}^n$, moving vectors around in $\mathbb{R}^n$, $\mathbf{u} \mapsto \mathbf{Au}$.
- That is, $\mathbf{A}$ defines a linear mapping. More on this in the next chapter.

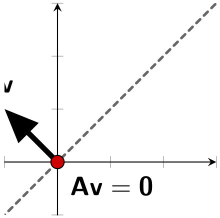
Visit https://www.geogebra.org/m/sbkscz46 to visualize how vectors are moved around by a matrix $\mathbf{A}$.

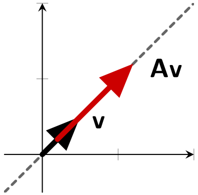
Eigenvalues and Eigenvectors

Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. It takes a vector and maps it to a vector along the line $x = y$ such that both coordinates in $\mathbf{Av}$ are the sum of the coordinates in $\mathbf{v}$.

$$\mathbf{A}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$







Observe that any vector $\mathbf{v}$ along the line $x = y$ is mapped to twice itself, $\mathbf{Av} = 2\mathbf{v}$, and it take any vector $\mathbf{v}$ along the line $x = -y$ to the origin, $\mathbf{Av} = \mathbf{0}$.

Eigenvalues and Eigenvectors

Definition

Let $\mathbf{A}$ be a square matrix of order n. A real number λ is an *eigenvalue* of $\mathbf{A}$ if there is a nonzero vector $\mathbf{v}$ in $\mathbb{R}^n$, $\mathbf{v} \neq \mathbf{0}$, such that

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

In this case, the nonzero vector $\mathbf{v}$ is called an *eigenvector* associated to λ.

Remark

- Geometrically, eigenvectors are the vectors that are being scaled (stretch, dilate, or reflect) when acted upon by $\mathbf{A}$, and eigenvalues are the amount to scale the eigenvectors.
- We require the eigenvector to be nonzero, $\mathbf{v} \neq \mathbf{0}$, for otherwise, the identity

  $$\mathbf{A}\mathbf{0} = \lambda\mathbf{0}$$

  holds for every real number λ, which means that every real number is an eigenvalue of $\mathbf{A}$. This renders the definition not meaningful and uninteresting.

Characteristic Polynomial

Definition

Let $\mathbf{A}$ be a square matrix of order n, the *characteristic polynomial* of $\mathbf{A}$, denoted as $\text{char}(\mathbf{A})$, is the degree n polynomial

$$\det(x\mathbf{I} - \mathbf{A}).$$

Example

1. Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The characteristic polynomial is $\det\left(\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) = \begin{vmatrix} x & -1 \\ -1 & x \end{vmatrix} = x^2 - 1$.
2. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. $\det\left(\begin{pmatrix} x-1 & -1 \\ -1 & x-1 \end{pmatrix}\right) = \begin{vmatrix} x-1 & -1 \\ -1 & x-1 \end{vmatrix} = (x-1)^2 - 1 = x(x-2)$.
3. Let $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 3 & 1 \end{pmatrix}$. $\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & 0 & 0 \\ 0 & x & -2 \\ 0 & -3 & x-1 \end{vmatrix} = (x-1)[x(x-1) - 6] = (x-1)(x+2)(x-3)$.

Finding Eigenvalues

Recall that λ is an eigenvalue if there is a nonzero $\mathbf{v}$ such that $\mathbf{Av} = \lambda\mathbf{v}$. Manipulating the equation, we have

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

which shows that $\mathbf{v}$ is a nontrivial solution to the homogeneous system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$.

Theorem

Let $\mathbf{A}$ be a square matrix of order n. $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathbf{A}$ if and only if the homogeneous system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

Observe that $\lambda\mathbf{I} - \mathbf{A}$ is a square matrix. And so, $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ has nontrivial solutions if and only if $(\lambda\mathbf{I} - \mathbf{A})$ is singular, which can be checked using its determinant. Hence,

Theorem

Let $\mathbf{A}$ be a square matrix of order n. λ is an eigenvalue of $\mathbf{A}$ if and only if λ is a root of the characteristic polynomial $\det(x\mathbf{I} - \mathbf{A})$.

Eigenvalue and Invertibility

Question: Can $\lambda = 0$ be an eigenvalue of $\mathbf{A}$?

Suppose 0 is an eigenvalue of $\mathbf{A}$. Let $\mathbf{v}$ be an eigenvector of $\mathbf{A}$ associated to eigenvalue 0. Then $\mathbf{v} \neq \mathbf{0}$ is a nonzero vector such that

$$\mathbf{Av} = \mathbf{v} = \mathbf{0}.$$

This shows that the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has nontrivial solutions, and hence, $\mathbf{A}$ is singular.

It is easy to see that the converse is true too, that is, if $\mathbf{A}$ is singular, then $\lambda = 0$ is an eigenvalue.

Theorem

A square matrix $\mathbf{A}$ is invertible if and only if $\lambda = 0$ is not an eigenvalue of $\mathbf{A}$.

We will add this to the equivalent statements of invertibility.

Equivalent Statements of Invertibility

Theorem (Equivalent Statements for Invertibility)

Let $\mathbf{A}$ *be a square matrix of order* n*. The following statements are equivalent.*

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}^T$ *is invertible.*

(iii) *(left inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{BA} = \mathbf{I}$*.*

(iv) *(right inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{AB} = \mathbf{I}$*.*

(v) *The reduced row-echelon form of* $\mathbf{A}$ *is the identity matrix.*

(vi) $\mathbf{A}$ *can be expressed as a product of elementary matrices.*

(vii) *The homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution.*

(viii) *For any* $\mathbf{b}$*, the system* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution.*

(ix) *The determinant of* $\mathbf{A}$ *is nonzero,* $\det(\mathbf{A}) \neq 0$*.*

(x) *The columns/rows of* $\mathbf{A}$ *are linearly independent.*

(xi) *The columns/rows of* $\mathbf{A}$ *spans* $\mathbb{R}^n$*.*

(xii) $\text{rank}(\mathbf{A}) = n$ *(*$\mathbf{A}$ *has full rank).*

(xiii) $\text{nullity}(\mathbf{A}) = 0$*.*

(xiv) 0 *is not an eigenvalue of* $\mathbf{A}$*.*

Algebraic Multiplicity

Let λ be an eigenvalue of $\mathbf{A}$. The algebraic multiplicity of λ is the largest integer r_λ such that

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda)^{r_\lambda} p(x),$$

for some polynomial $p(x)$. Alternatively, r_λ is the positive integer such that in the above equation, λ is not a root of $p(x)$.

Suppose $\mathbf{A}$ is an order n square matrix such that $\det(x\mathbf{I} - \mathbf{A})$ can be factorize into linear factors completely. Then we can write

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_k)^{r_k}$$

where $r_1 + r_2 + \cdots + r_k = n$, and $\lambda_1, \lambda_2, ..., \lambda_k$ are the distinct eigenvalues of $\mathbf{A}$. Then the algebraic multiplicity of λ_i is r_i for $i = 1, ..., k$.

Eigenvalues of Triangular Matrices

Theorem

The eigenvaules of a triangular matrix are the diagonal entries. The algebraic multiplicity of the eigenvalue is the number of times it appears as a diagonal entry of $\mathbf{A}$.

Proof.

We will prove for the case where $\mathbf{A}$ is an upper triangular matrix. The proof for lower triangular matrix is analogous.

Suppose $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$. Then

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{pmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ 0 & x - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x - a_{nn} \end{pmatrix} = (x - a_{11})(x - a_{22}) \cdots (x - a_{nn}).$$

□

Eigenspace

Recall that eigenvectors of $\mathbf{A}$ associated to eigenvalue λ are nontrivials solution to

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

Since the system is homogeneous, the set of all solutions is a subspace. We will call it the eigenspace of $\mathbf{A}$ associated to eigenvalue λ.

Definition

Let $\mathbf{A}$ be an order n square matrix. The *eigenspace* associated to an eigenvalue λ of $\mathbf{A}$ is

$$E_\lambda = \{\ \mathbf{v} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\ \} = \mathrm{Null}(\lambda \mathbf{I} - \mathbf{A}).$$

The *geometric multiplicity* of an eigenvalue λ is the dimension of its eigenspace,

$$\dim(E_\lambda) = \mathrm{nullity}(\lambda \mathbf{I} - \mathbf{A}).$$

Challenge

Let $\mathbf{A}$ be a $n \times n$ matrix.

1. Show that the characteristic polynomial of $\mathbf{A}$ is equal to the characteristic polynomial of $\mathbf{A}^T$. Hence $\mathbf{A}$ and $\mathbf{A}^T$ has the same eigenvalues.

2. Let λ be an eigenvalue of $\mathbf{A}$. Show that the geometric multiplicity of λ as an eigenvalue of $\mathbf{A}$ is equal to its geometric multiplicity as an eigenvalue of $\mathbf{A}^T$.

6.2 Diagonalizaton

Diagonalization

Definition

A square matrix $\mathbf{A}$ of order n is *diagonalizable* if there exists an invertible matrix $\mathbf{P}$ such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$$

is a diagonal matrix, OR

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}.$$

Diagonalization

Previous example:

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1}.$$

- $\begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. This shows that the first column of **P** is an eigenvector associated to eigenvalue 2, the (1,1) diagonal entry of **D**.
- $\begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ This shows that the second column of **P** is an eigenvector associated to eigenvalue 2, the (2,2) diagonal entry of **D**.
- $\begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 0 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ This shows that the third column of **P** is an eigenvector associated to eigenvalue 4, the (3,3) diagonal entry of **D**.

