MA283: Linear Algebra

70% Exam
30% Continuous Assessment (Homework)
10% Optional Project (Bonus)

Robert Davidson

Contents

1 Review of Matrix Algebra

Matrix Addition

If a matrix has m rows and n columns, we say it is $m \times n$.**Two matrices can only be added if they have the same size.**. In this case, we just add the entries in each position.

The $m \times n$ **zero** matrix is a matrix with all entries equal to 0. It is the **Identity element** for matrix addition (adding it to any matrix does not change the matrix)

Matrix Multiplication by a Scalar

This simply means multiplying each entry of the matrix by the scalar. For example:

$$\alpha \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} \alpha & 2\alpha \\ 3\alpha & 4\alpha \end{bmatrix}$$

Remark: Now that we have addition and scalar multiplication, we can subtract matrices ($A - B = A + (-1)B$), provided they are the same size.

Vector Space

With these operations of addition and scalar multiplication, the set of $m \times n$ matrices is a vector space. A **vector space** algebraic structure whose elements can be added, subtracted and multiplied by scalars.

Linear Combinations

Definition 1.1: Linear Combinations

Suppose $v_1, v_2, \ldots, v_k$ are elements that can be added together and multiplied by scalars.

A Linear Combination of $v_1, v_2, \ldots, v_k$ is an expression of the form:

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k$$

where $a_i \in \mathbb{R}$ are scalars, called **coefficients**.

Matrix-Vector Multiplication

Definition 1.2

Let A be a $m \times n$ matrix, and $\mathbf{v}$ be a column vector with n entries ($n \times 1$ matrix).

Then the matrix vector product Av is the column vector, with m entries, obtained by taking the linear combination of the columns of A with the entries of $\mathbf{v}$ as coefficients.

$$\begin{bmatrix} -1 & 2 & 4 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 7 \\ 6 \\ 9 \end{bmatrix} = 7 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + 6 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 9 \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 41 \\ 33 \end{bmatrix}.$$

Remark: Av, if defined, has the same number of rows as A and the same number of columns as $\mathbf{v}$.

Matrix-Matrix Multiplication

Definition 1.3

Let A and B be matrices of size $m \times p$ and $p \times n$, respectively.Write $v_1, \ldots v_n$ for the columns of B. Then the product AB is the $m \times n$ matrices whose columns are $Av_1, \ldots, Av_n$.

The entry at row i and column j of the matrix A is given by A_{ij}. The entry in the i, j position of the product AB is the ith entry of the vector Av_j, where the vector v_j is the jth column of B. In other words, the entry in the i, j position of the product AB is given by:

$$(AB)_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \ldots + A_{ip}B_{pj} = \sum_{k=1}^{p} A_{ik}B_{kj}$$

Definition 1.4

If A is $m \times p$ with rows $u_1, \dots, u_m$ and B is $p \times n$ with columns $v_1, \dots, v_n$, then the product AB is:

$$AB = \begin{bmatrix} u_1 \cdot v_1 & u_1 \cdot v_2 & \dots & u_1 \cdot v_n \\ u_2 \cdot v_1 & u_2 \cdot v_2 & \dots & u_2 \cdot v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_m \cdot v_1 & u_m \cdot v_2 & \dots & u_m \cdot v_n \end{bmatrix}$$

Example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad B = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix} \quad AB = \begin{bmatrix} 1 \cdot 7 + 2 \cdot 9 + 3 \cdot 11 & 1 \cdot 8 + 2 \cdot 10 + 3 \cdot 12 \\ 4 \cdot 7 + 5 \cdot 9 + 6 \cdot 11 & 4 \cdot 8 + 5 \cdot 10 + 6 \cdot 12 \end{bmatrix} = \begin{bmatrix} 58 & 64 \\ 139 & 154 \end{bmatrix}$$

For matrices A and B, the products AB and BA are generally not equal, even if they are both defined and even if both have the same size.

Linear Transformations

Definition 1.5

Let m and n be positive integers.

A **linear transformation** T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a function $T : \mathbb{R}^n \to \mathbb{R}^m$ that satisfies:

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$
- $T(\lambda \mathbf{u}) = \lambda T(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$

Matrix of a Linear Transformation

Suppose $T : \mathbb{R}^3 \to \mathbb{R}^2$ is the linear transformation:

$$T \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad T \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \quad T \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -6 \\ 7 \end{bmatrix}$$

Then for the vector in $\mathbb{R}^3$ with entries a, b, c:

$$T \begin{bmatrix} a \\ b \\ c \end{bmatrix} = aT \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + bT \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + cT \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -6 \\ -3 & 4 & 7 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Where the 2×3 matrix M_T is called the **standard matrix** of A. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ can be completely represented by an $m \times n$ matrix M_T.

Understanding the Matrix Representation

- The columns of matrix M_T are the images of the standard basis vectors $e_1, e_2, \dots, e_n$ under T.
- For any vector $v \in \mathbb{R}^n$, we calculate $T(v)$ by multiplying: $M_T \cdot v$.
- Therefore, matrix-vector multiplication is simply evaluating a linear transformation.

Correspondence: Any $m \times n$ matrix A defines a linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ by: $T_A(v) = Av$. Linear transformations include rotations, reflections and scaling

Efficiency of Representation: A remarkable property of linear transformations is their information efficiency:

- To completely define $T : \mathbb{R}^n \to \mathbb{R}^m$, we need only mn values.
- These values are the coordinates of the n transformed basis vectors in $\mathbb{R}^m$.
- This differs fundamentally from general continuous functions $f : \mathbb{R} \to \mathbb{R}$, which cannot be fully determined by their values at finitely many points.

Matrix multiplication is composition

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^p$ and $S : \mathbb{R}^p \to \mathbb{R}^m$ are linear transformations. Then the composition $S \circ T : \mathbb{R}^n \to \mathbb{R}^m$ is also a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ defined for $\mathbf{v} \in \mathbb{R}^n$ by:

$$S \circ T(\mathbf{v}) = S(T(\mathbf{v}))$$

To see how that the $m \times n$ matrix $M_{S \circ T}$ depends on the matrix $M_S(m \times p)$ and $M_T(p \times n)$ we look at the definition of $M_{S \circ T}$:

- The first column has coordinates $S \circ T(e_1) = S(T(e_1))$
- $T(e_1)$ is first column of M_T
- Then $S(T(e_1))$ is the matrix-vector product $M_S \cdot M_T(e_1)$
- Same for all other columns $\implies M_{S\circ T} = M_S \cdot M_T$

Thus, we conclude **matrix multiplication is composition of linear transformations**.

2 Systems of linear equations

2.1 Linear equations and Solution Sets

A linear equation in the variables x and y is an equation of the form

$$2x + y = 3$$

If we replace x and y with some numbers, the statement **becomes true or false**.

Definition 2.1: Solution to a linear equation

A pair, $(x_0, y_0) \in \mathbb{R}$, is a solution to an linear equation if setting $x = x_0$ and $y = y_0$ **makes the equation true.**

Definition 2.2: Solution set

The **solution set** is the set of all solutions to a linear equation.

$$a_1X_1 + a_2X_2 + \ldots + a_nX_n = b \quad \text{where } a_i, b \in \mathbb{R}$$

is an **affine hyperplane** in $\mathbb{R}^n$; geometrically resembles a copy of $\mathbb{R}^{n-1}$ inside $\mathbb{R}^n$.

2.1.1 Interpreting Linear Systems as Matrix Equations

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 5 \\ 3x & + & y & - & 2z & = & 9 \\ -x & + & 4y & + & 2z & = & 0 \end{array} \quad \Rightarrow \quad \begin{bmatrix} 1 & 2 & -1 \\ 3 & 1 & -2 \\ -1 & 4 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \\ 0 \end{bmatrix}$$

2.2 Elementary Row Operations

To solve a system of linear equations we associate an **augmented matrix** to the system of equations. For example:

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 5 \\ 3x & + & y & - & 2z & = & 9 \\ -x & + & 4y & + & 2z & = & 0 \end{array} \quad \Rightarrow \quad \left[\begin{array}{ccc|c} 1 & 2 & -1 & 5 \\ 3 & 1 & -2 & 9 \\ -1 & 4 & 2 & 0 \end{array}\right]$$

To solve, we can perform the following **Elementary Row Operations (EROs)**:

1. Multiply a row by a non-zero constant.
2. Add a multiple of one row to another row.
3. Swap two rows.

The goal of these operations is to transform the augmented matrix into **row echelon form** (REF) or **reduced row echelon form** (RREF).

2.2.1 REF and Strategy

We say a matrix is in **row echelon form** (REF) if:

- The first non zero entry in each row is a 1 (called the **leading 1**).
- If a column has a leading 1, then all entries below it are 0.
- The leading 1 in each row is to the right of the leading 1 in the previous row.
- All rows of 0s are at the bottom of the matrix.

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

Example of REF

We have produced a new system of equations. This is easily solved by back substitution.

