Computational Linear Algebra
EK103

Giacomo Cappelletto

21/1/2

Contents

Chapter 1

Basics

1.1 Vectors, Norms and Products

Note:-

Let us consider two vectors in $\mathbb{R}^3$:

$$u = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

We wish to compute their magnitudes (norms and norm-squared), the angle between them, and the plane that they span. These methods are directly applicable to computational tools such as MATLAB.

Definition 1.1.1: Norm of a Vector

For a vector $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, its norm is

$$\|x\| \;=\; \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

In many programming languages (including MATLAB), this is computed via `norm(x)`, while the square of the norm is $\|x\|^2 = x \cdot x = x_1^2 + \cdots + x_n^2$.
Norm squared is the result of the dot product of a vector with itself. For example, the norm squared of x is

$$\|x\|^2 \;=\; x \cdot x = x_1^2 + \cdots + x_n^2 = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}.$$

Example 1.1.1 (Norms and Norm-Squared of u and v)

$$\|u\| \;=\; \sqrt{1^2 + 1^2 + 1^2} \;=\; \sqrt{3}, \quad \|v\| \;=\; \sqrt{1^2 + (-1)^2 + 1^2} \;=\; \sqrt{3}.$$

Thus, both vectors have the same magnitude $\sqrt{3}$. Their squared norms are

$$\|u\|^2 \;=\; 3, \quad \|v\|^2 \;=\; 3.$$

In MATLAB notation, one could write:

- `norm(u)` or `norm(u,2)` for the norm of u.
- `dot(u,u)` or `norm(u)^2` for $\|u\|^2$.

Definition 1.1.2: Angle Between Two Vectors

The angle θ between two nonzero vectors u and v in $\mathbb{R}^n$ is given by

$$\theta \;=\; \arccos\Bigl(\frac{u \cdot v}{\|u\|\|v\|}\Bigr).$$

Example 1.1.2 (Angle Between u and v)
First, compute the dot product:

$$u \cdot v \;=\; (1)(1) + (1)(-1) + (1)(1) \;=\; 1 - 1 + 1 \;=\; 1.$$

Hence,

$$\theta \;=\; \arccos\Bigl(\frac{u \cdot v}{\|u\|\|v\|}\Bigr) \;=\; \arccos\Bigl(\frac{1}{\sqrt{3}\,\sqrt{3}}\Bigr) \;=\; \arccos\bigl(\tfrac{1}{3}\bigr).$$

In MATLAB, one could write:

```
theta = acos(dot(u,v)/(norm(u)*norm(v)));
```

Definition 1.1.3: Plane Spanned by Two Vectors

The plane containing vectors u and v and passing through the origin is given by

$$\{\; \alpha\, u + \beta\, v \;\mid\; \alpha, \beta \in \mathbb{R}\}.$$

An equivalent description is all points $x \in \mathbb{R}^3$ such that $x \cdot (u \times v) = 0$.

Example 1.1.3 (Plane Containing u and v)

- *Span form:*

$$\text{Plane} = \Bigl\{\, \alpha \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \;+\; \beta \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \;\Big|\; \alpha, \beta \in \mathbb{R} \Bigr\}.$$

- *Normal form:* The cross product

$$u \times v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 1 \\ 1 & -1 & 1 \end{vmatrix} = (2,\, 0,\, -2).$$

  Hence, the plane also can be described by the set of points $x = (x_1, x_2, x_3)$ for which

$$(2,\, 0,\, -2) \cdot (x_1, x_2, x_3) = 0 \quad \Longrightarrow \quad 2\,x_1 - 2\,x_3 = 0 \quad \Longrightarrow \quad x_1 = x_3.$$

In many computational environments, one simply keeps the span form or uses a symbolic package to compute the cross product and normal equation.

Definition 1.1.4: Cross Product

Construct a system of linear equations where the dot product of the vector is orthogonal to both the vectors in the matrix.

$$\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Example 1.1.4 (Finding the Plane Spanned by Two Vectors)

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

Then, we must formulate an equation for any vector perpendicular to the normal of the plane, i.e. the cross product of the two original vectors.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = 0$$

Hence,

$$y_1 - y_3 = 0$$

Definition 1.1.5: Dot product

We can take a vector $\vec{v}$ in $\mathbb{R}^n$ and a vector $\vec{w}$ in $\mathbb{R}^n$. Then, the dot product of $\vec{v}$ and $\vec{w}$ is defined as

$$\vec{v} \cdot \vec{w} = \vec{v}^T \vec{w} = \vec{w}^T \vec{v}$$

Therefore

$$\|\vec{v}\|^2 = \vec{v}^T \vec{v}$$

Definition 1.1.6: Scalar Multiplication

Scalar multiplication is the operation of multiplying a vector by a scalar. The result is a new vector with the same direction as the original vector, but with a magnitude that is the product of the original magnitude and the scalar.

$$t \cdot \vec{v} = t \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} t\, a_1 \\ t\, a_2 \\ \vdots \\ t\, a_n \end{bmatrix}$$

Where

$$\|t \cdot \vec{v}\| = \|\vec{v}\| \cdot t$$

Definition 1.1.7: Vector Addition

Vector addition is the operation of adding two vectors together. The result is a new vector that is the sum of the two original vectors.

$$\vec{v} + \vec{w} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix}$$

Definition 1.1.8: Matrix to Vector Multiplication

Matrix to vector multiplication is the operation of multiplying a matrix by a vector. The result is a new vector that is the result of the matrix-vector multiplication. Matrices are represented as $n \times m$, where n is the number of rows and m is the number of columns. The number of columns of the matrix must be equal to the number of rows of the vector.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix}$$

Definition 1.1.9: Matrix to Matrix Multiplication

Matrix to matrix multiplication is the non-commutative operation of multiplying two matrices together. The result is a new matrix that is the product of the two original matrices.
For two matrices A and B to be multiplied, the number of columns of A must be equal to the number of rows of B. If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then their product $C = AB$ is an $m \times p$ matrix. The element c_{ij} of the resulting matrix C is computed as:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

where a_{ik} is the element from the i-th row and k-th column of matrix A, and b_{kj} is the element from the k-th row and j-th column of matrix B.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix}$$

Example 1.1.5 (Matrix to Matrix Multiplication)
Consider the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}.$$

Their product $C = AB$ is computed as follows:

$$C = \begin{bmatrix} 1\cdot 7 + 2\cdot 10 & 1\cdot 8 + 2\cdot 11 & 1\cdot 9 + 2\cdot 12 \\ 3\cdot 7 + 4\cdot 10 & 3\cdot 8 + 4\cdot 11 & 3\cdot 9 + 4\cdot 12 \\ 5\cdot 7 + 6\cdot 10 & 5\cdot 8 + 6\cdot 11 & 5\cdot 9 + 6\cdot 12 \end{bmatrix} = \begin{bmatrix} 27 & 30 & 33 \\ 61 & 68 & 75 \\ 95 & 106 & 117 \end{bmatrix}.$$