Diagonalization

Theorem (Diagonalizability)

A $n \times n$ square matrix $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ has n linearly independent eigenvectors.

Proof.

First observe that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$ if and only if $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}$. Write $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ and $\mathbf{D} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix}$, then

$$\begin{pmatrix} \mathbf{A}\mathbf{u}_1 & \mathbf{A}\mathbf{u}_2 & \cdots & \mathbf{A}\mathbf{u}_n \end{pmatrix} = \mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D} = \begin{pmatrix} \mu_1\mathbf{u}_1 & \mu_2\mathbf{u}_2 & \cdots & \mu_n\mathbf{u}_n \end{pmatrix},$$

and thus by comparing the columns, we have $\mathbf{A}\mathbf{u}_i = \mu_i\mathbf{u}_i$. This shows that the columns of $\mathbf{P}$ are eigenvectors of $\mathbf{A}$. Now $\mathbf{P}$ is invertible if and only if its columns form a basis for $\mathbb{R}^n$. Hence, $\mathbf{A}$ is diagonalizable if and only if we can find a basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$. $\square$

Diagonalization

That is, $\mathbf{A}$ is diagonalizable if and only if we can find

$$\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}, \text{ and } \mathbf{D} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$

where μ_i is the eigenvalue associated to eigenvector $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$.

$\mathbf{P}$ is invertible if and only if $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$.

Note that μ_i may not be distinct.

Not Diagonalizable

Not all square matrices are diagonalizable. For example, consider

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

This is a triangular matrix, with only one eigenvalue $\lambda = 0$.

$$0\mathbf{I} - \mathbf{A} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

tells us that **A** has only 1 linearly independent eigenvector associated to the only eigenvalue $\lambda = 0$. Hence, **A** is not diagonalizable.

Not Diagonalizable

Consider $\mathbf{A} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$.

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-2 & 1 \\ -1 & x \end{vmatrix} = (x-1)^2;$$

$\mathbf{A}$ has only one eigenvalue $\lambda = 1$.

$$\mathbf{I} - \mathbf{A} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$$

tell us that $\mathbf{A}$ has only 1 linearly independent eigenvector. Hence, $\mathbf{A}$ is not diagonalizable.

Notice that in both examples above, the algebraic multiplicities are greater than the geometric multiplicities.

Independence of Eigenspaces

Suppose λ_1 and λ_2 are distinct eigenvalues. Let $\mathbf{v}_1$ be an eigenvector associated to eigenvalue λ_1. Then since $\lambda_1 \neq \lambda_2$,

$$\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_1 \neq \lambda_2\mathbf{v}_1,$$

$\mathbf{v}_1$ cannot be in the eigenspace associated to λ_2. This demonstrates that vectors from different eigenspaces are linearly independent. The proof of the following theorem is given in the appendix.

Theorem (Eigenspaces are linearly independent)

Let $\mathbf{A}$ be a $n \times n$ square matrix. Let λ_1 and λ_2 are distinct eigenvalues of $\mathbf{A}$, $\lambda_1 \neq \lambda_2$. Suppose $\{\mathbf{u}_1, ..., \mathbf{u}_k\}$ is a linearly independent subset of eigenspace associated to eigenvalue λ_1, and $\{\mathbf{v}_1, ..., \mathbf{v}_m\}$ is a linearly independent subset of of eigenspace associated to eigenvalue λ_2. Then the union $\{\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{v}_1, ..., \mathbf{v}_m\}$ is linearly independent.

Visualization of Eigenspaces

Click on the following link, https://www.geogebra.org/3d/u87k4uah, to visualize the independence of the he eigenspaces.

1. Let $\mathbf{A} = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 2 & 0 \\ 1 & 1 & 2 \end{pmatrix}$. It is a diagonalizable matrix.

2. The eigenvalues are $\lambda = 2, 3$, with algebraic and thus geometric multiplicities $r_2 = \dim(E_2) = 2$ and $r_3 = \dim(E_3) = 1$, respectively.

3. At the side, if we let $c_3 = 0$, then for any scalars c_1, c_2, $\mathbf{w}$ is a vector in the blue plane and $\mathbf{Aw} = 2\mathbf{w}$. If we let $c_1 = c_2 = 0$, $\mathbf{w}$ is a vector in alone the red line and $\mathbf{Aw} = 3\mathbf{w}$.

4. This shows that blue plane is the eigenspace E_2, and the red line is the eigenspace E_3.

5. It is clear from the plot that vectors in the blue plane and the red line are independent.

Equivalent Statements for Diagonalizability

The geometric multiplicity is bounded above by the algebraic multiplicity. The proof can be found in the appendix.

Theorem (Geometric Multiplicity is no greater than Algebraic multiplicity)

The geometric multiplicity of an eigenvalue λ of a square matrix $\mathbf{A}$ is no greater than the algebraic multiplicity, that is,

$$1 \leq \dim(E_\lambda) \leq r_\lambda.$$

Equivalent Statements for Diagonalizability

Let $\mathbf{A}$ be a $n \times n$ matrix. Let $\lambda_1, \lambda_2, ..., \lambda_p$ be the distinct eigenvalues of $\mathbf{A}$ with algebraic multiplicities $r_1, r_2, ..., r_p$ respectively. Let

$$\{\mathbf{v}_{i,1}, ..., \mathbf{v}_{i,k_i}\} \text{ be a basis for } E_{\lambda_i},$$

the eigenspace associated to eigenvalue λ_i, i.e. $\dim(E_{\lambda_i}) = k_i$. Collect the bases together, we get the set $\{\mathbf{v}_{1,1}, ..., \mathbf{v}_{1,k_1}, \mathbf{v}_{2,1}, ..., \mathbf{v}_{2,k_2}, \dots ..., \mathbf{v}_{p,1}, ..., \mathbf{v}_{p,k_p}\}$. Now

$$\begin{array}{ccccccccc}
k_1 & + & k_2 & + & \cdots & + & k_p & & \\
\| & & \| & & \cdots & & \| & & \\
\dim(E_{\lambda_1}) & + & \dim(E_{\lambda_2}) & + & \cdots & + & \dim(E_{\lambda_k}) & & \\
\text{I}\wedge & & \text{I}\wedge & & \cdots & & \text{I}\wedge & & \\
r_1 & + & r_2 & + & \cdots & + & r_k & \leq & n
\end{array}$$

If $\mathbf{A}$ is diagonalizable, then $k_1 + k_2 + \cdots + k_p = n$. Thus, necessarily $r_1 + r_2 + \cdots + r_p = n$ and $\dim(E_{\lambda_i}) = r_i$. On the other hand, if $r_1 + r_2 + \cdots + r_p = n$ and $\dim(E_{\lambda_i}) = r_i$, then $k_1 + k_2 + \cdots + k_p = n$ and since $\{\mathbf{v}_{1,1}, ..., \mathbf{v}_{1,k_1}, \mathbf{v}_{2,1}, ..., \mathbf{v}_{2,k_2}, \dots ..., \mathbf{v}_{p,1}, ..., \mathbf{v}_{p,k_p}\}$ is linearly independent, $\mathbf{A}$ has n linearly independent eigenvectors, and is thus diagonalizable.

Equivalent Statements for Diagonalizability

Theorem

Let $\mathbf{A}$ *be a square matrix of order* n*. The following statements are equivalent.*

(i) $\mathbf{A}$ *is diagonalizable.*

(ii) *There exists a basis* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *of* $\mathbb{R}^n$ *of eigenvectors of* $\mathbf{A}$*.*

(iii) *The characteristic polynomial of* $\mathbf{A}$ *splits into linear factors,*

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}},$$

where r_{λ_i} *is the algebraic multiplicity of* λ_i*, for* $i = 1, ..., k$*, and the eigenvalues are distinct,* $\lambda_i \neq \lambda_j$ *for all* $i \neq j$*, and the geometric multiplicity is equal to the algebraic multiplicity for each eigenvalue* λ_i*,*

$$\dim(E_{\lambda_i}) = r_{\lambda_i}.$$

Not Diagonalizable

A square matrix $\mathbf{A}$ is diagonalizable if

(i) The characteristic polynomial splits into linear factors,

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}},$$

(ii) and the algebraic multiplicity is equal to the geometric multiplicity,

$$r_\lambda = \dim(E_\lambda),$$

for every eigenvalue λ of $\mathbf{A}$.

To show that a square matrix $\mathbf{A}$ of order n is not diagonalizable, show that either

(i) $\det(x\mathbf{I} - \mathbf{A})$ does not split into linear factors, or

(ii) there exists an eigenvalue λ such that $\dim(E_\lambda) < r_\lambda$.

Exercise

Suppose $\mathbf{A}$ is a $n \times n$ matrix with $n > 1$. Show that if $\mathbf{A}$ has only 1 eigenvalue λ, then $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ is the scalar matrix, $\mathbf{A} = \lambda \mathbf{I}_n$.