Concept 2.1: Stategy for Obtaining REF

- Get a 1 as the top left entry
- Use this 1 to clear the entries below it
- Move to the next column and repeat
- Continue until all leading 1s are in place
- Use back substitution to solve the system

2.2.2 Row Reduced Echelon Form

A matrix is in **reduced row echelon form** (RREF) if:

- It is in REF
- The leading 1 in each row is the only non-zero entry in its column.

$$\begin{bmatrix} 1 & 0 & 0 & | & 2 \\ 0 & 1 & 0 & | & -1 \\ 0 & 0 & 1 & | & -1 \end{bmatrix}$$

Example of RREF

2.3 Leading variables and free variables

We'll start by an example:

$$\begin{array}{rcrcrcrcl} x_1 & - & x_2 & - & x_3 & + & 2x_4 & = & 0 \\ 2x_1 & + & x_2 & - & x_3 & + & 2x_4 & = & 8 \\ x_1 & - & 3x_2 & + & 2x_3 & + & 7x_4 & = & 2 \end{array} \Rightarrow \begin{bmatrix} 1 & -1 & -1 & 2 & | & 0 \\ 2 & 1 & -1 & 2 & | & 8 \\ 1 & -3 & 2 & 7 & | & 2 \end{bmatrix}$$

Solving this system of equations, we get:

$$\text{RREF: } \begin{bmatrix} 1 & 0 & 0 & 2 & | & 4 \\ 0 & 1 & 0 & -1 & | & 2 \\ 0 & 0 & 0 & 1 & | & 2 \end{bmatrix} \Rightarrow \begin{array}{rcrcl} x_1 & + & 2x_4 & = & 4 \\ x_2 & - & x_4 & = & 2 \\ x_3 & + & x_4 & = & 2 \end{array} \Rightarrow \begin{array}{rcl} x_1 & = & 4-2x_4 \\ x_2 & = & 2+x_4 \\ x_3 & = & 2-x_4 \end{array}$$

This RREF tells us how the **leading variables** (x_1, x_2, x_3) depend on the **free variable** (x_4). The free variable can take any value in $\mathbb{R}$. We write the solution set as:

$$x_1 = 4-2t, \quad x_2 = 2+t, \quad x_3 = 2-t, \quad x_4 = t \quad \text{where } t \in \mathbb{R}$$

$$(x_1, x_2, x_3, x_4) = (4-2t, 2+t, 2-t, t); \quad t \in \mathbb{R}$$

Definition 2.3: Leading and Free Variables

- **Leading variable** : A variable whose columns in the RREF contain a leading 1
- **Free variable** : A variable whose columns in the RREF do not contain a leading 1

2.4 Consistent and Inconsistent Systems

Consider the following system of equations:

$$\begin{array}{rcrcrcl} 3x & + & 2y & - & 5z & = & 4 \\ x & + & y & - & 2z & = & 1 \\ 5x & + & 3y & - & 8z & = & 6 \end{array} \Rightarrow \begin{bmatrix} 3 & 2 & -5 & | & 4 \\ 1 & 1 & -2 & | & 1 \\ 5 & 3 & -8 & | & 6 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & -2 & | & 1 \\ 0 & 1 & -1 & | & 1 \\ 0 & 0 & 0 & | & 1 \end{bmatrix} \text{ (REF)}$$

We can see the last row of the REF is:

$$0x + 0y + 0z = 1$$

This equation clearly has no solution, and hence the system has no solutions. We say the system is **inconsistent**. Alternatively, we say the system is **consistent** if it has at least one solution.

2.5 Possible Outcomes when solving a system of equations

- The system may be **inconsistent** (no solutions) - i.e:

$$[0\ 0\ \ldots\ 0 \mid a] \quad a \neq 0$$

- The system may be **consistent** which occurs if:
  - **Unique Solutions** each column (aside from the rightmost) contains a single leading 1. - i.e:

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

  - **Infinitely many solutions** at least one variable does not appear as a leading 1 in any row, making it a free variable - i.e:

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

2.6 Elementary Row Operations as Matrix Transformations

Elementary row operations may be interpreted as **matrix multiplication**. To see this, first we introduce the **Identity matrix:**

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The I_m Identity matrix is an $m \times m$ matrix with 1s on the diagonal and 0s elsewhere. We also introduce the $E_{i,j}$ matrix which has 1 in the (i, j) position and 0s elsewhere. For example:

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

Then:

$$I_3 + 4E_{1,2} = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Performing a row operation on A is the same as multiplying A by an appropriate matrix E on the left. These matrices are called **elementary matrices**. They are **always invertible,** and **their inverses are also elementary matrices**. The statement:

"every matrix can be reduced to RREF through EROs"

is equivalent to saying that

"for every matrix A with m rows, there exists a $m \times m$ matrix B which is a product of elementary matrices such that BA is in RREF."

2.6.1 Multiplying a Row by a Non-Zero Scalar

When multiplying row i of matrix A by a scalar $\alpha \neq 0$, we can use the matrix:

$$I_m + (\alpha - 1)E_{i,i}$$

This works because it modifies only the (i, i) entry of the identity matrix to be α while keeping all other entries unchanged. When multiplied with A, it scales row i by α and leaves all other rows intact.
Example: If $\alpha = 5$ and $i = 2$, then:

$$I_3 + 4E_{2,2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad (I_3 + 4E_{2,2})A = \begin{bmatrix} 1 & 2 & 3 \\ 20 & 25 & 30 \\ 7 & 8 & 9 \end{bmatrix}$$

2.6.2 Switching Two Rows

To swap rows i and k, we use:

$$S = I_m + E_{i,k} + E_{k,i} - E_{i,i} - E_{k,k}$$

This works by:

- Removing the 1's at positions (i, i) and (k, k) from the identity matrix
- Adding 1's at positions (i, k) and (k, i)

Example: Swapping rows 1 and 3:

$$S = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad SA = \begin{bmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{bmatrix}$$

2.6.3 Adding a Multiple of One Row to Another

To replace row k with row $k + \alpha \times$ row i, use:

$$I_m + \alpha E_{k,i}$$

This adds α times row i to row k while leaving all other rows unchanged because:

- For any row $j \neq k$, the corresponding row in this matrix is just the standard basis row
- Row k becomes the sum of the standard basis row k plus α times the standard basis row i

Example: Adding 3 times row 1 to row 2:

$$I_3 + 3E_{2,1} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad (I_3 + 3E_{2,1})A = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 11 & 15 \\ 7 & 8 & 9 \end{bmatrix}$$

Example 2.1

Write the inverse of an elementary matrix and show it is an elementary matrix.

Multiplying a row by a nonzero scalar:

- **Operation:** Multiply row i by $\alpha \neq 0$.
- **Elementary Matrix:** $E = I_m + (\alpha - 1)E_{i,i}$
- **Inverse:** To reverse the operation, multiply row i by $1\backslash\alpha$. Hence, the inverse is

$$E^{-1} = I_m + \left(\frac{1}{\alpha} - 1\right) E_{i,i} = I_m + \frac{1-\alpha}{\alpha} E_{i,i}.$$

Swapping two rows:

- **Operation:** Swap rows i and k.
- **Elementary Matrix:** $S = I_m - E_{i,i} - E_{k,k} + E_{i,k} + E_{k,i}$
- **Inverse:** Since swapping the same two rows twice returns them to their original positions,

$$S^{-1} = S.$$

Adding a multiple of one row to another:

- **Operation:** Add α times row i to row k.
- **Elementary Matrix:** $E = I_m + \alpha\, E_{k,i}$
- **Inverse:** To undo the operation, subtract α times row i from row k. Therefore,

$$E^{-1} = I_m - \alpha\, E_{k,i}.$$

Example 2.2

Prove that every invertible matrix in $M_n(\mathbb{R})$ is a product of elementary matrices.

Let A be an invertible matrix in $M_n(\mathbb{R})$. Since A is invertible, we can use Gaussian elimination to transform A into the identity matrix I_n.

Let $E_1, E_2, \ldots, E_k$ be the elementary matrices corresponding to the row operations used in the elimination process. Then, we have:

$$\begin{aligned} \text{Multiplying a row by a scalar:} &\quad I_n + (\alpha - 1)E_{i,i} \\ \text{Swapping two rows:} &\quad I_n + E_{i,k} + E_{k,i} - E_{i,i} - E_{k,k} \\ \text{Adding a multiple of one row to another:} &\quad I_n + \alpha\, E_{k,i} \end{aligned}$$

Applying these in sequence to A gives:

$$E_k \cdots E_2 E_1 A = I_n$$

Since $E_k \cdots E_2 E_1 = I_n$, we can multiply both sides by $(E_k \cdots E_2 E_1)^{-1}$ on the left to obtain:

$$A = (E_k \cdots E_2 E_1)^{-1} I_n = (E_k \cdots E_2 E_1)^{-1}$$

Using the property:

$$(E_k \cdots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_k^{-1},$$

we can express A as a product of elementary matrices:

$$A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$$

Since each E_i is an elementary matrix, its inverse is also an elementary matrix. Therefore, A can be expressed as a product of elementary matrices.