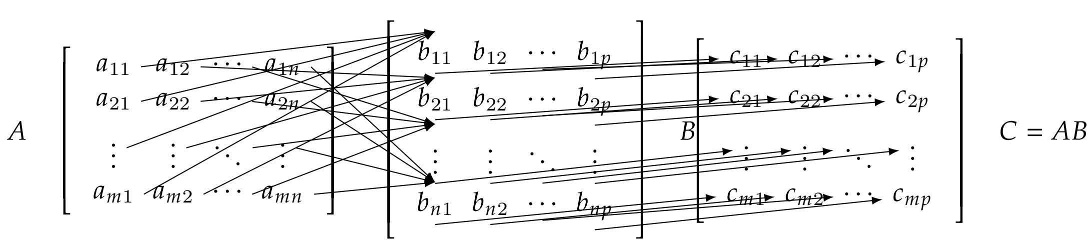


Figure 1.1: Matrix-Matrix Multiplication Diagram

Definition 1.1.10: Matrix Multiplication of a Matrix with Itself

When a matrix A is multiplied by its transpose A^T, the resulting matrix is a symmetric matrix. The element c_{ij} of the resulting matrix $C = AA^T$ is computed as:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} a_{jk}$$

where a_{ik} is the element from the i-th row and k-th column of matrix A, and a_{jk} is the element from the j-th row and k-th column of matrix A.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{bmatrix}$$

The zeros in this matrix represent orthogonality between the corresponding rows of the original matrix A. Specifically, if $c_{ij} = 0$, it means that the i-th row and the j-th row of matrix A are orthogonal to each other.

Example 1.1.6 (Orthogonality in Matrix Multiplication)
Consider the matrix

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

Its transpose is

$$A^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The product $C = AA^T$ is computed as follows:

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Notice that the resulting matrix is the identity matrix, which is symmetric and has zeros in all off-diagonal elements, indicating that the rows (and columns) of the original matrix A are orthogonal to each other.

1.1.1 Interpretation of vectors in $\mathbb{R}^{2,3}$

Definition 1.1.11: Position Vector

A position vector is a vector that describes the position of an object in space with reference to an origin.

Definition 1.1.12: Translational Vector

A translational vector is a vector that describes the displacement of an object in space with reference to an origin.

1.2 Rotation Matrices

Note:-

We are considering vectors in $\mathbb{R}^2$, denoted by

$$\vec{v} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

We also have a 2×2 matrix (linear operator or transformation) given by

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

Note:-

Applying A to an input vector $\vec{v}_i$ produces the output vector $\vec{v}_o$:

$$A\,\vec{v}_i = \vec{v}_o.$$

Explicitly,

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

Question 1: Transforming a region

Suppose we restrict the input vectors $\vec{v}_i$ to the square region in the (x_1, x_2)-plane with coordinates

$$0 \leqslant x_1 \leqslant 1, \quad 0 \leqslant x_2 \leqslant 1.$$

- How does A map this square region in the input space to a region in the (y_1, y_2)-plane?
- Geometrically, what does that image region look like?

Solution: We can answer this question *mechanically* by observing what happens to the corners of the square, or by spanning the space with two basis vectors:

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Each point in the square can be written as

$$x_1\,\vec{e}_1 + x_2\,\vec{e}_2 \quad \text{with } 0 \le x_1, x_2 \le 1.$$

Applying A to these basis vectors, we get

$$A\,\vec{e}_1 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad A\,\vec{e}_2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Hence, the image of the unit square (spanned by $\vec{e}_1$ and $\vec{e}_2$) is the parallelogram spanned by

$$A\,\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad A\,\vec{e}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

In other words, every point (x_1, x_2) in the original square is mapped to

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 + 2\,x_2 \\ x_2 \end{bmatrix},$$

with $0 \le x_1, x_2 \le 1$. Geometrically, this results in a parallelogram in the (y_1, y_2)-plane whose vertices are $(0,0)$, $(1,0)$, $(1+2,1)$, and $(2,1)$.

Note:-

Thus, restricting $\vec{v}_i$ to a square region in the input space restricts $\vec{v}_o$ to a parallelogram in the output space.

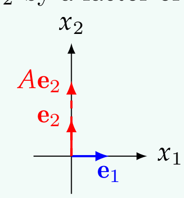
Algorithm 1: Mapping a unit square under the linear transformation A

Input: Input vector $\vec{v}_i = [x_1\ x_2]^T$ with $0 \le x_1, x_2 \le 1$
Output: Output vector $\vec{v}_o = [y_1\ y_2]^T$ in the parallelogram
/* Matrix A: */
$A \leftarrow \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$;
/* Apply A to input: */
$y_1 \leftarrow x_1 + 2\,x_2$;
$y_2 \leftarrow x_2$;
return $\vec{v}_o$;

1.2.1 Examples of 2D Matrix Transformations

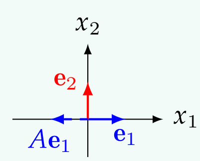
Example 1.2.1 ($A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$)

Transformation: Vertical stretch scaling x_2 by a factor of 2.



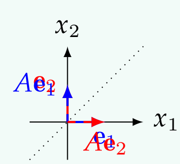
Example 1.2.2 ($A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$)

Transformation: Reflection about the vertical axis.

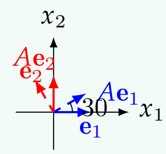


Example 1.2.3 ($A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$)

Transformation: Reflection about the line $x_1 = x_2$.

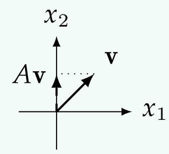


Example 1.2.4 ($A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$)



Transformation: Counterclockwise rotation by θ about the origin.

Example 1.2.5 ($A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$)



Transformation: Projection onto the vertical axis.

1.3 Mapping

$$[x_1,\ x_2] \longmapsto \text{point } (x_1, x_2).$$

Under the linear transformation given by

$$[x_1,\ x_2] \mapsto \begin{bmatrix} x_1 + 2x_2 \\ x_2 \end{bmatrix},$$

we see that points are "stretched to the right" in the x_1-direction, while the x_2 coordinate remains unchanged.

1.3.1 Method 1 – Vector Method

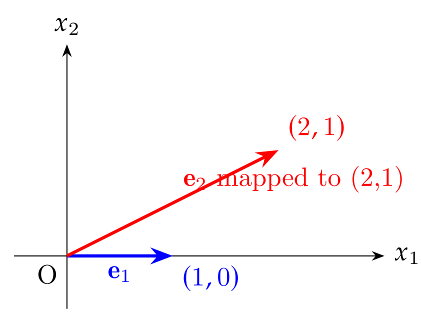
The matrix under consideration is

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

The columns of this matrix are:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

These are the images of the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$, respectively.