Hence, all non-scalar matrix with only 1 eigenvalue is not diagonalizable.

Distinct Eigenvalues

Theorem

If $\mathbf{A}$ *is a square matrix of order* n *with* n *distinct eigenvalues, then* $\mathbf{A}$ *is diagonalizable.*

Sketch of Proof.

Follows from

$$
\begin{array}{ccccccccccc}
n & = & 1 & + & 1 & + & \cdots & + & 1 & & \\
 & & \leq & & \leq & & \cdots & & \leq & & \\
 & & \dim(E_{\lambda_1}) & + & \dim(E_{\lambda_2}) & + & \cdots & + & \dim(E_{\lambda_n}) & & \\
 & & \leq & & \leq & & \cdots & & \leq & & \\
 & & r_1 & + & r_2 & + & \cdots & + & r_n & \leq & n
\end{array}
$$

$\square$

Algorithm to Diagonalization

Let $\mathbf{A}$ be an order n square matrix.

1. Compute the characteristic polynomial

$$\det(x\mathbf{I}-\mathbf{A}) = (x-\lambda_1)^{r_{\lambda_1}}(x-\lambda_2)^{r_{\lambda_2}}\cdots(x-\lambda_k)^{r_{\lambda_k}}.$$

   If the characteristic polynomial do not split into linear factors, $\mathbf{A}$ is not diagonalizable.

2. For each eigenvalue λ_i of $\mathbf{A}$, $i = 1, ..., k$, find a basis S_{λ_i} for the eigenspace, that is, find a basis S_{λ_i} for the solution space of the following linear system,

$$(\lambda_i\mathbf{I}-\mathbf{A})\mathbf{x} = \mathbf{0}.$$

   Compute first the eigenspace associated to eigenvalues with algebraic multiplicity greater than 1. If $\dim(E_\lambda) < r_\lambda$, $\mathbf{A}$ is not diagonalizable.

3. Let $S = \bigcup_{i=1}^{k} S_{\lambda_i}$. Then $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$.
4. Let $\mathbf{P} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{pmatrix}$, and $\mathbf{D} = \operatorname{diag}(\mu_1, \mu_2, ..., \mu_n)$, where μ_i is the eigenvalue associated to $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i\mathbf{u}_i$. Then

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}.$$

6.3 Orthogonally Diagonalizable

Orthogonally Diagonalizable

Definition

An order n square matrix $\mathbf{A}$ is orthogonally diagonalizable if

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$$

for some orthogonal matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$.

Remark

Note that since $\mathbf{P}$ is orthogonal, $\mathbf{P}^T = \mathbf{P}^{-1}$, then $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. That is, orthogonally diagonalizable matrices are also diagonalizable, except we need $\mathbf{P}$ to not only be invertible, but also an orthogonal matrix.

The Spectral Theorem

Theorem (Spectral theorem)

Let $\mathbf{A}$ be a $n \times n$ square matrix. $\mathbf{A}$ is orthgonally diagonalizable if and only if $\mathbf{A}$ is symmetric.

Suppose $\mathbf{A}$ is orthogonally diagonalizable. Write $\mathbf{A} = \mathbf{PDP}^T$. Then

$$\mathbf{A}^T = (\mathbf{PDP}^T)^T = (\mathbf{P}^T)^T\mathbf{D}^T\mathbf{P}^T = \mathbf{PDP}^T = \mathbf{A},$$

where we used the fact that diagonal matrices are symmetric.

The proof of the converse if beyond the scope of the syllabus.

Equivalent Statements for Orthogonally Diagonalizable

Theorem

Let **A** *be a square matrix of order* n. *The following statements are equivalent.*

(i) **A** *is orthogonally diagonalizable.*

(ii) *There exists an orthonormal basis* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *of* $\mathbb{R}^n$ *of eigenvectors of* **A**.

(iii) **A** *is a symmetric matrix.*

Orthogonally Diagonalization

A orthogonally diagonalizable if and only if

$$\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}, \text{ and } \mathbf{D} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$

where μ_i is the eigenvalue associated to eigenvector $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$.

Now **P** is orthogonal if and only if its columns $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.

However, in the algorithm to diagonalization, there is no guarantee that the basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ obtained is orthonormal. Does this mean that we need to apply the Gram-Schmidt process to all n vectors in the basis of $\mathbb{R}^n$ consisting of eigenvectors of **A**?

Visualization of Eigenspace

Visit the following link, https://www.geogebra.org/m/uqksb8h5, to visualize the eigenspaces.

- Here $\mathbf{A} = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix}$, from the previous example.
- The eigenvalues are $\lambda = 3, 6$.
- The blue plane is E_6, the eigenspace associated to eigenvalue $\lambda = 6$ and the red line is E_3, the eigenspace associated to eigenvalue $\lambda = 3$.
- When $c_1 = 0$, the purple vector $\mathbf{w}$ is a vector in E_6, and the green vector is $\mathbf{Aw} = 6\mathbf{w}$.
- When $c_2 = c_3 = 0$, the purple vector $\mathbf{w}$ is a vector in E_3, and the green vector is $\mathbf{Aw} = 3\mathbf{w}$.
- Observe that E_3 is orthogonal to E_6. However, $\mathbf{u}_2$ is not orthogonal to $\mathbf{u}_3$. The set $\{\mathbf{u}_2, \mathbf{v}_3\}$ is orthogonal. $\mathbf{v}_3$ was obtained via Gram-Schmidt process, see later.

Eigenspaces of a Symmetric Matrix are Orthogonal

Theorem (Eigenspaces of a symmetric matrix is orthogonal)

If $\mathbf{A}$ is a symmetric matrix, then the eigenspaces are orthogonal to each other. That is, suppose λ_1 and λ_2 are distinct eigenvalues of a symmetric matrix $\mathbf{A}$, $\lambda_1 \neq \lambda_2$, and $\mathbf{v}_i$ is an eigenvector associated to eigenvalue λ_i, for $i = 1, 2$. Then $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Proof.

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvalues associated to eigenvalues λ_1 and λ_2 of the symmetric matrix $\mathbf{A}$, respectively. Then

$$\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = (\mathbf{A}\mathbf{v}_1) \cdot \mathbf{v}_2 = (\mathbf{A}\mathbf{v}_1)^T \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{A}^T \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{A}\mathbf{v}_2 = \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2$$

$\Rightarrow (\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0 \Rightarrow \mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ since $\lambda_1 \neq \lambda_2$. $\square$

This means that, vectors belonging to different eigenspaces are orthogonal to each other.

Hence, we only need to perform Gram-Schmidt process to the eigenvectors within the same eigenspace.

Algorithm to orthogonal diagonalization

Let $\mathbf{A}$ be an order n symmetric matrix. Since $\mathbf{A}$ is symmetric, it is orthogonally diagonalizable.

1. Compute the characteristic polynomial
$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}}.$$
2. For each eigenvalue λ_i of $\mathbf{A}$, $i = 1, ..., k$, find a basis S_{λ_i} for the eigenspace, that is, find a basis S_{λ_i} for the solution space of the following linear system,
$$(\lambda_i\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}.$$
3. Apply Gram-Schmidt process to each basis S_{λ_i} of the eigenspace E_{λ_i} to obtain an orthonormal basis T_{λ_i}. Let $T = \bigcup_{i=1}^{k} T_{\lambda_i}$. Then $T = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.
4. Let $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, and $\mathbf{D} = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_n)$, where μ_i is the eigenvalue associated to $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i\mathbf{u}_i$. Then $\mathbf{P}$ is an orthogonal matrix, and
$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T.$$

6.4 Application of Diagonalization: Markov Chain

Powers of Diagonalizable Matrices

Theorem

Suppose $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. *Then* $\mathbf{A}^m = \mathbf{P}\mathbf{D}^m\mathbf{P}^{-1}$.

This can be proved by induction; the proof is left as an exercise. Note that we do not require **D** to be diagonal in the theorem.

Theorem

Let $\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$ *be a diagonal matrix. Then for any positive integer* m, $\mathbf{D}^m = \begin{pmatrix} d_1^m & 0 & \cdots & 0 \\ 0 & d_2^m & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n^m \end{pmatrix}$.

That is, the positive powers of a diagonal matrix is a diagonal matrix with entries the powers of the diagonal entries.

Powers of Diagonalizable Matrices

Corollary

Suppose $\mathbf{A}$ *is diagonalizable. Write* $\mathbf{A} = \mathbf{P} \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix} \mathbf{P}^{-1}$. *Then for any positive integer* $k > 0$,

$$\mathbf{A}^k = \mathbf{P} \begin{pmatrix} \mu_1^k & 0 & \cdots & 0 \\ 0 & \mu_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n^k \end{pmatrix} \mathbf{P}^{-1}.$$

Moreover, if $\mathbf{A}$ *is invertible, then the identity above holds for any integer* $k \in \mathbb{Z}$.

Markov Chain

Definition

(i) A vector $\mathbf{v} = (v_i)_n$ with nonnegative coordinates that add up to 1, $\sum_{i=1}^{n} v_i = 1$, is called a *probability vector*.

(ii) A *stochastic* matrix is a square matrix whose columns are probability vectors.