2.7 EROs and Inverses

Elementary Row Operations can be used to find the inverse of a square matrix. Consider a square matrix $A \in M_n(\mathbb{F})$ (that is, an $n \times n$ matrix over a field $\mathbb{F}$). If A is invertible, let

$$A^{-1} = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}$$

be its inverse, where each $\mathbf{v}_i$ is the ith column of A^{-1}. By definition of the matrix inverse, we have

$$A\,A^{-1} = A\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix} = \begin{bmatrix}A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n\end{bmatrix} = I_n,$$

the $n \times n$ identity matrix. This implies that

$$A\mathbf{v}_i = \mathbf{e}_i, \quad \text{for each } i = 1, 2, \ldots, n,$$

where $\mathbf{e}_i$ is the ith column of I_n (which has a 1 in the ith row and 0 everywhere else). In other words, each column $\mathbf{v}_i$ of A^{-1} is the unique solution to the linear system

$$A\mathbf{v}_i = \mathbf{e}_i.$$

To find A^{-1} effectively, we form the augmented matrix $[A \mid I_n]$ and apply EROs to trasnform A into I_n. When this is achieved, the augmented portion becomes A^{-1}. Thus, we have

$$\mathrm{RREF}\big([\,A \mid I_n\,]\big) = \big[\,I_n \mid A^{-1}\big].$$

Example 2.3

Find A^{-1} if $A = \begin{bmatrix}3 & 4 & -1\\ 1 & 0 & 3\\ 2 & 5 & -4\end{bmatrix}$.

We form a 3×6 matrix $A' = [A \mid I_3]$:

$$A' = \begin{bmatrix}3 & 4 & -1 & | & 1 & 0 & 0\\ 1 & 0 & 3 & | & 0 & 1 & 0\\ 2 & 5 & -4 & | & 0 & 0 & 1\end{bmatrix}$$

We apply the following EROs to A':

- $R_1 \leftrightarrow R_2$
- $R_2 \to R_2 - 3R_1$
- $R_3 \to R_3 - 2R_1$
- $R_3 \to R_3 + R - 2$
- $R_3 \leftrightarrow R_2$
- $R_3 \to R_3 - 4R_2$
- $R_3 \times (-\frac{1}{10})$
- $R_1 \to R_1 - 3R_3$

To obtain:

$$\begin{bmatrix}1 & 0 & 0 & | & \frac{3}{2} & -\frac{11}{10} & -\frac{6}{5}\\ 0 & 1 & 0 & | & -1 & 1 & 1\\ 0 & 0 & 1 & | & -\frac{1}{2} & \frac{7}{10} & \frac{2}{5}\end{bmatrix}$$

That is:

$$A^{-1} = \begin{bmatrix}\frac{3}{2} & -\frac{11}{10} & -\frac{6}{5}\\ -1 & 1 & 1\\ -\frac{1}{2} & \frac{7}{10} & \frac{2}{5}\end{bmatrix}$$

It is easily checked that $AA^{-1} = I_3$.

3 Vector Spaces and Subspace Structure

3.1 The Image and Kernel of a Linear Transformation

$T : \mathbb{R}^3 \to \mathbb{R}^3$ is the linear transformation defined with:

$$M_T = \begin{bmatrix}1 & 2 & 0\\ 2 & -1 & 5\\ 1 & 1 & 1\end{bmatrix}$$

The **image** of T is the subset of $\mathbb{R}^3$ consisting of all elements $T(\mathbf{v}), \mathbf{v} \in \mathbb{R}^3$. This is the set of all vectors of the form:

$$a\begin{bmatrix}1\\2\\1\end{bmatrix} + b\begin{bmatrix}2\\-1\\1\end{bmatrix} + c\begin{bmatrix}0\\5\\1\end{bmatrix}$$

In matrix terms, this is the **column space** of M_T.
The **kernel** of T is the set of all vectors $\mathbf{v} \in \mathbb{R}^3$ such that $T(\mathbf{v}) = \mathbf{0}$. This is the set of all column vectors, whose entries, a, b, c satisfies:

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 5 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + b \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} + c \begin{bmatrix} 0 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The kernel is a line and the image is a plane

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & -1 & 5 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \left[\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & -5 & 5 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right] \quad \Rightarrow \quad \left[\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The kernel (or nullspace) is $(2, 1, 1)t, t \in \mathbb{R}$, which is a line in $\mathbb{R}^3$. The fact that $(-2, 1, 1)$ is in the kernel of T, means that column 3 of M_T is a linear combination of columns 1 and 2.

$$-2 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Longrightarrow \quad \begin{bmatrix} 0 \\ 5 \\ 1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

It follows that every linear combination of all three columns of M_T is just a linear combination of columns 1 and 2.

The column space of M_T is:

$$\left\{ a \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + b \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \ : \ a, b \in \mathbb{R} \right\}$$

3.2 Subspaces

Definition 3.1

A non empty subset $\mathbf{V}$ of $\mathbb{R}^n$ is a **subspace** if:

- **Closed under addition**: $u + v \in \mathbb{V}$, $u, v \in \mathbf{V}$
- **Closed under scalar multiplication**: $\alpha u \in \mathbf{V}$, $u \in \mathbf{V}$, $\alpha \in \mathbb{R}$

Examples of subspaces

- $\{(x, y, z) \in \mathbb{R}^3 : x + y + z = 1\}$ is not a subspace of $\mathbb{R}^3$. the $[1, 0, 0]$ and $(0, 1, 0)$ vectors are in the set, but their sum $(1, 1, 0)$ is not in the set.
- $\{(x, y, z) \in \mathbb{R}^3 : (x, y, z) \cdot (1, 2, 3) = 0\}$ is a subspace of $\mathbb{R}^3$.
- $\{(x, y, z) \in \mathbb{R}^3 : (x, y, z) \cdot (1, 2, 3) \neq 0\}$ is not a subspace of $\mathbb{R}^3$.
- The kernel of any linear transformation is a subspace of $\mathbb{R}^n$.
- The image of any linear transformation is a subspace of $\mathbb{R}^n$.

3.3 The span : how to make subspaces

Definition 3.2

Let $S = \{v_1, \ldots, v_k\}$ be any finite subset of $\mathbb{R}^n$

The subset of $\mathbb{R}^n$ consisting of all linear combinations of the elements of S is a subspace of $\mathbb{R}^n$ and is called the **span** of S and is denoted by $\langle (S) \rangle$.

Proof that $\langle S \rangle$ is a subspace of $\mathbb{R}^n$

- **Closed under addition**:
  Let $u, v \in \langle S \rangle$. Then $u = a_1 v_1 + a_2 v_2 + \cdots + a_k v_k$ and $v = b_1 v_1 + b_2 v_2 + \cdots + b_k v_k$ for some $a_i, b_i \in \mathbb{R}$. We see that:
  $$u + v = (a_1 + b_1) v_1 + (a_2 + b_2) v_2 + \cdots + (a_k + b_k) v_k$$
  So S is closed under addition.
- **Closed under scalar multiplication**:
  Let $u \in \langle S \rangle$ and $\alpha \in \mathbb{R}$. We need to show that cu is a linear combination of $v_1, \ldots, v_k$. We have $u = a_1 v_1 + a_2 v_2 + \cdots + a_k v_k$ for some $a_i \in \mathbb{R}$. Then:
  $$cu = c(a_1 v_1 + a_2 v_2 + \cdots + a_k v_k) = (ca_1) v_1 + (ca_2) v_2 + \cdots + (ca_k) v_k$$
  so $cu \in \langle S \rangle$.

3.4 Spanning sets

Definition 3.3

Let V be a subspace of $\mathbb{R}^n$.

A subset S of V is a **spanning set** for V if $\langle S \rangle = V$.

This means that every element of V can be expressed as a linear combination of the elements of S.

Example
The set $\{e_1, e_2, e_3\}$ is a spanning set of $\mathbb{R}^3$. We know that:

$$e_1 = [1, 0, 0], \quad e_2 = [0, 1, 0], \quad e_3 = [0, 0, 1]$$

We can represent every element of $\mathbb{R}^3$ as a linear combination of e_1, e_2, e_3:

$$\begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} = 2e_1 - 3_2 + 4e_3$$

Remark A set S of three column vectors in $\mathbb{R}^3$ is a spanning set of $\mathbb{R}^3$ if and only if the three vectors are linearly independent.This occurs only if the 3×3 matrix whose columns are the three vectors has S **as an inverse**.