1.3.2 Method 2 – Vertex Method

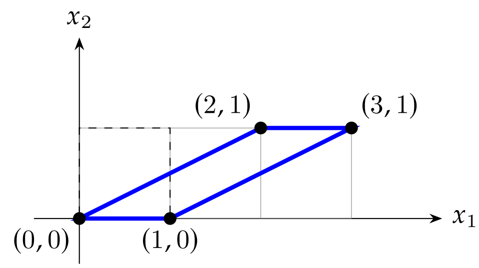
Consider the unit square with vertices

$$(0, 0), \quad (1, 0), \quad (1, 1), \quad (0, 1).$$

We apply A to each vertex:

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

Hence the new vertices are $(0,0)$, $(1,0)$, $(2,1)$, and $(3,1)$. Plotting these points yields the transformed parallelogram (the image of the original unit square).



Definition 1.3.1: Rules for Linear Transformations

1. **Straight lines remain straight.**
2. **Parallel lines remain parallel.**
3. **Distances along lines scale in a consistent, proportional way.**

Justification

1. Straight lines stay straight. A line in parametric form is:

$$r(t) = t\,\mathbf{v} + \mathbf{w}.$$

Applying A gives:

$$A(r(t)) = A(t\,\mathbf{v} + \mathbf{w}) = t\,A(\mathbf{v}) + A(\mathbf{w}),$$

which is again a parametric line.

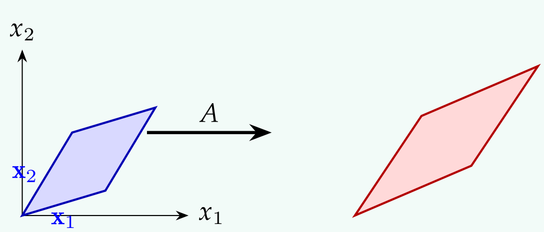
2. Parallel lines stay parallel. If two lines are parallel, their direction vectors are scalar multiples of each other. After applying A, the resulting direction vectors are $A(\mathbf{v})$ for each original direction $\mathbf{v}$. Since A is linear, any scalar multiples remain so, preserving parallelism.

3. Distances scale proportionally. For two points $\mathbf{v}_1$ and $\mathbf{v}_2$, the difference is $\mathbf{d} = \mathbf{v}_2 - \mathbf{v}_1$. Under A:

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

Thus the new distance $\|\mathbf{d}'\| = \|A(\mathbf{d})\|$ is a consistent transform of $\|\mathbf{d}\|$, depending on the nature of A.

Example 1.3.1 (A matrix A that acts on a parallelogram (spanned by two vectors $\mathbf{x}_1, \mathbf{x}_2$) will produce another parallelogram in the output plane.)



Here, lines remain lines, parallels remain parallel.

1.4 A System of Linear Equations

We have a matrix A of size $m \times n$, multiplying an unknown vector $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ (which belongs to $\mathbb{R}^n$), producing a result vector $\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$ in $\mathbb{R}^m$:

$$A_{m\times n} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}.$$

This can be viewed as a *system of m linear equations in n unknowns*. We are often interested in two main questions:

1. **Does a solution exist?** That is, can we find $x_1, x_2, \ldots, x_n$ so that $A\mathbf{x} = \mathbf{v}$? Geometrically, this asks if $\mathbf{v}$ lies in the *column space* (or image) of A.

2. **If at least one solution exists, is it unique or are there infinitely many?** Uniqueness is typically tied to whether the columns of A are linearly independent (and whether m, n are related in a way that gives a single solution). If there are *fewer* pivots than unknowns, or if the system is underdetermined, infinitely many solutions can occur.

Key intuition:

- The question of *existence* boils down to whether $\mathbf{v}$ is in the span of the columns of A.
- The question of *uniqueness* depends on whether those columns form a set of independent vectors and on the relationship between m and n.

1.5 How to Determine Consistency, Uniqueness, and the Number of Solutions for a Linear System

Note:-

We consider a linear system

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

and write it in *augmented matrix* form:

$$\begin{bmatrix} A \mid \mathbf{b} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}.$$

- We perform *elementary row operations* (EROs) on $\begin{bmatrix} A \mid \mathbf{b} \end{bmatrix}$ to attempt to solve the system. The EROs are:

  1. **Scaling (multiplying a row by a nonzero constant).**

2. **Row replacement** ($R_i = R_i + c\,R_j$ for $i \neq j$).
3. **Interchanging two rows** ($R_i \leftrightarrow R_j$).

These operations do not change the solution set of the system.

Row Echelon Form (REF) We say $\left[A \mid \mathbf{b}\right]$ is in *row echelon form* if:

- All rows of all zeros (if any) are at the bottom of the matrix.
- Each *pivot* (leftmost nonzero entry in a nonzero row) is strictly to the right of the pivot in the row above.

A pivot column in A corresponds to a *leading variable* (or *basic variable*), and any other column (apart from the augmented column) is called a *free column* (its corresponding variable is a *free variable*).
Consistency & Number of Solutions

- If, in the augmented matrix, there is a pivot in the last column (meaning a row of the form

$$[0 \quad 0 \quad \cdots \quad 0 \mid c], \quad c \neq 0,$$

) then the system is *inconsistent* (no solutions). This corresponds to an equation $0 = c$ where $c \neq 0$, which is impossible.

- If no such contradiction is found, then at least one solution exists (the system is *consistent*).
- The number of pivot columns in A (i.e. the number of leading variables) tells us whether solutions are unique or infinite:
    1. If the number of pivots equals the number of unknowns n, then there is exactly one solution (assuming no inconsistency).
    2. If the number of pivots is less than n, then there are free variables, implying infinitely many solutions (again, assuming no inconsistency).

Example 1.5.1 (Example: Augmented Matrix and REF)

$$\begin{bmatrix} 1 & 2 & 3 & 5 \\ 2 & 4 & 6 & 9 \\ 3 & 6 & 9 & 15 \end{bmatrix} \longrightarrow \text{(REF)}.$$

One performs row operations to get an upper-triangular or echelon form. If the last row becomes something like

$$[0 \;\; 0 \;\; 0 \mid c], \quad c \neq 0,$$

then there is no solution. Otherwise, we identify pivot columns, read off the relationships among variables, and find a general solution (unique or infinite).

Definition 1.5.1: Free Columns and Basic Columns

Once an augmented matrix is in REF, we label each pivot column (in A) as a *basic column*, and any other column (except the last augmented column) as a *free column*.

- If x_j corresponds to a free column, we may label x_j as a parameter (e.g. $t_1, t_2, \ldots$).
- The variables in pivot columns can then be written in terms of these parameters.

In this manner, we get a general solution describing the entire solution set to $A\mathbf{x} = \mathbf{b}$.

Algorithm to Convert $\left[A \mid \mathbf{b}\right]$ to $\mathrm{REF}(A)$

1. *Select a candidate row*: Choose the topmost row *among those not yet having a pivot* in which a pivot might appear.