(iii) A *Markov chain* is a sequence of probability vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots$, together with a stochastic matrix $\mathbf{P}$ such that

$$\mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \mathbf{x}_2 = \mathbf{P}\mathbf{x}_1, \quad \ldots \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}, \ldots$$

When a Markov chain of vectors in $\mathbb{R}^n$ describes a system or a sequence of experiments, the entries in $\mathbf{x}_k$ list, respectively, the probabilities that the system is in each of n possible states, or the probabilities that the outcome of the experiment is one of n possible outcomes. For this reason, $\mathbf{x}_k$ is often called a state vector.

Observe that $\mathbf{x}_k = \mathbf{P}^k \mathbf{x}_0$.

Challenge

Definition

A *steady-state vector*, or *equilibrium vector* for a stochastic matrix $\mathbf{P}$ is a probability vector that is an eigenvector associated to eigenvalue 1.

Theorem

Let $\mathbf{P}$ *be a* $n \times n$ *stochastic matrix and*

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \dots \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}$$

be a Markov chain for some probability vector $\mathbf{x}_0$*. If the Markov chain converges, it will converge to an equilibrium vector.*

Proof.

Exercise. Hint:

(i) Show that 1 is always an eigenvalue of a stochastic matrix.

(ii) Show that if $\mathbf{v}$ is a probability vector and $\mathbf{P}$ a stochastic matrix, then $\mathbf{P}\mathbf{v}$ is also a probability vector.

(iii) Show that if the Markov chain do converge, then the state vectors will converge to an equilibrium vector.

□

Google PageRank Algorithm

- Assume a set S of sites contain key words on a topic of common interest.
- Need an algorithm to rank the sites, so that the sites with the highest rank appear first.
- In 1996, a new search engine "Google" was developed by Larry Page and Sergey Brin. This engine is based on the PageRank algorithm, which involves the use of a dominant eigenvector of some matrix.
- The underlying assumption is that more important websites are likely to receive more links from other websites.

Adjacency Matrix and Probability Transition Matrix

Suppose the set S contains n sites. We define the adjacency matrix for S for be the order n square matrix $\mathbf{A} = (a_{ij})$ where

$$a_{ij} = \begin{cases} 1 & \text{if site } j \text{ has an outgoing link to site } i; \\ 0 & \text{if site } j \text{ does not have an outgoing link to site } i. \end{cases}$$

$$\mathbf{A} = \begin{array}{c} \\ \\ \end{array} \begin{array}{cc} \begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array} & \\ \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} & \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} \end{array}$$

- s_1 references s_2, s_3 and s_4,
- s_2 references s_4 only,
- s_3 references s_1 and s_4,
- s_4 references s_1 and s_3.

Observe that

- The sum of the entries in the i-th row of is the number of incoming links to site i from other sites.
- The sum of the entries of the j-th column is the number of outgoing links on site j-th to other sites.

Adjacency Matrix and Probability Transition Matrix

From the adjacency matrix $\mathbf{A}$, we define the probability transition matrix $\mathbf{P} = (p_{ij})$ by dividing each entry of $\mathbf{A}$ by the sum of the entries in the same column; that is

$$p_{ij} = \frac{a_{ij}}{\sum_{k=1}^{n} a_{kj}}.$$

Using $\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$, we have

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{pmatrix}.$$

Observe that this is a stochastic matrix.

Discussion

- We may diagonalize the probability transition matrix $\mathbf{P}$ to obtain the outcome in the long run for the different starting state vectors.
- However, since $\mathbf{P}$ is a stochastic matrix, and the starting state vectors are probability vectors, if the Markov chain converges, it will converge to an equilibrium vector.
- Moreover, if the probability transition matrix $\mathbf{P}$ in the Google PageRank algorithm is a regular stochastic matrix (see below for definition), it will always converge to a unique equilibrium vector.

Definition

A stochastic matrix is *regular* if for some positive integer $k > 0$, the matrix power $\mathbf{P}^k$ has positive entries,

$$\mathbf{P}^k = (a_{ij})_n, \quad a_{ij} > 0 \text{ for all } i, j = 1, ..., n.$$

Algorithm to Compute Equilibrium vector

Let $\mathbf{P}$ be a $n \times n$ stochastic matrix.

1. Find an eigenvector $\mathbf{u}$ associate to eigenvalue $\lambda = 1$, that is, find a nontrivial solution to the homogeneous system $(\mathbf{I} - \mathbf{P})\mathbf{x} = \mathbf{0}$.

2. Write $\mathbf{u} = (u_i)$. Then
$$\mathbf{v} = \frac{1}{\sum_{k=1}^{n} u_k}\mathbf{u}$$
will be an equilibrium vector. Indeed, the i-th coordinate of $\mathbf{v}$ is $\frac{u_i}{\sum_{k=1}^{n} u_k}$ and hence, the sum of the coordinates of $\mathbf{v}$ is
$$\sum_{i=1}^{n} \frac{u_i}{\sum_{k=1}^{n} u_k} = \frac{\sum_{i=1}^{n} u_i}{\sum_{k=1}^{n} u_k} = 1.$$

Exercise

Let $\mathbf{P}$ be a $n \times n$ stochastic matrix. Show that $\mathbf{v}$ is an equilibrium vector of $\mathbf{P}$ if and only if it is a solution to the system

$$\begin{pmatrix} & \mathbf{P}-\mathbf{I}_n & & \\ 1 & 1 & \cdots & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Here $\begin{pmatrix} & \mathbf{P}-\mathbf{I}_n & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}$ is the $(n+1) \times n$ matrix whose first n rows are the matrix $\mathbf{P}-\mathbf{I}_n$, and the last row has all entries 1.

6.5 Application of Orthogonal Diagonalization: Singular Value Decomposition

Introduction

All non-square matrices are non-diagonalizable, much less orthogonally diagonalizable. However, we can still factorize any $m \times n$ $\mathbf{A}$ into

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T,$$

where $\mathbf{U}$ is an order m orthogonal matrix, $\mathbf{V}$ an order n orthogonal matrix, and the matrix $\mathbf{\Sigma}$ has the form

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{D} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix},$$

for some diagonal matrix $\mathbf{D}$ of order r, where $r \leq \min\{m, n\}$.

Singular Values

Let $\mathbf{A}$ be a $m \times n$ matrix. Then since $\mathbf{A}^T\mathbf{A}$ is an order n symmetric matrix, we may orthogonally diagonalize it. Let $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}^T\mathbf{A}$.

Let μ_i be the eigenvalue associated to $\mathbf{v}_i$, for $i = 1, ..., n$, not necessarily distinct.

Lemma

The eigenvalue μ_i of $\mathbf{A}^T\mathbf{A}$ is nonnegative.

Proof.

$$\|\mathbf{A}\mathbf{v}_i\|^2 = (\mathbf{A}\mathbf{v}_i)^T(\mathbf{A}\mathbf{v}_i) = \mathbf{v}_i^T\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \mu_i\mathbf{v}_i^T\mathbf{v}_i = \mu_i,$$

where third equality follows from the fact that μ_i is an eigenvalue of $\mathbf{A}^T\mathbf{A}$, and the forth equality follows from the fact that $\mathbf{v}_i$ is a unit vector. $\square$

Singular Values

Reordering if necessary, we may assume that

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n \geq 0.$$

The singular values of $\mathbf{A}$ are

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0,$$

where $\sigma_i = \sqrt{\mu_i}$, i=1,...,n, is the square root of the eigenvalues of $\mathbf{A}^T\mathbf{A}$, arranged in decreasing order. Let r be the largest integer such that $1 \leq r \leq n$ and $\sigma_i > 0$ for all $i \leq r$, that is

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = 0 = \cdots = \sigma_m = 0$$

Define the matrix $m \times n$ matrix $\mathbf{\Sigma}$ to be

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{D} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}, \text{ where } \mathbf{D} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix}.$$

Exercise

1. Show that
$$\mathbf{\Sigma}^T\mathbf{\Sigma} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_r & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \mu_n \end{pmatrix},$$
where μ_i, $i = 1, ..., n$, is the eigenvalues of $\mathbf{A}^T\mathbf{A}$; that is $\mathbf{A}^T\mathbf{A} = \mathbf{P}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{P}^T$, where $\mathbf{P} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix}$ and $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}^T\mathbf{A}$.

2. Show that $\mathbf{A}\mathbf{v}_i \neq \mathbf{0}$ for all $i \leq r$ and $\mathbf{A}\mathbf{v}_i = \mathbf{0}$ for all $i > r$.

Singular Value Decomposition

Suppose $\mathbf{A}$ is a $m \times n$ matrix. Let $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}^T\mathbf{A}$. Let

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

be the nonzero singular values of $\mathbf{A}$. Define

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i, \quad i = 1, .., r.$$

Lemma

$\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ *is an orthonormal basis for the column space of* $\mathbf{A}$, *and* $\mathrm{rank}(\mathbf{A}) = r$.

Proof.