Questions about spanning sets

- Does $\mathbb{R}^3$ have a spanning set fewer than three vectors?
  - **No.** A spanning set for $\mathbb{R}^3$ must contain at least three linearly independent vectors, since the dimension of $\mathbb{R}^3$ is 3. Fewer than three vectors cannot span all of $\mathbb{R}^3$.
- Does every spanning set of $\mathbb{R}^3$ have three vectors?
  - **No.** A spanning set can have more than three vectors, but not necessarily exactly three. Redundant vectors (linearly dependent ones) can be included, so a spanning set might have more than three vectors.
- Does every spanning set of $\mathbb{R}^{\nvDash}$ contain one with exactly three elements?
  - **Yes.** Every spanning set of $\mathbb{R}^3$ contains a basis, and since the dimension is 3, there exists a subset of exactly three linearly independent vectors that still span $\mathbb{R}^3$.
- If V is a subspace of $\mathbb{R}^3$ does V have a spanning set with at most three elements?
  - **Yes.** Any subspace of $\mathbb{R}^3$ has a basis, and since $\mathbb{R}^3$ has dimension 3, the basis of any of its subspaces can have at most 3 elements. Hence, every subspace can be spanned by at most three vectors.
- If V is a proper subspace of $\mathbb{R}^3$, does V have a spanning set with fewer than three elements?
  - **Yes.** A proper subspace of $\mathbb{R}^3$ has dimension less than 3, so it can be spanned by fewer than three vectors.

3.5 Linear Dependence and Linear Independence

Definition 3.4

A set of at least two vectors in $\mathbb{R}^n$ is **linearly dependent** if one of its elements is a linear combination of the others.

A set of vectors in $\mathbb{R}^n$ is **linearly independent** if it is not linearly dependent.

For a subset $\{v_1, \ldots, v_k\}$ of $\mathbb{R}^n$, suppose that v_k is a linear combination of $\{v_1, \ldots, v_{k-1}\}$. Then every linear combination of $\{v_1, \ldots, v_k\}$ **is already a linear combination** of $v_1, \ldots, v_{k-1}$:

$$\langle v_1, \ldots, v_k \rangle = \langle v_1, \ldots, v_{k-1} \rangle$$

If we are interested in the span of $\{v_1, \ldots, v_k\}$, we can throw away v_k and this wouldn't change the span.

Linear independence means that throwing away any element of the set **shrinks the span**

Example 3.1

$$\begin{array}{rcrcrcl} x_1 & + & 3x_2 & + & 5x_3 & - & 9x_4 = 5 \\ 3x_1 & - & x_2 & - & 5x_3 & + & 13x_4 = 5 \\ 2x_1 & - & 3x_2 & - & 8x_3 & + & 18x_4 = 1 \end{array} \quad \Rightarrow \quad \left[\begin{array}{cccc|c} 1 & 3 & 5 & -9 & 5 \\ 3 & -1 & -5 & 13 & 5 \\ 2 & -3 & -8 & -18 & 51 \end{array}\right] \quad \Rightarrow \quad \left[\begin{array}{cccc|c} 1 & 0 & -1 & 3 & 2 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The three equations of the system form a linearly dependent set. One row was eliminated by adding a linear combination of the other two rows. Thus, all the information in the system was contained in the first two equations.

The non-zero rows of the RREF are linearly independent, they span the rowspace of the matrix. The rowspace is the subspace of $\mathbb{R}^5$ spanned by the rows of the matrix.

3.5.1 Test for linear independence

A set is linearly independent if none of its elements is a linear combination of the others. While this makes sense, to use it as a test would mean checking every element. We have an alternative formulation, which is easier to check:

"A set of vectors is linearly independent if the only way to write the zero vector as a linear combination of the vectors in the set is to use all zero coefficients."

To decide if the set $\{v_1, \dots, v_k\}$ is linearly independent, try to write the zero vector as a linear combination of the vectors in the set:

$$\sum_{i=1}^{k} \alpha_i v_i = \alpha_1 v_1 + \cdots + \alpha_k v_k = 0 \quad \text{for } \alpha_i \in \mathbb{R}$$

If $\forall i \to a_i = 0$, then the set is linearly independent. If not, the set is linearly dependent.

Example 3.2

Decide whether the set $\{[1, 0, 1], [1, 0, -1], [1, 1, 1]\}$ is linearly independent or dependent.

To solve, we use ERO and find:

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad a = b = c = 0$$

The set is linearly independent

3.6 Finite Dimensional Spaces

Definition 3.5

A vector space V is finite dimensional if it contain a finite spanning set.

This means a set $\{v_1, \dots, v_k\}$ of elements, with the property that every element of V is a linear combination of $v_1, \dots, v_k$.

Examples

- $\mathbb{R}^n$ is finite dimensional with $\{e_1, \dots, e_n\}$ as a spanning set. The dimension of $\mathbb{R}^n$ is n.
- $M_{m\times n}(\mathbb{R})$ is finite dimensional,with $\{E_{ij}\}_{1\le i\le m, 1\le j\le n}$ as a spanning set with mn elements.
- An example of an infinite dimensional space is the set, $\mathbb{R}[x]$, of all polynomials with real coefficients. This set is infinite dimensional because it contains an infinite number of linearly independent vectors, such as $\{1, x, x^2, \dots\}$.

3.7 Basis

Definition 3.6

A **basis** for a vector space is a **linearly independent spanning set**.

- A basis is a minimal spanning set, one in which every element is needed and does not contain a smaller spanning set.

- Example: $\{e_1, \dots, e_n\}$ is a basis for $\mathbb{R}^n$.
- $\{(1,3),(1,4)\}$ is a basis for $\mathbb{R}^2$.
- If S is a finite spanning set of a vector space V, then S contains a basis of V. If S is not linearly independent, then some $v \in S$ is a linear combination of the other elements of S. Throwing away v leaves a smaller set that still spans V. This process can be repeated until a basis is obtained.

3.8 Steinitz Replacement Lemma

Lemma 3.1

Let V be a vector space that has a basis with n elements.

Then every linearly independent set with n elements in V is a basis for V.

Proof (for $n = 3$)
Suppose $B = \{b_1, b_2, b_3\}$ is a basis of V and let $\{y_1, y_2, y_3\}$ be a linearly independent subset of V.

Step 1.
$y_1 = a_1 b_1 + a_2 b_2 + a_3 b_3$ for scalars a_1, a_2, a_3, not all zero. We can assume (after maybe relabelling the b_i), that $a_1 \neq 0$. Then

$$b_1 = a_1^{-1} y_1 - a_1^{-1} a_2 b_2 - a_1^{-1} a_3 b_3.$$

So $b_1 \in \langle y_1, b_2, b_3 \rangle$ and $\{y_1, b_2, b_3\}$ spans V. (Note that we have to use the fact that we can divide by non-zero scalars to write b_1 as a linear combination of y_1, b_2, b_3.)

Step 2.
Now $y_2 \in \langle y_1, b_2, b_3 \rangle$ and y_2 is not a scalar multiple of y_1 (because $\{y_1, y_2, y_3\}$ is linearly independent).
So b_2 (or b_3) has non-zero coefficient in any description of y_2 as a linear combination of y_1, b_2, b_3.
Replace again: $\{y_1, y_2, b_2\}$ spans V.

Step 3.
Same reasoning: we can replace b_2 with y_3 to conclude $\{y_1, y_2, y_3\}$ spans V.

Conclusion: $\{y_1, y_2, y_3\}$ is a basis of V.

3.9 Recap of span, linear independence and basis

Let V be a vector space, e.g. $V = \mathbb{R}^n$ and S be a finite subset of V. Let V be a vector space (e.g. $V = \mathbb{R}^n$). Let S be a (finite) subset of V.

1. S is a spanning set of V (or S spans V) if every element of V is a linear combination of the elements of S.
2. The span of S, denoted $\langle S \rangle$, is the set of all linear combinations of elements of S, a subspace of V.
3. S is linearly independent if no element of S is a linear combination of the other elements of S.
   Equivalently, if no proper subset of S spans $\langle S \rangle$.
4. S is a basis of V if S is linearly independent **AND** S spans V.
   A basis is a minimal spanning set.
   A basis is a maximal linearly independent set.
5. Every finite spanning set of V contains a basis of V.
6. Every linearly independent subset of V can be extended to a basis of V (we have not proved this yet!).

3.10 Consequences of the replacement theorem

Theorem 3.1

Let V be a vector space that has a basis with n elements.

Then ever linearly independent set with n elements in V is a basis for V.

If V has a spanning set with n elements, a linearly independent set in V cannot have more than n elements.

If V has a linearly independent set with n elements, a spanning set in V must have at least n elements. More concisely:

Concept 3.1

The number of elements of a linearly independent set cannot exceed the number in a spanning set. Every spanning set has at least as many elements as the biggest independent set.

3.11 Every basis has the same number of elements

Let V be a finite dimensional vector space and let B and B' be the bases of V. Then:

- B is linearly independent and B' is a spanning set, so B has **at most** as many elements as B'.
- B is a spanning set and B' is linearly independent, so B has **at least** as many elements as B'.

It follows that B and B' have the same number of elements.