2. *Pivot search and possible row swap*: Among this candidate row and those below it, find a row having a leftmost nonzero entry in the desired pivot column. Interchange (swap) that row with the candidate row if needed, placing a nonzero entry where your pivot should be.

3. *Declare the pivot and eliminate below*: Scale the pivot row (if desired) so that the pivot becomes 1. Then use row replacement to produce zeros below that pivot in the same column.

4. Move to the next row down and next column to the right, and repeat until you have a row echelon form.

One can then further use row replacement operations to clear the entries *above* each pivot, yielding the *reduced* row echelon form (RREF). However, for most solution purposes, REF is already sufficient to read off whether solutions exist, how many, and so on.

1.5.1 FRR

Algorithm 2: Forward Row Reduction (Forward Elimination)

Input: Matrix A of size $m \times n$ and vector $\mathbf{b}$ of size m

Output: Upper triangular matrix A and modified vector $\mathbf{b}$

```
/* Forward elimination process */
for k ← 1 to min(m, n) do
    for i ← k + 1 to m do
        if A_kk ≠ 0 then
            f ← A_ik / A_kk;
            for j ← k to n do
                A_ij ← A_ij − f · A_kj;
            end
            b_i ← b_i − f · b_k;
        end
    end
end
return A, b;
```

1.6 Back Row Reduction

Note:-

Back row reduction, also known as back substitution, is a method used to solve a system of linear equations that has been transformed into an upper triangular form through Gaussian elimination. This method involves solving the equations starting from the last row and moving upwards.

Definition 1.6.1: Back Row Reduction

Consider a system of linear equations represented in matrix form as $A\mathbf{x} = \mathbf{b}$, where A is an upper triangular matrix:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

The solution is obtained by solving the last equation first and then substituting the obtained values into the preceding equations.

Example 1.6.1 (Example of Back Row Reduction)

Consider the following upper triangular system:

$$\begin{cases} 2x_1 + 3x_2 + 4x_3 = 5, \\ 0x_1 + 6x_2 + 7x_3 = 8, \\ 0x_1 + 0x_2 + 9x_3 = 10. \end{cases}$$

We start with the last equation:

$$9x_3 = 10 \quad \Rightarrow \quad x_3 = \frac{10}{9}.$$

Next, we substitute x_3 into the second equation:

$$6x_2 + 7\left(\frac{10}{9}\right) = 8 \quad \Rightarrow \quad 6x_2 + \frac{70}{9} = 8 \quad \Rightarrow \quad 6x_2 = 8 - \frac{70}{9} \quad \Rightarrow \quad x_2 = \frac{2}{9}.$$

Finally, we substitute x_2 and x_3 into the first equation:

$$2x_1 + 3\left(\frac{2}{9}\right) + 4\left(\frac{10}{9}\right) = 5 \quad \Rightarrow \quad 2x_1 + \frac{6}{9} + \frac{40}{9} = 5 \quad \Rightarrow \quad 2x_1 = 5 - \frac{46}{9} \quad \Rightarrow \quad x_1 = \frac{1}{9}.$$

Thus, the solution is:

$$x_1 = \frac{1}{9}, \quad x_2 = \frac{2}{9}, \quad x_3 = \frac{10}{9}.$$

Algorithm 3: Back Row Reduction (Back Substitution)

Input: Upper triangular matrix A of size $n \times n$ and vector $\mathbf{b}$ of size n
Output: Solution vector $\mathbf{x}$ of size n

/* Initialize solution vector */

for $i \leftarrow n$ **to** 1 **do**
  $x_i \leftarrow b_i$;
  for $j \leftarrow i+1$ **to** n **do**
    $x_i \leftarrow x_i - A_{ij} \cdot x_j$;
  end
  $x_i \leftarrow x_i / A_{ii}$;
end
return $\mathbf{x}$;

Example 1.6.2 (Full SLE solving)

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

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

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

given m=0

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

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

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

Chapter 2

Linear Combinations

2.1 Matrix–Vector Product as Linear Combination

$$Ax = \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} = b.$$

Intuition: "Scaling each column of A by its corresponding entry in x."

In other words, to compute Ax, you take x_1 times the first column of A plus x_2 times the second column of A.

Example 2.1.1 (A Specific Choice of x)

If

$$A = \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix},$$

then the first column is

$$a_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \text{and} \quad a_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}.$$

A linear combination of these columns is

$$x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} = b.$$

As an example, if $x_1 = -\frac{1}{3}$ and $x_2 = 1$,

$$-\frac{1}{3} \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} - 2 \\ -1 + 2 \end{bmatrix} = \begin{bmatrix} -\frac{7}{3} \\ 1 \end{bmatrix}.$$

In the notes, a similar combination yields $\begin{bmatrix} -2 \\ -2 \end{bmatrix}$. The key point is that any pair (x_1, x_2) gives a vector in $\mathbb{R}^2$.

Definition 2.1.1: Span

Span of a set of vectors is the collection of all linear combinations of those vectors. Concretely, for

$$a_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad a_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix},$$

we write

$$\text{Span}\{\, a_1, a_2\} = \Big\{ x_1 a_1 + x_2 a_2 \,\Big|\, x_1, x_2 \in \mathbb{R} \Big\}.$$

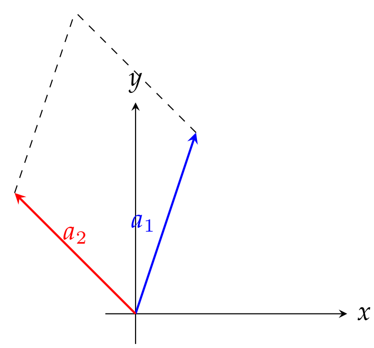
Any vector b not in this span means $Ax = b$ has no solution.

2.2 Linear Combinations and the Span

Vectors "outside" the span of $\{a_1, a_2\}$ are precisely those b for which the system $Ax = b$ is not consistent. Equivalently, they are not expressible as a linear combination of a_1 and a_2.

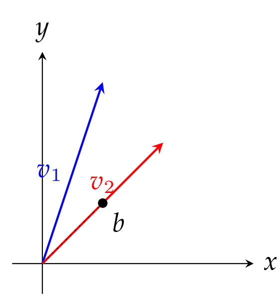
2.3 Geometric Plots in the Notes

1. *First Plot (showing the columns a_1 and a_2):*



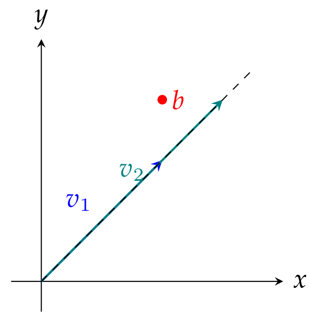
   Any linear combination $x_1 a_1 + x_2 a_2$ lands in the parallelogram structure (and its extensions) formed by these two vectors.