By construction, $\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ is an orthonormal set,

$$\mathbf{u}_i \cdot \mathbf{u}_j = \frac{1}{\sigma_i\sigma_j}(\mathbf{A}\mathbf{v}_i)^T(\mathbf{A}\mathbf{v}_j) = \frac{1}{\sigma_i\sigma_j}\mathbf{v}_i^T\mathbf{A}^T\mathbf{A}\mathbf{v_j} = \frac{\mu_j}{\sigma_i\sigma_j}\mathbf{v}_i \cdot \mathbf{v}_j = \begin{cases} 0 & \text{if } i \neq j \\ \frac{\mu_i}{\sigma_i\sigma_i} = 1 & \text{if } i = j \end{cases}$$

since $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ is an orthonormal set.

Singular Value Decomposition

Continue of proof.

Recall that $\text{Col}(\mathbf{A}) = \{\mathbf{Av} \mid \mathbf{v} \in \mathbb{R}^n\}$. Hence, by construction, $\mathbf{u}_i \in \text{Col}(\mathbf{A})$, and hence $\text{span}\{\mathbf{u}_1, ..., \mathbf{u}_r\} \subseteq \text{Col}(\mathbf{A})$. Now given any $\mathbf{v} \in \mathbb{R}^n$, write $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$. Then

$$\begin{aligned} \mathbf{Av} &= c_1\mathbf{Av}_1 + c_2\mathbf{Av}_2 + \cdots + c_n\mathbf{Av}_n \\ &= \sigma_1 c_1 \left(\frac{1}{\sigma_1}\mathbf{Av}_1\right) + \sigma_2 c_2 \left(\frac{1}{\sigma_2}\mathbf{Av}_2\right) + \cdots + \sigma_r c_r \left(\frac{1}{\sigma_r}\mathbf{Av}_r\right) + \mathbf{0} + \cdots + \mathbf{0} \\ &= \sigma_1 c_1 \mathbf{u}_1 + \sigma_2 c_2 \mathbf{u}_2 + \cdots \sigma_r c_r \mathbf{u}_r \end{aligned}$$

This shows that $\text{Col}(\mathbf{A}) \subseteq \text{span}\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ too, therefore there are equal. Hence, $\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ is an orthonormal basis for the column space of $\mathbf{A}$, which proves that $\text{rank}(\mathbf{A}) = r$. $\square$

Singular Value Decomposition

Using the notations from above, extend $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ for $\mathbb{R}^m$ (if $r \neq m$). Define

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{pmatrix},$$

it is an order m orthogonal matrix. Define

$$\mathbf{V} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix},$$

then $\mathbf{V}$ is an order n orthogonal matrix. Let $\mathbf{\Sigma}$ be the matrix defined by the nonzero singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$. Then

$$\mathbf{A} = \mathbf{U\Sigma V}^T.$$

Proof.

Since $\mathbf{V}$ is orthogonal, suffice to show that $\mathbf{AV} = \mathbf{U\Sigma}$, but by construction,

$$\mathbf{AV} = \begin{pmatrix} \mathbf{Av}_1 & \cdots & \mathbf{Av}_r & \mathbf{Av}_{r+1} & \cdots & \mathbf{Av}_n \end{pmatrix} = \begin{pmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0} & \cdots & 0 \end{pmatrix} = \mathbf{U\Sigma}.$$

$\square$

Algorithm to Singular Value Decomposition

Let $\mathbf{A}$ be a $m \times n$ matrix with $\text{rank}(\mathbf{A}) = r$.

1. Find the eigenvalues of $\mathbf{A}^T\mathbf{A}$. Arrange the nonzero eigenvalues in descending order (counting multiplicity)

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_r > 0 = \mu_{r+1} = \cdots = \mu_n,$$

   and let $\sigma_i = \sqrt{\mu_i}$, $i = 1, ..., r$. Set

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{D} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}, \text{ where } \mathbf{D} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix}.$$

2. Find an orthogonal basis for each eigenspace, and let $\mathbf{v}_i$ be the unit vector associated to μ_i. Set

$$\mathbf{V} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix}.$$

3. Let $\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$ for $i = 1, ..., r$. Extend $\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, ..., \mathbf{u}_r, \mathbf{u}_{r+1}, ..., \mathbf{u}_m\}$ of $\mathbb{R}^m$, that is, solve for $\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r \end{pmatrix}^T \mathbf{x} = \mathbf{0}$ and find an orthonormal basis for the solution space. Let

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$$

Challenge

Let $\mathbf{A}$ be a $m \times n$ matrix. Prove the following statements.

1. $\text{rank}(\mathbf{A}) = n$ if and only if all the singular values of $\mathbf{A}$ are positive.

2. $\text{rank}(\mathbf{A}) = m$ if and only if all the singular values of $\mathbf{A}^T$ are positive.

Appendix

Independence of Eigenspaces

Theorem

Let $\mathbf{A}$ *be a* $n \times n$ *matrix. Let* λ_1 *and* λ_2 *are distinct eigenvalues of* $\mathbf{A}$, $\lambda_1 \neq \lambda_2$. *Suppose* $\{\mathbf{u}_1, ..., \mathbf{u}_k\}$ *is a linearly independent subset of eigenspace associated to eigenvalue* λ_1, *and* $\{\mathbf{v}_1, ..., \mathbf{v}_m\}$ *is a linearly independent subset of of eigenspace associated to eigenvalue* λ_2. *Then* $\{\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{v}_1, ..., \mathbf{v}_m\}$ *is linearly independent.*

Sketch of Proof.

Suppose $c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m = \mathbf{0}$. Multiply both sides of the equation by $\mathbf{A}$, λ_1, and λ_2, respectively, we have

$$\begin{aligned}
\mathbf{0} &= \mathbf{A}(c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m) = c_1\lambda_1\mathbf{u}_1 + \cdots + c_k\lambda_1\mathbf{u}_k + d_1\lambda_2\mathbf{v}_1 + \cdots + d_m\lambda_2\mathbf{v}_m \quad (1)\\
\mathbf{0} &= \lambda_1(c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m) = c_1\lambda_1\mathbf{u}_1 + \cdots + c_k\lambda_1\mathbf{u}_k + d_1\lambda_1\mathbf{v}_1 + \cdots + d_m\lambda_1\mathbf{v}_m \quad (2)\\
\mathbf{0} &= \lambda_2(c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m) = c_1\lambda_2\mathbf{u}_1 + \cdots + c_k\lambda_2\mathbf{u}_k + d_1\lambda_2\mathbf{v}_1 + \cdots + d_m\lambda_2\mathbf{v}_m \quad (3)
\end{aligned}$$

Take equation(1) - equation(2), we have

$$\mathbf{0} = (\lambda_2 - \lambda_1)(d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m).$$

Since $(\lambda_2 - \lambda_1) \neq 0$, we can conclude that $d_1 = \cdots = d_m = 0$. Take equation(1)-equation(3), we too can conclude that $c_1 = \cdots = c_k = 0$. $\square$

Similar Matrices

Definition

Two square matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be *similar* if there exists an invertible matrix $\mathbf{P}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}.$$

Example

Diagonalizable matrices are similar to diagonal matrices.

Similar Matrices

Lemma

Suppose $\mathbf{A}$ *and* $\mathbf{B}$ *are similar matrices, then they have the same characteristic polynomial,*

$$\det(x\mathbf{I} - \mathbf{A}) = \det(x\mathbf{I} - \mathbf{B}).$$

Proof.

Let $\mathbf{P}$ be such that $\mathbf{A} = \mathbf{PBP}^{-1}$. Then

$$\begin{aligned}
\det(x\mathbf{I} - \mathbf{B}) &= \det(x\mathbf{I} - \mathbf{B})\det(\mathbf{P})\det(\mathbf{P})^{-1} \\
&= \det(\mathbf{P})\det(x\mathbf{I} - \mathbf{B})\det(\mathbf{P}^{-1}) \\
&= \det(\mathbf{P}(x\mathbf{I} - \mathbf{B})\mathbf{P}^{-1}) \\
&= \det(\mathbf{P}x\mathbf{I}\mathbf{P}^{-1} - \mathbf{PBP}^{-1}) \\
&= \det(x\mathbf{I} - \mathbf{A})
\end{aligned}$$

$\square$

Geometric multiplicity is no greater than algebraic multiplicity

Theorem (Geometric multiplicity is no greater than algebraic multiplicity)

The geometric multiplicity of an eigenvalue λ of a square matrix $\mathbf{A}$ is no greater than the algebraic multiplicity, that is,

$$1 \leq \dim(E_\lambda) \leq r_\lambda.$$

Proof.

Let $\mathbf{A}$ be a $n \times n$ matrix. Let λ be an eigenvalue of $\mathbf{A}$ and E_λ be the associated eigenspace. Suppose $\dim(E_\lambda) = k$. Let $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for the eigenspace E_λ. Extend this set to be a basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}_{k+1}, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$. Let $\mathbf{Q} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, it is an invertbile matrix. Note that

$$\begin{aligned} \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{pmatrix} &= \mathbf{I} = \mathbf{Q}^{-1}\mathbf{Q} = \mathbf{Q}^{-1}\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Q}^{-1}\mathbf{u}_1 & \mathbf{Q}^{-1}\mathbf{u}_2 & \cdots & \mathbf{Q}^{-1}\mathbf{u}_n \end{pmatrix}, \end{aligned}$$

that is, $\mathbf{Q}^{-1}\mathbf{u}_i = \mathbf{e}_i$ for all $i = 1, ..., n$. Let $\mathbf{B} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$.