Definition 3.7

The dimension of V is the number of elements in a basis of V.

Note: Every vector space that has a finite spanning set has a finite basis (since we can discard elements from a finite spanning set until a basis remains).

Examples:

- The set $\{1, x, x^2, x^3\}$ is a basis for the vector space P_3 of all polynomials of degree at most 3 with real coefficients.

  It is linearly independent because the only way to write the zero polynomial as

  $$a_3x^3 + a_2x^2 + a_1x + a_0 = 0$$

  is by taking $a_0 = a_1 = a_2 = a_3 = 0$.

  Another basis of P_3, preferable for some applications, consists of the first four Legendre polynomials:

  $$\{1,\ x,\ \tfrac{1}{2}(3x^2 - 1),\ \tfrac{1}{2}(5x^3 - 3x)\}.$$

- The **row space** of an $m \times n$ matrix is the subspace of $\mathbb{R}^n$ spanned by its rows. When we reduce a matrix to row-reduced echelon form (RREF), we are computing a basis of its row space.
- In $\mathbb{R}^2$, the reflection in the line $y = 2x$ sends:

  $$(1, 0) \mapsto \left(-\tfrac{3}{5}, \tfrac{4}{5}\right), \quad (0, 1) \mapsto \left(\tfrac{4}{5}, \tfrac{3}{5}\right).$$

  Its standard matrix is:

  $$\begin{bmatrix} -\frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}.$$

  The same reflection sends:

  $$(1, 2) \mapsto (1, 2), \quad (2, -1) \mapsto (-2, 1).$$

  It is easier to describe this transformation in terms of the basis:

  $$\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}.$$

3.12 Row rank and column rank

Let A be an $m \times n$ matrix.

The **row rank** of A, denoted r, is the dimension of the row space of A—the subspace of $\mathbb{R}^n$ spanned by the rows of A.

The **column rank** of A, denoted c, is the dimension of the column space of A—the subspace of $\mathbb{R}^m$ spanned by the columns of A. Equivalently, it is the dimension of the image of the linear transformation represented by A.

- The row rank is the number of linearly independent rows in A.
- The column rank is the number of linearly independent columns in A.

Since the row rank is at most m and the column rank at most n, both values can be strictly less than m or n, respectively.

3.13 Row rank = column rank

Theorem 3.2

The row rank and column rank are the same for every matrix

Therefore, we refer to this common value simply as the **rank** of the matrix.

Let A be an $m \times n$ matrix. The row rank r is the number of non-zero rows in the RREF of A, and the column rank c is the number of linearly independent columns of A. To show that $c \leq r$, consider a basis for the row space of A and arrange its vectors as the rows of an $r \times n$ matrix P. Since every row of A is a linear combination of the rows of P, there exists an $m \times r$ matrix Q such that

$$A = QP.$$

It follows that each column of A is a linear combination of the r columns of Q, implying that $\dim(\text{col}(A)) \leq r$. Hence, $c \leq r$.
Conversely, to show that $r \leq c$, take a basis for the column space of A and arrange its vectors as the columns of an $m \times c$ matrix P'. Since every column of A is a linear combination of the columns of P', there exists a $c \times n$ matrix Q' such that

$$A = P'Q'.$$

Therefore, each row of A is a linear combination of the c rows of Q', and $\dim(\text{row}(A)) \leq c$. Hence, $r \leq c$.
Combining both inequalities, we conclude that $r = c$, i.e., the row rank and column rank are equal.

Example 3.3

Step 1: Determine the Rank. Reduce A to its reduced row echelon form (RREF):

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 1 \end{bmatrix} \implies \text{RREF}(A) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

There are 2 non-zero rows, so the row rank is $r = 2$. Examining the columns, we observe that the third column is a linear combination of the first two, so the column rank $c = 2$ as well.

Step 2: Show that $c \leq r$ via $A = QP$. Choose a basis for the row space from the non-zero rows of RREF

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

Thus, each column of A is a linear combination of the $r = 2$ columns of Q, implying that $\dim(\text{col}(A)) \leq r$.

Step 3: Show that $r \leq c$ via $A = P'Q'$. Take a basis for the column space of A

$$A = P'Q' = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 1 \end{bmatrix}.$$

So, each row of A is a linear combination of the $c = 2$ rows of Q', and $\dim(\text{row}(A)) \leq c$.

Conclusion. Since $c \leq r$ and $r \leq c$, it follows that $r = c = 2$. This example confirms that the row rank equals the column rank.

3.14 Coordinates

Lemma 3.2

If $\{b_1, \ldots, b_n\}$ is a basis of a vector space V, then every element of V has a unique expression of a linear combination of $b_1, \ldots, b_n$:

Proof: Suppose, for some $v \in V$, that:

$$v = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n, \quad a_i \in \mathbb{R}$$
$$v = a_1' b_1 + a_2' b_2 + \cdots + a_n' b_n, \quad a_i' \in \mathbb{R}$$

Then:

$$0_v = (a_1 - a_1')b_1 + (a_2 - a_2')b_2 + \cdots + (a_n - a_n')b_n$$

Since B is linearly independent, we have:

$$a_i - a_i' = 0 \implies a_i = a_i', \quad \forall\, i$$

Example 3.4

In $\mathbb{R}^2$, the standard coordinates of $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$ are $(4, 3)$.

With respect to the basis, $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$ the coordinates of $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$ are $(2, -1)$. Which is saying:

$$\begin{bmatrix} 4 \\ 3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

3.15 Coordinates with respect to different bases

Let B be the ordered basis of $\mathbb{R}^3$ with elements:

$$b_1 = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \quad b_3 = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix} \quad \Longrightarrow \quad B = \begin{bmatrix} 1 & 2 & 4 \\ 0 & -1 & 0 \\ 4 & 0 & 2 \end{bmatrix}$$

Given an element of $\mathbb{R}^3$, say v, how do we find the B-coordinates of v?

$$v = \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}$$

We know

$$v = 2e_1 - 3e_2 + 4e_3 = [v]_B = 2[e_1]_B - 3[e_2]_B + 4[e_3]_B.$$

To find $[e_1]_B$:

$$e_1 = xb_1 + yb_2 + zb_3 = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix} x + \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} y + \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix} z = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 2 & 4 \\ 0 & -1 & 0 \\ 4 & 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

This is saying that $[e_1]_B$ is the first column of the inverse of the matrix B. Write a matrix P which has b_1, b_2, b_3 as columns:

$$P = \begin{bmatrix} 1 & 2 & 4 \\ 0 & -1 & 0 \\ 4 & 0 & 2 \end{bmatrix} \quad \Longrightarrow \quad P^{-1} = \begin{bmatrix} -\frac{1}{7} & -\frac{2}{7} & \frac{2}{7} \\ 0 & -1 & 0 \\ \frac{2}{7} & \frac{4}{7} & -\frac{1}{14} \end{bmatrix}$$

To find the B-coordinates of any $v \in \mathbb{R}^3$, we can multiply v on the left by P^{-1}:

$$\begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}_B = P^{-1}v = \begin{bmatrix} -\frac{1}{7} & -\frac{2}{7} & \frac{2}{7} \\ 0 & -1 & 0 \\ \frac{2}{7} & \frac{4}{7} & -\frac{1}{14} \end{bmatrix} \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{12}{7} \\ 3 \\ -\frac{10}{7} \end{bmatrix}$$

This is saying that $v = \frac{12}{7}b_1 + 3b_2 - \frac{10}{7}b_3$. P^{-1} is called the **change of basis matrix** from the standard basis to the basis B.

3.16 The Rank Nullity Theorem

The Rank-Nullity Theorem **relates the dimensions of the kernel, image and domain** of a linear transformation. The dimension of the image of a linear transformation is called the **rank** and the dimension of the the kernel is called the **nullity.** The rank of T is equal to the rank of matrix X, since the image of T is the column space of this matrix.

Theorem 3.3

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Where V and W are finite-dimensional vector spaces, over some field $\mathbb{F}$. Then:

$$\dim(\ker(T)) + \operatorname{rank}(T) = n$$

Informally: The rank-nullity theorem says the full dimension of the domain must be accounted for in the combination of the kernel and image.

Proof:

1. Write k for $\dim(\ker(T))$ and let $\{b_1, \ldots, b_k\}$ be a basis of $\ker(T)$.
2. Extend this to a basis: $\{b_1, \ldots, b_k, b_{k+1}, \ldots, b_n\}$ of $\mathbb{R}^n$.
3. Since T sends each b_i to 0, the image under T of every element of $\mathbb{R}^n$ is a linear combination of $T(c_{k+1}), \ldots, T(c_n)$.