2. *Second Plot (showing a vector not in the span)*:



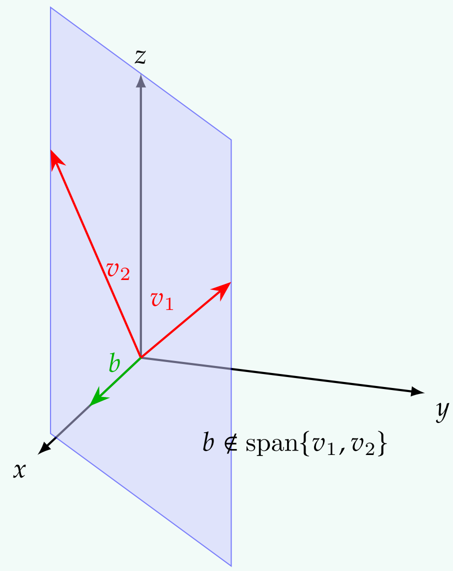
   If v_1 and v_2 do not cover $(1, 1)$ by any linear combination, then $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is not in their span, hence no solution exists for $Ax = b$.

3. *Plot illustrating parallel vectors*:



Here, v_1 and v_2 are multiples of each other, so their span is just the single dashed line. The vector b off that line cannot be written as any combination of v_1 and v_2.

Example 2.3.1 ($\vec{b}$ not in span of $\vec{v}_1, \vec{v}_2$)



2.3.1 Key Conclusions

- Span$\{a_1, a_2\}$ is the set of all vectors b for which $Ax = b$ has a solution.
- If b is not in that span, there is no solution.
- Geometrically, if a_1 and a_2 are not multiples of each other, their span is a 2D plane through the origin in $\mathbb{R}^2$. If they are multiples, the span is just a single line, and most vectors in $\mathbb{R}^2$ lie outside that line (no solution).
- $span a_1$ is a line or point. $span a_1, a_2$ is a plane, line or point. and so on

2.3.2 Dropping a vector to retain the same span

Question 2: Drop vector a_3 to retain the same span as a_1, a_2 and a_3.

$$a_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$a_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$a_3 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

Solution: yes, $span a_1, a_2, a_3 = span a_1, a_2$ because a_3 is a linear combination of a_1 and a_2 $(a_1 + a_2 a_3)$

2.4 Evaluating Linear Dependence Procedurally

Example 2.4.1 (Finding linear dependence of a set of vectors.)

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

Since v_3 is a linear combination of v_1 and v_2, we can drop v_3 to retain the same span as v_1 and v_2.

$$span v_1, v_2, v_3 = span v_1, v_2$$

As

$$v_3 = 2v_1 - v_2$$

Note:-

Working towards a procedure for evaluating linear dependence of a set of vectors.

$$-v_1 + 2v_2 - v_3 = 0$$

$$\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

We know that one solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}$$

Since this SLE has one particular solution as well as the trivial one, then it must have infinitely many solutions.
Then if $A\vec{x} = 0$ has infinitely many solutions, then the set of vectors (columns) in A is linearly dependent.
Conversely, if $A\vec{x} = 0$ has only one solution ($\vec{x} = \vec{0}$), then the set of vectors (columns) in A is linearly independent.

Definition 2.4.1: Linear Independence

We say a set of vectors $\{v_1, v_2, \cdots, v_n\}$ is *linearly independent* if the only solution to the SLE

$$\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

is the trivial solution, i.e. $\vec{x} = \vec{0}$.

$$A' = \begin{bmatrix} 1 & 3 & 5 & 0 \\ 2 & 2 & 2 & 0 \\ 3 & 1 & -1 & 0 \end{bmatrix} \xrightarrow{R_2=R_2-2R_1, \quad R_3=R_3-3R_1} \begin{bmatrix} 1 & 3 & 5 & 0 \\ 0 & -4 & -8 & 0 \\ 0 & -8 & -16 & 0 \end{bmatrix} \xrightarrow{R_3=R_3-2R_2} \begin{bmatrix} 1 & 3 & 5 & 0 \\ 0 & -4 & -8 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \text{REF}(A')$$

1 Free variable $\implies$ Linearly Independent

$$\text{REF}(A') \xrightarrow{R_2=R_2/(-4)} \begin{bmatrix} 1 & 3 & 5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_1=R_1-3R_2} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \text{RREF}(A') \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Then

$$\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} t_3 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

And therefore the linear combination in its homogeneous form implies

$$t_3(\vec{v}_1 - 2\vec{v}_2 + \vec{v}_3) = \vec{0}$$

2.5 Identity Matrix

Definition 2.5.1: Identity Matrix

special Square matrix where the diagonal elements are 1 and the off-diagonal elements are 0.

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Note that $I_n\vec{x} = \vec{x}$ and $I_n' = I_n$ and therefore is called a **symmetric matrix**

Question 3: Matrix Multiplication that gives Identity matrix

$$A = \begin{bmatrix} 1 & 0.7 \\ 0 & 1 \end{bmatrix}$$

Find a matrix B such that

$$AB = I$$

Solution: Since A makes a horizontal shear of 0.7 then B should do the opposite

$$\therefore B = \begin{bmatrix} 1 & -0.7 \\ 0 & 1 \end{bmatrix} = A^{-1}$$

Question 4: Matrix Multiplication that gives Identity matrix 2

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

Find a matrix B such that

$$AB = I$$

Solution: Since A makes a reflection about the x_2 axis, B must do the same

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

Question 5: Matrix Multiplication that gives Identity matrix

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

Find a matrix B such that

$$AB = I$$

Solution: A flattens all vertical components. Geometrically speaking, A^{-1} doesn't exist, as shown here

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$$

Observe that $spanA$ is the line $x_2 = 0$

Definition 2.5.2: General Case for inverses of square matrices

For a matrix A which is $n \times n$, A^{-1} exists if and only if

$$span\{A\} = \mathbb{R}^n$$

Definition 2.5.3: Finding A^{-1}

Get the Augmented matrix

$$A'' = [\, A \,|\, I \,]$$

where A , I are in the form $n \times n$ and then perform RREF on this augmented matrix. Once done, you will have a matrix in the form

$$RREF(A'') = [\, I \,|\, A^{-1} \,]$$

Example 2.5.1 (Finding A^{-1})

$$A = \begin{bmatrix} 1 & 0.7 \\ 0 & 1 \end{bmatrix}$$

$$A'' = \begin{bmatrix} 1 & 0.7 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Then RREF

$$\xrightarrow{R_1 = R_1 - 0.7R_2} \begin{bmatrix} 1 & 0 & 1 & -0.7 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$\therefore A^{-1} = \begin{bmatrix} 1 & -0.7 \\ 0 & 1 \end{bmatrix}$$

Example 2.5.2 (Finding A^{-1})