Geometric multiplicity is no greater than algebraic multiplicity

Continue of Proof.
Then

$$\begin{aligned}
\mathbf{B} &= \mathbf{Q}^{-1}\mathbf{A}\begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{pmatrix} = \mathbf{Q}^{-1}\begin{pmatrix}\mathbf{A}\mathbf{u}_1 & \mathbf{A}\mathbf{u}_2 & \cdots & \mathbf{A}\mathbf{u}_n\end{pmatrix} \\
&= \mathbf{Q}^{-1}\begin{pmatrix}\lambda\mathbf{u}_1 & \cdots & \lambda\mathbf{u}_k & \mathbf{A}\mathbf{u}_{k+1}\cdots & \mathbf{A}\mathbf{u}_n\end{pmatrix} \\
&= \begin{pmatrix}\lambda\mathbf{Q}^{-1}\mathbf{u}_1 & \cdots & \lambda\mathbf{Q}^{-1}\mathbf{u}_k & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_{k+1}\cdots & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_n\end{pmatrix} \\
&= \begin{pmatrix}
\begin{matrix}\lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0\end{matrix} & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_{k+1}\cdots & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_n
\end{pmatrix}.
\end{aligned}$$

Geometric multiplicity is no greater than algebraic multiplicity

Continue of Proof.

This means that $\det(x\mathbf{I} - \mathbf{B}) = (x - \lambda)^k p(x)$ for some polynomial $p(x)$. By since $\mathbf{A}$ and $\mathbf{B}$ are similar matrices,

$$\det(x\mathbf{I} - \mathbf{A}) = \det(x\mathbf{I} - \mathbf{B}) = (x - \lambda)^k p(x).$$

This means that the algebraic multiplicity of the eigenvalue λ of $\mathbf{A}$ is no less than k, that is,

$$r_\lambda \geq k = \dim(E_\lambda).$$

$\square$

Regular Stochastic Matrix

Definition

A stochastic matrix is *regular* if for some positive integer $k > 0$, the matrix power $\mathbf{P}^k$ has positive entries,

$$\mathbf{P}^k = (a_{ij})_n, \quad a_{ij} > 0 \text{ for all } i, j = 1, ..., n.$$

Lemma

Let $\mathbf{A} = (a_{ij})_n$ *be a* $n \times n$ *stochastic matrix with positive entries,* $a_{ij} > 0$ *for all* $i, j = 1, ..., n$*. Then geometric multiplicity of eigenvalue* 1 *is* 1*,* $\dim(E_1) = 1$*.*

Proof.

Write $\mathbf{A} = (a_{ij})_n$. We will show that the geometric multiplicity of $\lambda = 1$ as an eigenvalue of $\mathbf{A}^T$ is 1. Let $\mathbf{x} = (x_i)$ be an eigenvector of $\mathbf{A}^T$ associated to eigenvalue 1. By taking a multiple of $\mathbf{x}$ if necessary, we may assume that $\mathbf{x}$ has some coordinates with positive entries. Let $1 \leq m \leq n$ be the coordinate of $\mathbf{x}$ such that x_m is the largest, $x_m \geq x_i$ for all $i = 1, ..., n$. Now comparing the m-th coordinate of $\mathbf{A}^T\mathbf{x} = \mathbf{x}$, we have

$$a_{1m}x_1 + a_{2m}x_2 + \cdots + a_{nm}x_n = x_m.$$

Regular Stochastic Matrix

Continue of Proof.

Note that necessarily $x_m \neq 0$ and hence, dividing by x_m, we have

$$a_{1m}\frac{x_1}{x_m} + a_{2m}\frac{x_2}{x_m} + \cdots + a_{nm}\frac{x_n}{x_m} = 1.$$

Note that $\frac{x_i}{x_m} \leq 1$ for all $i = 1, ..., n$ and so $a_{im}\frac{x_i}{x_m} \leq a_{im}$. Suppose $x_m > x_j$ for some $j = 1, ..., n$. Then since $a_{jm} > 0$, $a_{jm}\frac{x_j}{x_m} < a_{jm}$, and thus

$$1 = a_{1m}\frac{x_1}{x_m} + a_{2m}\frac{x_2}{x_m} + \cdots + a_{jm}\frac{x_j}{x_m} + \cdots + a_{nm}\frac{x_n}{x_m} < a_{1m} + a_{2m} + \cdots + a_{jm} + \cdots + a_{nm} = 1,$$

a contradiction. Hence, $x_i = x_m$ for all $i = 1, ..., n$, that is $\mathbf{x} = \alpha \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ for some real number α, which shows that the geometric multiplicity of eigenvalue 1 as an eigenvalue of $\mathbf{A}^T$ is 1. The result therefore follows from the fact that the geometric multiplicity of 1 as an eigenvalue of $\mathbf{A}^T$ is equal to its geometric multiplicity as an eigenvalue of $\mathbf{A}$. $\square$

Regular Stochastic Matrix

Lemma

Let $\mathbf{A}$ *be a* $n \times n$ *square matrix and* $\mathbf{v}$ *an eigenvector of* $\mathbf{A}$ *associated to eigenvalue* λ*. Then for any positive integer* k*,* $\mathbf{v}$ *is an eigenvector of* $\mathbf{A}^k$ *associated to eigenvalue* λ^k*.*

The proof is left as an exercise. Together with the previous lemma, this shows that if $\mathbf{P}$ is a regular stochastic matrix, then the geometric multiplicity of eigenvalue 1 is 1.

Lemma

Suppose $\mathbf{P}$ *is a regular stochastic matrix. Then for any probability vector* $\mathbf{x}_0$ *the Markov chain* $\{\mathbf{x}_0, \mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, ..., \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k+1}\}$ *will converge.*

The proof of the lemma regulars knowledge of Jordan block form, which is beyond the syllabus of the course.

Regular Stochastic Matrix

Theorem

If $\mathbf{P}$ is an $n \times n$ regular stochastic matrix, then $\mathbf{P}$ has a unique equilibrium vector. Moreover, if $\mathbf{x}_0$ is any probability vector and $\mathbf{x}_{k+1} = \mathbf{P}\mathbf{x}_k$ for $k = 0, 1, \ldots$, then the Markov chain $\{\mathbf{x}_k\}$ converges to the unique equilibrium vector.

Proof.

Since $\mathbf{P}$ is a regular stochastic matrix, the geometric multiplicity of the eigenvalue 1 is 1, and thus the equilibrium vector is unique. Also, since the Markov chain will converge, it will converge to the unique equilibrium vector. $\square$

MA1522: Linear Algebra for Computing

Chapter 7: Linear Transformation

7.1 Introduction to Linear Transformation

Geometric Interpretation of Matrix Multiplication

Given a $m \times n$ matrix $\mathbf{A}$, we can think of it as mapping vectors $\mathbf{v}$ from $\mathbb{R}^n$ to a vector $\mathbf{Av}$ in $\mathbb{R}^m$.

Example

1. Consider the matrix $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$. It maps $\mathbb{R}^2$ to the plane in $\mathbb{R}^3$ defined by $z = 0$, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$.

2. The matrix $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ projects vectors in $\mathbb{R}^3$ onto the $z = 0$ plane and identifies it with $\mathbb{R}^2$, $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$. That is, it can be interpreted as the map $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix}$.

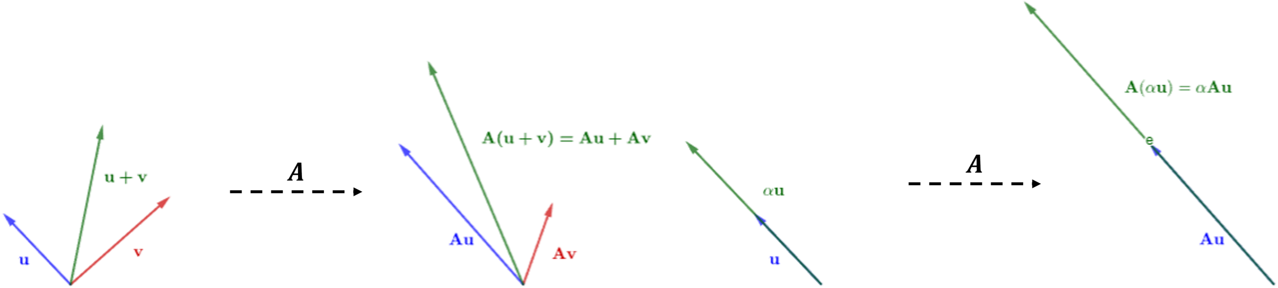
3. The zero matrix $\mathbf{A} = \mathbf{0}_{m \times n}$ sends any vector in $\mathbb{R}^n$ to the zero vector $\mathbf{0}$ in $\mathbb{R}^m$,

$$\mathbf{Av} = \mathbf{0} \quad \text{for all } \mathbf{v} \text{ in } \mathbb{R}^n.$$

Geometric Interpretation of Matrix Multiplication

Recall that matrix multiplication commutes with scalar multiplication, and is distributive, for all vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ and scalar α,

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



Or equivalently, matrix multiplication is linear, for all vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ and scalars α, β,

$$\mathbf{A}(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha\mathbf{A}\mathbf{u} + \beta\mathbf{A}\mathbf{v}.$$

Geometrically, this means that the mapping of a linear combination is the linear combination of the mapping.