4. Also, $\{T(c_{k+1}), \ldots T(c_n)\}$ is a linearly independent subset of $\mathbb{R}^m$. To see this suppose for some scalar $a_{k+1}, \ldots, a_n$ that $a_{k+1}T(c_{k+1}) + \cdots + a_nT(c_n) = 0$. Then:

$$a_{k+1}c_{k+1} + \cdots + a_nc_n \in \ker T \implies a_{k+1}c_{k+1} + \cdots + a_nc_n \in \langle b_1, \ldots, b_k \rangle.$$

   Since $\{b_1, \ldots, b_k, c_{k+1}, \ldots, c_n\}$ is linearly independent in $\mathbb{R}^n$, this means that $a_{k+1}c_{k+1} + a_{k+2}c_{k+2} + \cdots + a_nc_n = 0$ for each $a_j = 0$.
5. It follows that $\{T(c_{k+1}), \ldots, T(c_n)\}$ is a basis for the image of T, so this image has dimensions $n - k$ as required.

3.17 Linear transformations and change of basis

Definition 3.8

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and let $B = \{b_1, \ldots, b_n\}$ be a basis of $\mathbb{R}^n$.

The **matrix of** T **with respect to the basis** B is the $n \times n$ matrix that has the B-coordinates of $T(b_1), \ldots, T(b_n)$ as its n columns. This matrix M satisfies

$$[T(v)]_b = M[v]_B, \quad \forall v \in \mathbb{R}^n.$$

Example 3.5

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation defined by $v \to Av$ and B be the ordered basis of $\mathbb{R}^3$ with elements:

$$A = \begin{bmatrix} -2 & 2 & 1 \\ 4 & 5 & -1 \\ -4 & -8 & 3 \end{bmatrix} \quad ||| \quad b_1 = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \quad b_3 = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix} \quad \implies \quad B = \begin{bmatrix} 1 & 2 & 4 \\ 0 & -1 & 0 \\ 4 & 0 & 2 \end{bmatrix}$$

A diagonal representation of the matrix A' of T with respect to the basis B

$$T(b_1) = \begin{bmatrix} -2 & 2 & 1 \\ 4 & 5 & -1 \\ -4 & -8 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 8 \end{bmatrix} = 2b_1 \implies [T(b_1)]_B = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

$$T(b_2) = \begin{bmatrix} -2 & 2 & 1 \\ 4 & 5 & -1 \\ -4 & -8 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -6 \\ 3 \\ 0 \end{bmatrix} = -3b_2 \implies [T(b_2)]_B = \begin{bmatrix} 0 \\ -3 \\ 0 \end{bmatrix}$$

$$T(b_3) = \begin{bmatrix} -2 & 2 & 1 \\ 4 & 5 & -1 \\ -4 & -8 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -7 \\ 14 \end{bmatrix} = 7b_3 \implies [T(b_3)]_B = \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix}$$

The matrix of A' of T with respect to B is **diagonal**. For describing this transformation, T, the basis B is preferable to the standard basis.

$$A' = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

This means for any $v \in \mathbb{R}^3$:

$$\underbrace{[T(v)]_B}_{\text{B-coordinates of } T(v)} = \underbrace{A'[v]_B}_{\text{matrix-vector product}}$$

3.18 More on Change of Basis

Let P be the matrix with the basis vectors from B as columns. As we've seen, P^{-1} is the change of basis matrix from the standard basis to the basis B.

For any element, $v \in \mathbb{R}^n$, its B-coordinates are given by:

$$[v]_B = P^{-1}v$$

Equivalently, if we start with the B-coordinates, then the standard coordinates of v are given by:

$$v = P[v]_B$$

So P itself, is the change of basis matrix from the basis B to the standard basis.

3.19 Similarity (The relation of A and A')

Starting with A, the matrix $T : \mathbb{R}^3 \to \mathbb{R}^3$ with respect to the standard basis, to we find A', the matrix of T with respect to the basis Bm we:

- **Take a vector $v \in \mathbb{R}^3$ written in B-coordinates as the column** $[v]_B$.
- **Convert to standard coordinates:** take the product $P[v]_B$.
- **Apply T:** Left multiply by A, to get $AP[v]_B$. This column is the standard coordinates of $T(v)$.
- **Convert to B-coordinates:** left multiply by P^{-1} (change of basis matrix from standard to B) to get $P^{-1}AP[v]_B$. This column has the B-coordinates of $T(v)$.
- **Conclusion:** For any element $v \in \mathbb{R}^3$, the B-coordinates of $T(v)$ are given by:

$$[T(v)]_B = P^{-1}AP[v]_B$$

The B matrix of T is: $P^{-1}AP$, where P has the elements of B as columns.

3.20 Similar Matrices

Definition 3.9

Two matrices A and B are **similar** if there exists an invertible matrix P such that:

$$B = P^{-1}AP$$

Notes:

- Two distinct matrices are similar only if they **represent the same linear transformation** with respect to different bases.
- We can't tell if two square matrices are similar just by looking. Instead, we look at the **trace** of the matrix, the sum of the diagonal elements, if two matrices have the same trace, they are similar.
- Similar matrices have some features in common, such as the same determinant, the same eigenvalues, etc (more on this later).
- Our example showed that the matrix:

$$A = \begin{bmatrix} -2 & 2 & 1 \\ 4 & 5 & -1 \\ -4 & -8 & 3 \end{bmatrix}$$

  is similar to the diagonal matrix:

$$\text{diag}(2, -3, 7) = A' = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

  In this situation, we say that A is **diagonalizable**.

3.21 Two interpretations of diagonalizability

Interpretation 1:
From the diagonal form of A' we have:

$$T(b_1) = 2b_1, \quad T(b_2) = -3b_2, \quad T(b_3) = 7b_3$$

This means that each of the basis elements b_1, b_2, b_3 is mapped by T to a scalar multiple of itself. - each of them is an eigenvector of T

Interpretation 2:
We can rearrange the version of $P^{-1}AP = A'$ to get:

$$AP = P'A$$

Bearing in mind that that:

$$P = \begin{bmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{bmatrix} \quad \text{and} \quad A' = \text{diag}(2, -3, 7)$$

this is saying that:

$$A \begin{bmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 7 \end{bmatrix} = \begin{bmatrix} | & | & | \\ Ab_1 & Ab_2 & Ab_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ 2b_1 & -3b_2 & 7b_3 \\ | & | & | \end{bmatrix}$$

This means that $Ab_1 = 2b_1$, $A_b2 = -3b_2$, $Ab_3 = 7b_3$, so that $B = \{b_1, b_2, b_3\}$ is a basis of $\mathbb{R}^3$ consisting of eigenvectors of A.

3.22 Eigenvalues and diagonalizability

Definition 3.10

An **eigenvector** of a square matrix A is a non zero column vector v, for which there exists a scalar λ such that:

$$Av = \lambda v$$

The scalar λ is called the **eigenvalue** of A corresponding to the eigenvector v.

The eigenvalues of A are the roots of its **characteristic polynomial**:

$$\det(xI_n - A) = 0$$

The **eigenspace** corresponding to a particular eigenvalue, λ, is the set of all vector, v, satisfying: $Av = \lambda v$. It is a subspace of the relevant $\mathbb{R}^n$ of dimension at least 1.

The matrix $A \in M_n(\mathbb{R})$ is **diagonalizable** only if $\mathbb{R}^n$ has a basis consisting of eigenvectors of A. In this case $P^{-1}AP$ is diagonal, where P is a matrix whose n columns are linearly independent eigenvectors of A. The diagonal entries of $P^{-1}AP$ are the corresponding eigenvalues.

3.23 Two examples of non-diagonalizabilty

For $A \in M_n(\mathbb{R})$, it does not always happen that $\mathbb{R}^n$ has a basis consisting of eigenvectors of A.

Example 1:

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

The matrix A is diagonalizable in $M_2(\mathbb{C})$, but not in $M_2(\mathbb{R})$. The matrix represents an anti-clockwise rotation of $90°$ in $\mathbb{R}^2$. It does not fix any line in $\mathbb{R}^2$. Its characteristic polynomial is:

$$\det(xI_2 - A) = \det \begin{bmatrix} x & 1 \\ -1 & x \end{bmatrix} = x^2 + 1 = 0$$

Example 2:

$$B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

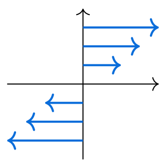
The matrix B is not diagonalizable in $M_2(\mathbb{C})$. The matrix represents **horizontal shear**. Its characteristic polynomial is:

$$\det(xI_2 - B) = \det \begin{bmatrix} x-1 & -1 \\ 0 & x-1 \end{bmatrix} = (x-1)^2 = 0$$

but its 1-dimensional eigenspace only consists of the X-axis. It does not have two linearly independent eigenvectors.

3.24 A Shear in $\mathbb{R}^2$

$$B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$



The linear transformation T described by B send $(x, y) \in \mathbb{R}^2$ to $(x + y, y)$. This is a **horizontal shear**: it shifts every point horizontally by its y-coordinate.

For every point $v \in \mathbb{R}^2$, $T(v)$ is on the same horizontal line as v. It follows that $T(v)$ is a scalar multiply of v, only if v lies on the X-axis; in this case, $T(v) = v$.