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

Check if $span\{A\} = \mathbb{R}^3$

$$\xrightarrow{\substack{R_2 = R_2 - 2R_1 \\ R_3 = R_3 - 3R_1}} \begin{bmatrix} 1 & 3 & 2 \\ 0 & -4 & -3 \\ 0 & -8 & -3 \end{bmatrix} \xrightarrow{R_3 = R_3 - 2R_2} \begin{bmatrix} 1 & 3 & 2 \\ 0 & -4 & -3 \\ 0 & 0 & 3 \end{bmatrix}$$

Here A has 3 pivots, so it is invertible. Next, make the augmented matrix $A'' = [\,A\,|\,I\,]$ and perform RREF

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

$$\begin{bmatrix} 1 & 3 & 2 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 & 1 & 0 \\ 3 & 1 & 3 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\substack{R_2 = R_2 - 2R_1 \\ R_3 = R_3 - 3R_1}} \begin{bmatrix} 1 & 3 & 2 & 1 & 0 & 0 \\ 0 & -4 & -3 & -2 & 1 & 0 \\ 0 & -8 & -3 & -3 & 0 & 1 \end{bmatrix} \xrightarrow{R_3 = R_3 - 2R_2} \begin{bmatrix} 1 & 3 & 2 & 1 & 0 & 0 \\ 0 & -4 & -3 & -2 & 1 & 0 \\ 0 & 0 & 3 & 1 & -2 & 1 \end{bmatrix}$$

$$\xrightarrow{R_2 = -\frac{1}{4}R_2} \begin{bmatrix} 1 & 3 & 2 & 1 & 0 & 0 \\ 0 & 1 & \frac{3}{4} & \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & 0 & 3 & 1 & -2 & 1 \end{bmatrix} \xrightarrow{R_1 = R_1 - 3R_2} \begin{bmatrix} 1 & 0 & -\frac{1}{4} & -\frac{1}{2} & \frac{3}{4} & 0 \\ 0 & 1 & \frac{3}{4} & \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & 0 & 3 & 1 & -2 & 1 \end{bmatrix}$$

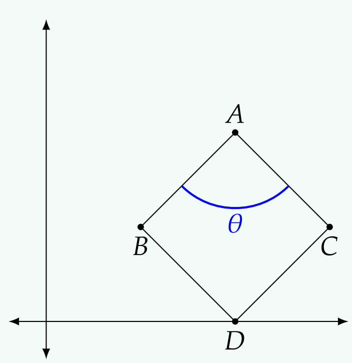
$$\xrightarrow{R_3 = \frac{1}{3}R_3} \begin{bmatrix} 1 & 0 & -\frac{1}{4} & -\frac{1}{2} & \frac{3}{4} & 0 \\ 0 & 1 & \frac{3}{4} & \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \xrightarrow{\substack{R_1 = R_1 + \frac{1}{4}R_3 \\ R_2 = R_2 - \frac{3}{4}R_3}} \begin{bmatrix} 1 & 0 & 0 & -\frac{5}{12} & \frac{7}{12} & \frac{1}{12} \\ 0 & 1 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

Therefore

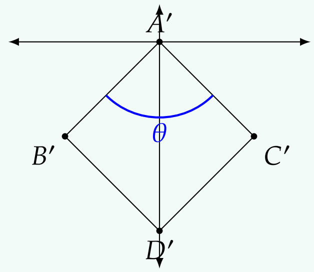
$$A^{-1} = \begin{bmatrix} -\frac{5}{12} & \frac{7}{12} & \frac{1}{12} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

Example 2.5.3 (Review)

$$A(5,5), \quad B\left(\frac{5}{2}, \frac{5}{2}\right), \quad C\left(\frac{15}{2}, \frac{5}{2}\right), \quad D(5,0)$$



$$A'(0,0), \quad B'\left(-\frac{5}{2}, -\frac{5}{2}\right), \quad C'\left(\frac{5}{2}, -\frac{5}{2}\right), \quad D'(0,-5)$$



$$\vec{AB} = \vec{v_2} = \begin{bmatrix} -\frac{5}{2} \\ -\frac{5}{2} \end{bmatrix}, \vec{AC} = \vec{v_1} = \begin{bmatrix} \frac{5}{2} \\ -\frac{5}{2} \end{bmatrix}$$

$$\vec{v_1}^T \vec{v_2} = -\frac{25}{4} + \frac{25}{4} = 0 = \left|\vec{v_1}\right| \left|\vec{v_2}\right| \cos(\theta)$$

$$\therefore \theta = 90^\circ$$

2.5.1 ERP's and Identity Matrices

Let A'' be a matrix. Consider the augmented matrix:

$$[A \mid I] : m \times 2m$$

Perform a sequence of elementary row operations (ERPs) on A to turn it into the identity matrix I:

$$E_p \ldots E_2 E_1 A'' = B\,[A \mid I] = [BA \mid BI]$$

where B is an $m \times m$ matrix.
If we achieve $BA = I$, then:

$$B = A^{-1}$$

Question 6: Thinklet

Consider a set $\mathcal{S}$ of vectors in $\mathbb{R}^m$ who's tips form a straight line through the origin.
If we linearly combine a set of vectors in $\mathcal{S}$, is the result guaranteed to be in $\mathcal{S}$?

Solution: DUH

Chapter 3

Vector Spaces

Definition 3.0.1: Vector Space

A vector space is a set of vectors in $\mathbb{R}^m$ for which the following hold:

1. The set is closed under addition: that is, if $\vec{v}_1, \vec{v}_2$ are both in the set, then $\vec{v}_1 + \vec{v}_2$ is also in the set.
2. The set is closed under scalar multiplication: that is, if $\vec{v}$ is in the set and $c \in \mathbb{R}$, then $c\vec{v}$ is also in the set.

Definition 3.0.2: Basis of a Vector Space

A basis for a vector space in $\mathbb{R}^m$ is a set of independent vectors that span the vector space. The maximum number of vectors that could be in that basis is m.
m is the **dimension** of the vector space.

Example 3.0.1 (Examples of Vector Spaces)

$\mathcal{S}$	Basis	Dimension
$\{\vec{0}\}$	$\varnothing$	0
$\left\{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} t\right\}$	$\left\{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}\right\}$	1
$\left\{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} t_1 + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} t_2\right\}$	$\left\{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}\right\}$	2
$\left\{\sum_{i=1}^{k} \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix} t_i\right\}$	$\left\{\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \ldots, \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix}\right\}$	k

3.1 Column Space

Definition 3.1.1: Column space of matrix A

The *column space* of an $m \times n$ matrix A, denoted as $\text{Col}(A)$, is the subspace of $\mathbb{R}^m$ spanned by the columns of A. Formally,

$$\text{Col}(A) = \text{span}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)$$

where $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are the column vectors of A. This means that $\text{Col}(A)$ consists of all possible linear combinations of the columns of A.