Linear Transformation

Definition

A mapping (function) $T : \mathbb{R}^n \to \mathbb{R}^m$, is a *linear transformation* if for all vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$, and scalars α, β,

$$T(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v}).$$

The Euclidean space $\mathbb{R}^n$ is called the *domain* of the mapping, and the Euclidean space $\mathbb{R}^m$ is called the *codomain* of the mapping.

Remarks

Equivalently, a mapping $T : \mathbb{R}^n \to \mathbb{R}^m$, is a linear transformation if it satisfies the following properties.

(i) For any vector $\mathbf{u}$ in $\mathbb{R}^n$ and scalar α,

$$T(\alpha \mathbf{u}) = \alpha T(\mathbf{u}).$$

(ii) For any vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$,

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}).$$

By induction, we have that for any vectors $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k$ in $\mathbb{R}^n$ and scalars $c_1, c_2, ..., c_k$,

$$T(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k) = c_1 T(\mathbf{u}_1) + c_2 T(\mathbf{u}_2) + \cdots + c_k T(\mathbf{u}_k).$$

The previous discussion shows that every matrix defines a linear transformation by multiplication,

$$\mathbf{A} \mapsto T_{\mathbf{A}}; \quad T_{\mathbf{A}}(\mathbf{u}) = \mathbf{A}\mathbf{u} \text{ for all } \mathbf{u} \text{ in } \mathbb{R}^n.$$

It will be shown later that this identification is one-to-one and onto, that is, every linear transformation is defined by multiplication of some matrix.

Not a Linear Transformation

Observe that by linearity, a linear transformation must map the zero vector $\mathbf{0}_n$ in $\mathbb{R}^n$ to the zero vector $\mathbf{0}_m$ in $\mathbb{R}^m$, $T(\mathbf{0}_n)) = \mathbf{0}_m$. Hence, together with equivalent definition of linear transformation, we have the following.

A mapping $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is not a linear transformation if any of the following statements hold.

(i) $\mathbf{T}$ does not map the zero vector to the zero vector, $\mathbf{T}(\mathbf{0}) \neq \mathbf{0}$.

(ii) There is a scalar α and a vector $\mathbf{u}$ in $\mathbb{R}^n$ such that $\mathbf{T}(\alpha\mathbf{u}) \neq \alpha\mathbf{T}(\mathbf{u})$.

(iii) There are vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ such that $\mathbf{T}(\mathbf{u} + \mathbf{v}) \neq \mathbf{T}(\mathbf{u}) + \mathbf{T}(\mathbf{v})$.

Challenge

Find a mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$T(\alpha \mathbf{u}) = \alpha T(\mathbf{u})$$

for all scalar α and vector $\mathbf{u}$ in $\mathbb{R}^n$, but is not a linear transformation.

Standard Matrix

Theorem

A mapping $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^m$ *is a linear transformation if and only if there is a unique* $m \times n$ *matrix* $\mathbf{A}$ *such that*

$$T(\mathbf{u}) = \mathbf{A}\mathbf{u} \quad \text{for all vectors } \mathbf{u} \text{ in } \mathbb{R}^n.$$

The matrix $\mathbf{A}$ *is given by*

$$\mathbf{A} = \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{pmatrix},$$

where $E = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ *is the standard basis for* $\mathbb{R}^n$*. That is, the* i*-th column of* $\mathbf{A}$ *is* $T(\mathbf{e}_i)$*, for* $i = 1, ..., n$*.*

Proof.

We have shown that a $m \times n$ matrix $\mathbf{A}$ defines a linear transformation by matrix multiplication.

Conversely, suppose $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. For any $\mathbf{u} = (u_i)$ in $\mathbb{R}^n$, write $\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + \cdots + u_n\mathbf{e}_n$. By linearity,

$$T(\mathbf{u}) = u_1 T(\mathbf{e}_1) + u_2 T(\mathbf{e}_2) + \cdots + u_n T(\mathbf{e}_n) = \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \mathbf{A}\mathbf{u}.$$

The uniqueness of $\mathbf{A}$ is left as an exercise.

Standard Matrix

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The unique $m \times n$ matrix $\mathbf{A}$ such that

$$T(\mathbf{u}) = \mathbf{A}\mathbf{u} \quad \text{for all } \mathbf{u} \text{ in } \mathbb{R}^n$$

is called the *standard matrix*, or *matrix representation* of T.

Question

1. Is the mapping $T : \mathbb{R}^3 \to \mathbb{R}^3$,
$$T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}\right)\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$
a linear transformation? If it is, find its standard matrix.

2. Is $T\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ for some constants $a_1, a_2, ..., a_n$ a linear transformation? If it is, find its standard matrix.

3. What is the standard matrix of the following linear transformation $\mathbf{T}\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$?

Representation of Linear Transformation with Respect to a Basis

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ be a basis for $\mathbb{R}^n$. The *representation of T with respect to basis S*, denoted as $[T]_S$, is defined to be the $m \times n$ matrix

$$[T]_S = \begin{pmatrix} T(\mathbf{u}_1) & T(\mathbf{u}_2) & \cdots & T(\mathbf{u}_n) \end{pmatrix}.$$

The standard matrix or matrix representation of T is the representation of T with respect to the standard matrix,

$$\mathbf{A} = [T]_E.$$

Representation of Linear Transformation with Respect to a Basis

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ be a basis for $\mathbb{R}^n$. Then for any vector $\mathbf{v}$ in $\mathbb{R}^n$,

$$T(\mathbf{v}) = [T]_S[\mathbf{v}]_S,$$

that is, the image $T(\mathbf{v})$ is the product of the representation of T with respect to basis S with the coordinates $\mathbf{v}$ with respect to basis S. Moreover, if $\mathbf{P}$ is the transition matrix from the standard basis E of $\mathbb{R}^n$ to basis S, then the standard matrix $\mathbf{A}$ of T is given by

$$\mathbf{A} = [T]_S\mathbf{P}.$$

This means that we are able to compute the standard matrix of T if we know the image of T on a basis of $\mathbb{R}^n$, and thus from $\mathbf{A}$, we are able to reconstruct the formula for T. In fact, this is a equivalence statement; that is, we can reconstruct the formula for T if and only if we have the image of T on a basis.

Representation of Linear Transformation with Respect to a Basis

Proof.

Given any vector $\mathbf{v}$ in $\mathbb{R}^n$, write $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n$. Then $[\mathbf{v}]_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$ and by linearity,

$$T(\mathbf{v}) = c_1 T(\mathbf{u}_1) + c_2 T(\mathbf{u}_2) + \cdots + c_n T(\mathbf{u}_n) = \begin{pmatrix} T(\mathbf{u}_1) & T(\mathbf{u}_2) & \cdots & T(\mathbf{u}_n) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = [T]_S[\mathbf{v}]_S.$$

Next, by definition of $\mathbf{P}$, $\mathbf{Pv} = [v]_S$. Hence, for any vector $\mathbf{v}$ in $\mathbb{R}^n$,

$$\mathbf{Av} = T(\mathbf{v}) = [T]_S[\mathbf{v}]_S = [T]_S\mathbf{Pv}.$$

Since this is true for any $\mathbf{v}$ in $\mathbb{R}^n$, we have the identity $\mathbf{A} = [T]_S\mathbf{P}$. $\square$

7.2 Range and Kernel of Linear Transformation

Range of Linear Transformation

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The *range* of T is

$$R(T) = T(\mathbb{R}^n) = \{\ \mathbf{v} \in \mathbb{R}^m \mid \mathbf{v} = T(\mathbf{u}) \text{ for some } \mathbf{u} \in \mathbb{R}^n\ \}.$$

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The range of T is a subspace.

Let $\mathbf{A}$ be the standard matrix of T. Recall that $T(\mathbf{u}) = \mathbf{A}\mathbf{u}$ for all $\mathbf{u}$ in $\mathbb{R}^n$. Hence,

$$R(T) = \{\ \mathbf{v} = T(\mathbf{u}) \mid \mathbf{u} \in \mathbb{R}^n\ \} = \{\ \mathbf{v} = \mathbf{A}\mathbf{u} \mid \mathbf{u} \in \mathbb{R}^n\ \} = \mathrm{Col}(\mathbf{A}).$$

That is, the range of T is the column space of its standard matrix, and therefore is a subspace of the codomain $\mathbb{R}^m$. The abstract proof can be found in the appendix.

Range of Linear Transformation

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The *rank* of T is the dimension of the range of T

$$\text{rank}(T) = \dim(\text{R}(T)).$$

Let $\mathbf{A}$ be the standard matrix of T. Since the range of T is the column space of $\mathbf{A}$, $\text{R}(T) = \text{Col}(\mathbf{A})$, therefore

$$\text{rank}(T) = \dim(\text{R}(T)) = \dim(\text{Col}(\mathbf{A})) = \text{rank}(\mathbf{A}).$$

Kernel of Linear Transformation

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The set of all vectors in $\mathbb{R}^n$ that maps to the zero vector $\mathbf{0}$ by T is called the *kernel* of T, and is denoted as

$$\ker(T) = \{\ \mathbf{u} \in \mathbb{R}^n \mid T(\mathbf{u}) = \mathbf{0}\ \}.$$

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The kernel of T is a subspace.