The characteristic polynomial of B (and T) is:

$$\det(xI_2 - B) = \det \begin{bmatrix} x-1 & -1 \\ 0 & x-1 \end{bmatrix} = (x-1)^2 = (\lambda - 1)^2 = 0$$

The only eigenvalue is $\lambda = 1$, and it has **algebraic multiplicity** 2, meaning it appears twice as a root of the characteristic polynomial.

Its **geometric multiplicity** is only 1, meaning its corresponding eigenspace is only 1-dimensional - just the line $y = 1$.

3.25 Eigenvector for distinct eigenvalues are independent

Theorem 3.4

Let $A \in M_n\mathbb{R}$ and $v_1, \ldots, v_k$ be eigenvector of A in $\mathbb{R}^n$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ of A.

Then $\{v_1, \ldots, v_k\}$ is a linearly independent subset of $\mathbb{R}^n$.

Proof (for $k = 3$): Note that no two v_1, v_2, v_3 are multiples of each other, since they correspond to distinct eigenvalues.

Suppose: $a_1v_1 + a_2v_2 + a_3v_3 = 0$ for some $a_i \in \mathbb{R}$. We want to show that $a_1 = a_2 = a_3 = 0$.
Multiplying on the left by A gives:

$$a_1Av_1 + a_2Av_2 + a_3Av_3 = 0 \implies a_1\lambda_1v_1 + a_2\lambda_2v_2 + a_3\lambda_3v_3 = 0$$

and multiplying the same equation by λ_1 gives:

$$a_1\lambda_1v_1 + a_2\lambda_1v_2 + a_3\lambda_1v_3 = 0$$

Subtract to get:

$$a_2(\lambda_1 - \lambda_2)v_2 + a_3(\lambda_1 - \lambda_3)v_3 = 0$$

Since v_2 and v_3 are linearly independent, and $lambda_1 - \lambda_2 \neq 0$ and $\lambda_1 - \lambda_3 \neq 0$, it follows that $a_2 = a_3 = 0$ and hence $a_1 = 0$.

3.26 At most n distinct eigenvalues

The following consequence of the previous theorem shows that a matrix can have at most n distinct eigenvalues. We already knew this, since the eigenvalues are roots of a polynomial of degree n.

Theorem 3.5: Corollary

Let $A \in M_N(\mathbb{R})$. Then A has at most n distinct eigenvalues in $\mathbb{R}$.

Proof:
If A has k distinct eigenvalues, with corresponding eigenvectors $v_1, \ldots, v_k \in \mathbb{R}^n$, then k cannot exceed the dimension of $\mathbb{R}^n$, since $\{v_1, \ldots, v_k\}$ is a linearly independent set in $\mathbb{R}^n$.

Theorem 3.6: Corollary

If $A \in M_n(\mathbb{R})$ has n distinct eigenvalues, then A is diagonalizable.

Proof:
A set consisting of one eigenvector for each of the n eigenvalues is linearly independent, so hence is a basis.

3.27 Notes about determinants and characteristic polynomials

1. The **characteristic polynomial** of the square matrix $A \in M_n(\mathbb{R})$ is the determinant of $\lambda I_n - A$.
2. If t is a root of this polynomial, the t-eigenspace of A is the nullspace of the matrix $tI_n - A$.
3. The determinant of a diagonal or upper triangular matrix is the product of the entries on its main diagonal.
4. A square matrix is **block diagonal** if its non-zero entries are all contained in square blocks along its diagonal. The determinant of a block diagonal matrix is the product of the determinants of its diagonal blocks.
5. Similar matrices have the same characteristic polynomial and the same eigenvalues and eigenspace dimensions, since they represent the same linear transformation.

3.28 Multiplicity of Eigenvalues

Suppose that A and B are similar square matrices, so that $B = P^{-1}AP$ for some invertible matrix P. Then

$$\begin{aligned}
\det(xI - B) &= \det(xI - P^{-1}AP) \\
&= \det(P^{-1}(xI)P - P^{-1}AP) \\
&= \det\left(P^{-1}(xI - A)P\right) \\
&= \det P^{-1} \det(xI - A) \det P \\
&= \det(xI - A).
\end{aligned}$$

If two matrices have the same characteristic polynomial, they are not necessarily similar. For example $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ have the same characteristic polynomial but are not similar.

3.29 Multiplicity of Eigenvalues

Let λ be an eigenvalue of a matrix $A \in M_n(\mathbb{R})$. The **algebraic multiplicity** of λ is the number of times that λ occurs as a root of the characteristic polynomial. The **geometric multiplicity** is the dimension of the t-eigenspace of A.

Example 3.6

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

The matrix has two distinct eigenvalues, 3 and 4. Both have algebraic multiplicity 2; the characteristic polynomial is $(\lambda - 3)^2(\lambda - 4)^2$. The 3-eigenspace has dimension 2, its elements are $\begin{bmatrix} a \\ b \\ 0 \\ 0 \end{bmatrix}$, for $a, b \in \mathbb{R}$.

The 4-eigenspace only has dimension 1, its elements are $\begin{bmatrix} 0 \\ 0 \\ c \\ 0 \end{bmatrix}$, for $c \in \mathbb{R}$.

This A is not diagonalizable since it does not have four independent eigenvectors.

3.30 Geometric multiplicity $\leq$ Algebraic multiplicity

Theorem 3.7

The geometric multiplicity of an eigenvalue is at most equal to its algebraic multiplicity.

Proof:
Suppose that t has geometric multiplicity k as an eigenvalue of the square matrix $A \in M_n(\mathbb{R})$, and let $\{v_1, \ldots, v_k\}$ be a basis for the t-eigenspace of A.

Extend this to a basis $\mathcal{B}$ of $\mathbb{R}^n$, and let P be the matrix whose columns are the elements of $\mathcal{B}$. Then the first k columns of $P^{-1}AP$ have t in the diagonal position and zeros elsewhere.

It follows that $\lambda - t$ occurs at least k times as a factor of $\det(\lambda I_n - P^{-1}AP)$, so the algebraic multiplicity of t is at least k.

Theorem 3.8: Corollary

A matrix is diagonalizable if and only if the geometric multiplicity of each of its eigenvalues is equal to the algebraic multiplicity

3.31 Inner Products

In $\mathbb{R}^n$, the dot product of the vectors $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ is defined as:

$$x \cdot y = x_1 y_1 + x_2 y_2 = x^T = y^T = y \cdot x$$

When can interpret the length $(||x||)$ of x as the length of the line segment from origin to (x_1, x_2), which by the Pythagorean theorem is:

$$\sqrt{x_1^2 + x_2^2} = \sqrt{x \cdot x}$$

Once we have a concept of length, we can define distance $(d(x, y))$ between two vectors x, y as the length of the line segment from x to y, which is:

$$d(x, y) = ||x - y||$$

Similarly, from the Cosine Rule:

$$x \cdot y = ||x|| \cdot ||y|| \cdot \cos(\theta)$$

where θ is the angle between x and y.
We say that x and y are **orthogonal** (or $x \perp y$) if $x \cdot y = 0$.

So the scalar product encodes geometric information in $\mathbb{R}^2$, and also provides a way to define length, distance and orthogonality. The scalar product is an example of an **inner product**.

3.32 Real Inner Products

A **Inner Product** on a vector space V is a function from $V \times V$ to $\mathbb{R}$ that assigns an element of $\mathbb{R}$ to every ordered pair of elements of V, and has the following properties.

1. **Symmetry**: $\langle x, y\rangle = \langle y, x\rangle$ for all $x, y \in V$
2. **Linearity in both slots (bilinearity)**: For all $x, y, z \in V$ and all $a, b \in \mathbb{R}$, we have $\langle ax + by, z\rangle = a\langle x, z\rangle + b\langle y, z\rangle$ and $\langle x, ay + bz\rangle = a\langle x, y\rangle + b\langle x, z\rangle$.
3. **Non-negativity**: $\langle x, x\rangle \geq 0$ for all $x \in V$, and $\langle x, x\rangle = 0$ only if $x = 0_V$.

The ordinary scalar product on $\mathbb{R}^n$ is the best known example of an inner product.

3.32.1 Examples of Inner Products

1. The ordinary scalar product on $\mathbb{R}^n$.
2. Let C be the vector space of all continuous real-valued functions on the interval $[0, 1]$. The analogue of the ordinary scalar product on C is the inner product given by
$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx, \quad \text{for } f, g \in C.$$
3. On the space $M_{m\times n}(\mathbb{R})$, the *Frobenius inner product* or *trace inner product* is defined by
$$\langle A, B\rangle = \mathrm{trace}(A^T B).$$
Note that $\mathrm{trace}(A^T B)$ is the sum over all positions (i, j) of the products $A_{ij}B_{ij}$. So this is closely related to the ordinary scalar product, if the matrices A and B were regarded as vectors with mn entries over $\mathbb{R}$.