3.2 Linear Combinations and Column Space

3.2.1 Basic Definitions and Concepts

Definition 3.2.1: Matrix Notation

An $m \times n$ matrix A is a rectangular array of numbers with m rows and n columns. We can represent A as a collection of column vectors:

$$A = [\vec{v}_1 \; \vec{v}_2 \; \ldots \; \vec{v}_n]$$

where each $\vec{v}_i$ is a vector in $\mathbb{R}^m$.

Definition 3.2.2: Column Space (Col A)

The column space of a matrix A, denoted as Col A, is the set of all linear combinations of the columns of A. Formally, if $A = [\vec{v}_1 \; \vec{v}_2 \; \ldots \; \vec{v}_n]$, then

$$\text{Col } A = \text{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} = \{t_1\vec{v}_1 + t_2\vec{v}_2 + \cdots + t_n\vec{v}_n \mid t_1, t_2, \ldots, t_n \in \mathbb{R}\}$$

3.2.2 Dimension and Basis of the Column Space

Question 7: What is the dimension of Col A?

The dimension of Col A is equal to the number of pivot columns in the reduced row echelon form (REF) of A.

Solution: The dimension corresponds to the number of linearly independent columns in matrix A.

Question 8: What is the basis for Col A?

The basis for Col A is formed by the column vectors of the **original** matrix A that correspond to the pivot columns in the REF of A. These are referred to as the "basic columns."

Solution:

1. Find REF(A).
2. Identify the pivot columns in REF(A).
3. Select the corresponding columns from the **original* matrix A. These columns form the basis.**

3.2.3 Parametric Form of the Column Space

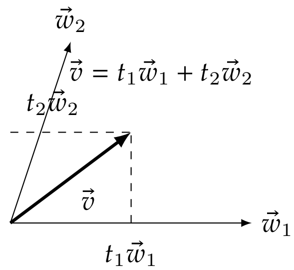
Question 9: What is the parametric form of Col A?

If the basis for Col A is $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$, then the parametric form of any vector $\vec{v}$ in Col A is given by:

$$\vec{v} = t_1\vec{w}_1 + t_2\vec{w}_2 + \cdots + t_k\vec{w}_k$$

where $t_1, t_2, \ldots, t_k$ are scalar parameters (real numbers).

Solution: This expresses any vector in the column space as a linear combination of the basis vectors.



3.2.4 Geometry of the Column Space

Question 10: What is the geometry of Col A?

The geometric interpretation of Col A depends on its dimension:

- If $\dim(\text{Col } A) = 0$, Col A is a point (the origin, $\vec{0}$).
- If $\dim(\text{Col } A) = 1$, Col A is a line through the origin.
- If $\dim(\text{Col } A) = 2$, Col A is a plane through the origin.
- If $\dim(\text{Col } A) = 3$, Col A is a volume (all of $\mathbb{R}^3$) through the origin.
- In general, if A is an $m \times n$ matrix, and $\dim(\text{Col } A) = k$, Col A is a k-dimensional subspace of $\mathbb{R}^m$.

Solution: The dimension indicates the "degrees of freedom" or the number of independent directions spanned by the column vectors.

3.2.5 Example

Example 3.2.1 (Example Calculation)
Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 4 & 6 \end{bmatrix}$$

Find the REF of A, the dimension of Col A, a basis for Col A, the parametric form of Col A, and the geometry of Col A.

Solution: First, find the reduced row echelon form (REF) of A:

$$\text{REF}(A) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & -4 & -8 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

- **Dimension of Col A:** The REF has two pivot columns (the first and second columns). Therefore, $\dim(\text{Col } A) = 2$.

- **Basis for Col A:** The pivot columns in REF(A) are the first and second columns. We take the corresponding columns from the *original* matrix A to form the basis:

$$\text{Basis(Col } A) = \left\{ \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \right\}$$

- **Parametric Form of Col A:** Any vector $\vec{v}$ in Col A can be written as a linear combination of the basis vectors:

$$\vec{v} = t_1 \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + t_2 \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

where t_1 and t_2 are any real numbers.

- **Geometry of Col A:** Since dim(Col A) = 2, Col A is a plane in $\mathbb{R}^3$ that passes through the origin.

3.3 Nullspace of a Matrix

Definition 3.3.1: Nullspace

The nullspace of a matrix A, denoted by Nul A, is the set of all solutions to the homogeneous equation $A\vec{x} = \vec{0}$.

Note:-

The solution space for $A\vec{x} = \vec{0}$ is a vector space. The nullspace of A resides in $\mathbb{R}^n$, where n is the number of columns of A.

Question 11: Dimension of Nullspace

What is the dimension of Nul A?

Solution: The dimension of Nul A is equal to the number of free columns in the reduced row echelon form (RREF) of A. We write this as Dim(Nul A) = # of free columns in RREF(A).

Question 12: Basis for Nullspace

What is the basis for Nul A?

Solution: To find the basis for Nul A, we first find the RREF of A. If the solution space of $A\vec{x} = \vec{0}$ is given by

$$\vec{x} = t_1\vec{w}_1 + t_2\vec{w}_2 + \cdots + t_k\vec{w}_k$$

where k is the number of free columns in RREF(A), and $t_1, t_2, \ldots, t_k$ are arbitrary scalars, then the basis of Nul A is the set $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$.

Question 13: Parametric Form of Nullspace

What is the parametric form of Nul A?

Solution: The parametric form of Nul A is given by the linear combination of the basis vectors:

$$\vec{v} = t_1\vec{w}_1 + t_2\vec{w}_2 + \cdots + t_k\vec{w}_k$$

where $\vec{w}_1, \vec{w}_2, ..., \vec{w}_k$ are the basis vector for the nullspace.

Example 3.3.1 (Example)

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 4 & 6 \end{bmatrix}$$

Find the nullspace, its dimension, basis, and the geometric interpretation.
First, find the RREF of A:

$$\text{REF}(A) = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & 0 & 0 \end{bmatrix}$$

Further reducing to RREF:

$$\text{RREF}(A) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

- The first column corresponds to a basic variable.
- The second column corresponds to a basic variable.
- The third column corresponds to a free variable.

The dimension of the nullspace is the number of free variables:

$$\text{Dim}(\text{N}(A)) = 1$$

To find the solution space for $A\vec{x} = \vec{0}$, we use the RREF:
Let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. The RREF gives us the following equations:

$$\begin{aligned} x_1 - x_3 &= 0 \\ x_2 + 2x_3 &= 0 \end{aligned}$$

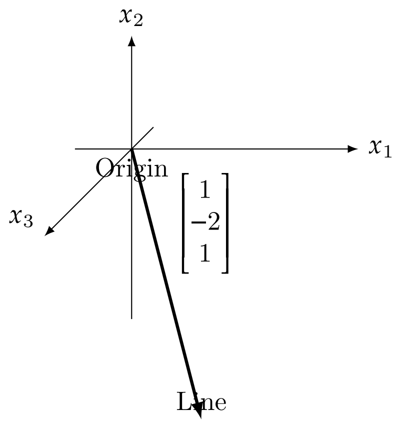
x_3 is free. Let $x_3 = t_3$. Then $x_1 = t_3$ and $x_2 = -2t_3$. So,

$$\vec{x} = \begin{bmatrix} t_3 \\ -2t_3 \\ t_3 \end{bmatrix} = t_3 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Therefore, the basis for Nul A is:

$$\text{Basis}(\text{Nul } A) = \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$$

The geometry of Nul A is a line in $\mathbb{R}^3$ passing through the origin and in the direction of the vector $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$.