Let $\mathbf{A}$ be the standard matrix of T. Then

$$\ker(T) = \{\ \mathbf{u} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{u} = T(\mathbf{u}) = \mathbf{0}\ \} = \text{Null}(\mathbf{A}).$$

That is, the kernel of T is the nullspace of its standard matrix, and is thus a subspace.

Kernel of Linear Transformation

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The *nullity* of T is the dimension of the kernel of T,

$$\text{nullity}(T) = \dim(\ker(T)).$$

Let $\mathbf{A}$ be the standard matrix of T. Then

$$\text{nullity}(T) = \dim(\ker(T)) = \dim(\text{Null}(\mathbf{A})) = \text{nullity}(\mathbf{A}).$$

Injectivity of Linear Transformation

Definition

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is *injective*, or *one-to-one* if for every vector $\mathbf{v}$ in the range of T, $\mathbf{v} \in \mathrm{R}(T)$, there is a unique $\mathbf{u}$ in $\mathbb{R}^n$ such that $T(\mathbf{u}) = \mathbf{v}$.

Alternatively, T is injective if whenever $T(\mathbf{u}_1) = T(\mathbf{u}_2)$, then $\mathbf{u}_1 = \mathbf{u}_2$.

Theorem

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is injective if and only if the kernel is trivial, $\ker(T) = \{\mathbf{0}\}$.

Let $\mathbf{A}$ be the standard matrix of T. Then recall that since the general solution to the consistent system $\mathbf{Ax} = \mathbf{v}$ is a particular solution plus the general solution to the homogeneous system $\mathbf{Ax} = \mathbf{0}$, $\mathbf{Ax} = \mathbf{v}$ has a unique solution if and only if $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution. Hence, T is injective if and only if $\mathbf{Ax} = \mathbf{v}$ has a unique solution for every $\mathbf{v}$ in $\mathrm{R}(T) = \mathrm{Col}(\mathbf{A})$, if and only if $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution, or $\ker(T) = \mathrm{Null}(\mathbf{A}) = \{\mathbf{0}\}$. The abstract proof is given in the appendix. We will add this to the equivalent statements for full rank matrices.

Full Rank Equals Number of Columns

Theorem

Suppose $\mathbf{A}$ *is a* $m \times n$ *matrix. The following statements are equivalent.*

(i) $\mathbf{A}$ *is full rank, where the rank is equal to the number of columns,* $\text{rank}(\mathbf{A}) = n$.

(ii) *The rows of* $\mathbf{A}$ *spans* $\mathbb{R}^n$, *Row*$(\mathbf{A}) = \mathbb{R}^n$.

(iii) *The columns of* $\mathbf{A}$ *are linearly independent.*

(iv) *The homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution, that is,* $\text{Null}(\mathbf{A}) = \{\mathbf{0}\}$.

(v) $\mathbf{A}^T\mathbf{A}$ *is an invertible matrix of order* n.

(vi) $\mathbf{A}$ *has a left inverse.*

(vii) *The linear transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ *defined by* $\mathbf{A}$ *is injective.*

Exercise

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that if T is injective, then necessary $n \leq m$.

Surjectivity of Linear Transformation

Definition

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is *surjective* or *onto* if for every $\mathbf{v}$ in the codomain $\mathbb{R}^m$, there exists a $\mathbf{u}$ in the domain $\mathbb{R}^n$ such that $T(\mathbf{u}) = \mathbf{v}$.

Alternatively, T is surjective if the range is the codmain, $R(T) = \mathbb{R}^m$, which is equivalent to $\text{rank}(T) = m$. This means that if $\mathbf{A}$ is the standard matrix of T, then $\mathbf{A}$ is full rank, where the rank is equal to its number of rows. We will add this to the equivalent statements for full rank matrices.

Full Rank Equals Number of Rows

Theorem

Suppose $\mathbf{A}$ *is a* $m \times n$ *matrix. The following statements are equivalent.*

(i) $\mathbf{A}$ *is full rank, where the rank is equal to the number of rows,* $\text{rank}(\mathbf{A}) = m$.

(ii) *The columns of* $\mathbf{A}$ *spans* $\mathbb{R}^m$, $Col(\mathbf{A}) = \mathbb{R}^m$.

(iii) *The rows of* $\mathbf{A}$ *are linearly independent.*

(iv) *The linear system* $\mathbf{Ax} = \mathbf{b}$ *is consistent for every* $\mathbf{b} \in \mathbb{R}^m$.

(v) $\mathbf{AA}^T$ *is an invertible matrix of order* m.

(vi) $\mathbf{A}$ *has a right inverse.*

(vii) *The linear transformation* T *defined by* $\mathbf{A}$ *is surjective.*

Exercise

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that if T is surjective, then necessary $n \geq m$.

Equivalent Statements of Invertibility

Theorem (Equivalent Statements for Invertibility)

Let $\mathbf{A}$ be a square matrix of order n. The following statements are equivalent.

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}^T$ *is invertible.*

(iii) *(left inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{BA} = \mathbf{I}$.

(iv) *(right inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{AB} = \mathbf{I}$.

(v) *The reduced row-echelon form of* $\mathbf{A}$ *is the identity matrix.*

(vi) $\mathbf{A}$ *can be expressed as a product of elementary matrices.*

(vii) *The homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution.*

(viii) *For any* $\mathbf{b}$, *the system* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution.*

(ix) *The determinant of* $\mathbf{A}$ *is nonzero,* $\det(\mathbf{A}) \neq 0$.

(x) *The columns/rows of* $\mathbf{A}$ *are linearly independent.*

(xi) *The columns/rows of* $\mathbf{A}$ *spans* $\mathbb{R}^n$.

(xii) $\text{rank}(\mathbf{A}) = n$ *(*$\mathbf{A}$ *has full rank).*

(xiii) $\text{nullity}(\mathbf{A}) = 0$.

(xiv) 0 *is not an eigenvalue of* $\mathbf{A}$.

(xv) *The linear transformation* T *defined by* $\mathbf{A}$ *is injective.*

(xvi) *The linear transformation* T *defined by* $\mathbf{A}$ *is surjective.*

Exercise

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is bijective if it is both injective and surjective.

Show that $T : \mathbb{R}^n \to \mathbb{R}^n$ is bijective if and only if there is a linear transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$T(S(\mathbf{x})) = \mathbf{x} \quad \text{and} \quad S(T(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

Appendix

Range of Linear Transformation is a Subspace

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The range of T is a subspace.

Proof.

(i) Since $T(\mathbf{0}) = \mathbf{0}$, the range of T contains the zero vector, $\mathbf{0} \in \mathrm{R}(T)$.

(ii) Suppose now $\mathbf{v}_1, \mathbf{v}_2$ are in the range of T. This means that there are some $\mathbf{u}_1, \mathbf{u}_2$ in $\mathbb{R}^n$ such that

$$T(\mathbf{u}_1) = \mathbf{v}_1 \quad \text{and} \quad T(\mathbf{u}_2) = \mathbf{v}_2.$$

Therefore, for any scalars α, β,

$$T(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2) = \alpha T(\mathbf{u}_1) + \beta T(\mathbf{u}_2) = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2,$$

which shows that $\alpha\mathbf{v}_1 + \beta\mathbf{v}_2$ is also in the range of T.

$\square$

Kernel of Linear Transformation is a Subspace

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The kernel of T is a subspace.

Proof.

(i) Since $T(\mathbf{0}) = \mathbf{0}$, it is clear that the zero vector is in the kernel of T, $\mathbf{0} \in \ker(T)$.

(ii) Suppose now $\mathbf{u}_1, \mathbf{u}_2$ are in the kernel of T, $T(\mathbf{u}_i) = \mathbf{0}$ for $i = 1, 2$. Then for any scalars α, β,

$$T(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2) = \alpha T(\mathbf{u}_1) + \beta T(\mathbf{u}_2) = \alpha\mathbf{0} + \beta\mathbf{0} = \mathbf{0},$$

which shows that $\alpha\mathbf{u}_1 + \beta\mathbf{u}_2$ is in the kernel of T too.

$\square$

Injectivity of Linear Transformation

Theorem

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is injective if and only if the kernel is trivial, $\ker(T) = \{\mathbf{0}\}$.

Proof.

Suppose T is injective. Then for any $\mathbf{u}$ in the kernel of T, $T(\mathbf{u}) = \mathbf{0} = T(\mathbf{0})$, which shows that $\mathbf{u} = \mathbf{0}$ by the injectivity of T.

Conversely, suppose $\ker(T) = \{\mathbf{0}\}$. Let $\mathbf{u}_1$ and $\mathbf{u}_2$ are such that $T(\mathbf{u}_1) = T(\mathbf{u}_2)$. Then by linearity,

$$\mathbf{0} = T(\mathbf{u}_1) - T(\mathbf{u}_2) = T(\mathbf{u}_1 - \mathbf{u}_2),$$

which shows that $\mathbf{u}_1 - \mathbf{u}_2$ is in the kernel of T. Since the kernel is the zero space, necessarily $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0}$, or $\mathbf{u}_1 = \mathbf{u}_2$. $\square$