It is possible for a single vector space to have many different inner products defined on it, and if there is any risk of ambiguity we need to specify which one we are considering.

3.33 Length, Distance and Scaling in an Inner Product Space

Definition 3.11

We define the **length** or **norm** of any vector v by:
$$||v|| = \sqrt{\langle v, v\rangle}$$

Definition 3.12

We define the distance between two vectors, u and v, by:
$$d(u, v) = ||u - v||$$

Concept 3.2: Scaling

Every vector, v, and scalar, c, satisfy:
$$||cv|| = |c| \cdot ||v||$$
Since:
$$||cv|| = \sqrt{\langle cv, cv\rangle} = \sqrt{c^2\langle v, v\rangle} = |c| \cdot ||v||$$
So we can adjust the norm of any element of V while preserving its direction, by multiplying by a scalar.

Definition 3.13

If v is a non-zero vector in an inner product space V, then
$$\hat{v} := \frac{1}{||v||}v$$
is a unit vector in the direction of v, called the **normalization** of v.

3.34 Orthogonality in an inner product space

Let V be a vector space with an inner product $\langle\cdot,\cdot\rangle$ (such as the ordinary scalar product).

Definition 3.14

We say that the vectors u and v are **orthogonal** (with respect to $\langle\cdot,\cdot\rangle$) if $\langle u, v\rangle = 0$.

All these definitions are consistent with "typical" geometrically motivated concepts of distance and orthogonality.

Examples

1. $(2, 5)$ and $(5, -2)$ are orthogonal with respect to the ordinary scalar product in $\mathbb{R}^2$.
2. $\sin \pi x$ and $\cos \pi x$ are orthogonal with respect to the scalar product on the space of continuous functions on $[0, 1]$ defined in Lecture 18; this is saying that

$$\int_0^1 \sin(\pi x)\cos(\pi x)\,dx = 0 \qquad \left(= \frac{1}{2\pi}\sin^2(\pi x)\Big|_0^1 \right).$$

3.35 Orthogonal Projection

Lemma 3.3

Let u and v be non-zero vectors in an inner product space V. Then it is possible to write (in a unique way) $v = au + v'$, where a is a scalar and v' is orthogonal to u.

- If v is orthogonal to u, take $a = 0$ and $v' = v$.
- If v is a scalar multiple of u, take $au = v$ and $v' = 0$.
- Otherwise, to solve for a and v' in the equation $v = au + v'$ (with $u \perp v'$), take the inner product with u on both sides. Then

$$\langle u, v\rangle = a\langle u, u\rangle + 0 \Longrightarrow a = \frac{\langle u, v\rangle}{\|u\|^2}, \quad v' = v - \frac{\langle u, v\rangle}{\|u\|^2}u.$$

We can verify directly that the two components in this expression are orthogonal to each other.

Example

$$\text{Write } u = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad v = \begin{bmatrix} 6 \\ -2 \end{bmatrix} \quad \text{Then } u = \begin{bmatrix} \frac{3}{2} \\ \frac{3}{4} \end{bmatrix} + \begin{bmatrix} \frac{3}{2} \\ -\frac{11}{4} \end{bmatrix}$$

3.36 Orthogonal projection of one vector on another

Definition 3.15

For non-zero vectors u and v in an inner product space V, the vector

$$\frac{\langle u, v\rangle}{\|u\|^2}u$$

is called the projection of v on the 1-dimensional space spanned by u.
It is denoted by $\mathrm{proj}_u(v)$ and it has the property that $v - \mathrm{proj}_u(v)$ is orthogonal to u

Lemma 3.4

$\mathrm{proj}_u(v)$ is the unique element of $\langle u\rangle$ whose distance from v is minimal.

Proof: Let au be a scalar multiple of u. Then

$$d(au, v)^2 = \langle au - v, au - v\rangle = a^2\langle u, u\rangle - 2a\langle u, v\rangle + \langle v, v\rangle$$

Regarded as a quadratic function of a, this has a minimum when its derivative is 0, i.e. when

$$2a\langle u, u\rangle - 2\langle u, v\rangle = 0, \quad \text{when } a = \frac{\langle u, v\rangle}{\|u\|^2}.$$

3.37 Orthogonal Bases (the Gram-Schmidt process)

Concept 3.3

Every finite-dimensional inner product space has an **orthogonal basis**

We can start with any basis $\{b_1, \ldots, b_n\}$, and adjust the elements one by one (by subtracting off orthogonal projections of later vectors on earlier ones). The process ends with an orthogonal basis $\{v_1, \ldots, v_n\}$.

1. Set $v_1 = b_1$, and

$$v_2 = b_2 - \mathrm{proj}_{v_1}(b_2) = b_2 - \frac{\langle v_1, b_2\rangle}{\langle v_1, v_1\rangle}v_1.$$

Then the pairs b_1, b_2 and v_1, v_2 span the same space, and $v_1 \perp v_2$.

2. Write $v_3 = b_3 - \mathrm{proj}_{v_1}(b_3) - \mathrm{proj}_{v_2}(b_3)$. Then $\{v_1, v_2, v_3\}$ and $\{b_1, b_2, b_3\}$ span the same space, and $v_3 \perp v_1$, $v_3 \perp v_2$. To see this look at $\langle v_3, v_1\rangle$ and $\langle v_3, v_2\rangle$, noting that

$$v_3 = b_3 - \frac{\langle b_3, v_1\rangle}{\langle v_1, v_1\rangle} v_1 - \frac{\langle b_3, v_2\rangle}{\langle v_2, v_2\rangle} v_2.$$

3. Continue: at the kth step, form v_k by subtracting from b_k its projections on $v_1, \ldots, v_{k-1}$.

3.38 Orthogonal projection on a subspace

The result of this process is a basis $\{v_1, \ldots, v_n\}$ whose elements satisfy

$$\langle v_i, v_j\rangle = 0 \text{ for } i \neq j$$

We can adjust this basis to a **orthonormal basis** (consisting of orthogonal unit vectors) by replacing each v_i with its normalization $\hat{v}_i$. From the Gram–Schmidt process, we have

Theorem 3.9

If V is a finite-dimensional inner product space, then V has an orthogonal (or orthonormal) basis.

Now let W be a subspace of V, and let $v \in V$. The **orthogonal projection** of v on W, denoted $\mathrm{proj}_W(v)$, is defined to be the unique element u of W for which

$$v = u + v',$$

and $v' \perp w$ for all $w \in W$.

3.39 Calculating the projection on a subspace

Example 3.7: I

$\mathbb{R}^3$, find the unique point of the plane $W : x + 2y - z = 0$ that is nearest to the point $v : (1, 2, 2)$.

First find an orthogonal basis for W: for example $\{b_1, b_2\}$, where

$$b_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$$

Then

$$\begin{aligned} \mathrm{proj}_W(v) &= \mathrm{proj}_{b_1}(v) + \mathrm{proj}_{b_2}(v) \\ &= \frac{\langle b_1, v\rangle}{\langle b_1, b_1\rangle} b_1 + \frac{\langle b_2, v\rangle}{\langle b_2, b_2\rangle} b_2 \\ &= \frac{3}{2} b_1 - \frac{3}{3} b_2 \\ &= \left(\frac{3}{2}, 0, \frac{3}{2}\right) - (1, -1, -1) = \left(\frac{1}{2}, 1, \frac{5}{2}\right) \end{aligned}$$

3.40 proj$_{(v)}$ is the nearest point of W to v

Let $u = \mathrm{proj}_W(v)$ and let w be any element of W. Note that $v - u$ is orthogonal to both w and u, hence to $w - u$. Then

$$\begin{aligned} d(v, w)^2 &= \langle v - w, v - w\rangle \\ &= \langle (v - u) + (u - w), (v - u) + (u - w)\rangle \\ &= \langle v - u, v - u\rangle + 2\langle v - u, u - w\rangle + \langle u - w, u - w\rangle \\ &= \langle v - u, v - u\rangle + \langle u - w, u - w\rangle \\ &\geq d(v, u)^2, \end{aligned}$$

with equality only if $w = u = \mathrm{proj}_W(v)$.

3.41 Recall

Let u and v be elements in an inner product space V. The **projection of v on u** is

$$\boxed{\operatorname{proj}_u v = \frac{\langle u, v\rangle}{\langle u, u\rangle} u}$$

Then $\operatorname{proj}_u v$ is the unique nearest element to v of the 1-dimensional subspace spanned by u (in the distance determined by the inner product), and

$$\boxed{\langle v', u\rangle = 0, \quad v' \perp u, \quad \text{where } v' = v - \operatorname{proj}_u v.}$$

So v is the sum of its projection on u and a component orthogonal to u.

In general, if W is any subspace of V and $v \in V$, then there is a unique nearest element in W to v, called $\operatorname{proj}_W v$ or the **projection of v on W**, and $v - \operatorname{proj}_W v$ is orthogonal to **every element of W**.