3.4 Row Space of a Matrix

Definition 3.4.1: Row Space

The row space of a matrix A, denoted as Row A, is defined as the column space of the transpose of A, i.e., Row A = Col A^T.

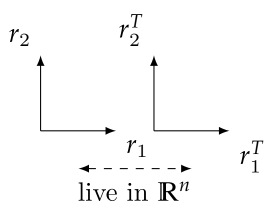
Let's consider a matrix A and its transpose A^T:

$$A = \begin{bmatrix} - & r_1 & - \\ - & r_2 & - \\ & \vdots & \\ - & r_m & - \end{bmatrix} \qquad A^T = \begin{bmatrix} | & | & & | \\ r_1^T & r_2^T & \cdots & r_m^T \\ | & | & & | \end{bmatrix}$$

Where $r_1, r_2, \ldots, r_m$ represent the rows of matrix A. The rows of A become the columns of A^T.

Note:-

The row vectors $r_1, r_2, \ldots, r_m$ live in $\mathbb{R}^n$.



3.5 Dimension and Basis of Row Space

Question 14: What is the dimension of Row A?

What is the dimension of the row space of a matrix A?

Solution: The dimension of Row A is equal to the number of pivots in the row echelon form (REF) of A. It is also equal to the number of pivots in the REF of A^T.

$$\text{Dim(Row } A) = \#\text{ pivots in REF}(A) = \#\text{ pivots in REF}(A^T)$$

Question 15: What is the Basis of Row A

How to find the basis for the Row Space of A

Solution:

1. Identify all the rows of REF(A) that have pivots.
2. The basis vectors for Row A are:
   - Vectors corresponding to rows (with pivots in REF(A)) of A.

   OR

   - Vectors corresponding to rows (with pivots in REF(A)) of REF(A).

   OR

   - Vectors corresponding to rows (with pivots in REF(A)) of RREF(A).

Example 3.5.1 (Example: Finding the Basis and Dimension)

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 2 & 3 \end{bmatrix}$$

Find the dimension and a basis for Row A.

Solution: First, find the REF of A:

$$\text{REF}(A) = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The REF of A has one pivot (in the first row).

Dimension of Row A:

$$\text{Dim}(\text{Row } A) = 1$$

Basis for Row A:

- Using rows of A with pivots in REF(A): [1 2 3]
- Using rows of REF(A) with pivots: [1 2 3]

Algorithm 4: Finding Dimension and Basis of Row Space

Input: Matrix A
Output: Dimension of Row A, Basis of Row A

```
/* Compute the Row Echelon Form (REF) of A */
REF_A ← REF(A);
/* Count the number of pivots in REF(A) */
num_pivots ← Number of pivots in REF_A;
/* The dimension of Row A is the number of pivots */
dim_row_A ← num_pivots;
/* Initialize an empty set to store basis vectors */
basis_row_A ← ∅;
/* Identify rows with pivots in REF(A) */
foreach row r in REF_A do
    if row r has a pivot then
        /* Add corresponding row from original matrix A to basis */
        Let r′ be the corresponding row in matrix A. basis_row_A ← basis_row_A ∪ {r′};
    end
end
return dim_row_A, basis_row_A;
```

Chapter 4

Properties of Matrices

4.1 Determinant of a Square Matrix

Consider a square Matrix

$$A : m \times m$$

If

$$A^{-1} exists \Leftrightarrow \dim(Col(A)) = \dim(Row(A)) = m$$

$$\therefore Rank(A) = m \text{ and } \mathrm{Nul}(A) = \{0\}$$

And most importantly, since the columns of A are linearly independent,

$$det(ref(A)) \neq 0$$

To make REF(A) unique for A, we require that in constructing REF(A), we follow three rules:

1. Use the "REPLACE" ERP freely.
2. When using the "SWAP" ERP, apply the "SCALE" ERP to one of the rows by -1.
3. Don't use the "SCALE" ERP for any other purpose than the one in Rule 2.

4.1.1 Cofactor Algorithm for Computing Determinant

Given a square matrix $A = [a_{ij}]$ of size $n \times n$, the determinant can be computed by expanding along any row i (where $1 \leqslant i \leqslant n$) as follows:

1. **Base Case**: If $n = 1$, then $\det(A) = a_{11}$.
2. **Recursive Case**: For $n > 1$:

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} \, a_{ij} \, \det(M_{ij}),$$

   where:

   - M_{ij} is the **minor matrix** obtained by deleting row i and column j from A,
   - $(-1)^{i+j} \det(M_{ij})$ is the **cofactor** of the element a_{ij}.

Steps:

- Choose a row i (commonly the first row, $i = 1$, but any row works).
- For each element a_{ij} in row i:
  - Compute the minor M_{ij}.

- Recursively calculate $\det(M_{ij})$.
- Multiply a_{ij}, $(-1)^{i+j}$, and $\det(M_{ij})$.

- Sum the results for all columns j in row i.

Say B which has the Form

$$\begin{bmatrix} a_1 & a_2 & \dots \\ b_1 & b_2 & \dots \end{bmatrix}$$

And delimits an area in $\mathbb{R}^2$, then the absolute value of the determinant of a matrix A which applies a transformation to B gives the scaling of the area of B

4.1.2 Properties of Determinants

- $det(A \cdot B) = det(A) \cdot det(B)$
- $det(A^{-1}) = \frac{1}{det(A)}$
- $det(A^T) = det(A)$
- $det(I) = 1$
- $det(E_{REPLACE} \cdot A) = det(A)$
- $det(E_{SCALE=\alpha} \cdot A) = \alpha \cdot det(A)$
- $det(A) = 0 \Leftrightarrow det(REF(A)) = 0$

4.2 EigenVectors and EigenValues

Given $A \in \mathbb{R}^{n \times n}$

$$A\vec{x} = \vec{0} \leftarrow \vec{x} \in Null(A)$$

Can rewrite As

$$A\vec{x} = 0\vec{x}$$

Generalise

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

where the solution for $\vec{x}$ also defines a vector space, called an **EigenSpace**, where any vector in this space is called and **EigenVector**, and has a corresponding value of λ called an **EigenValue**.

Example 4.2.1 (Eigenvectors)

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

is $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ an eigenvector for A?

$$\begin{bmatrix} -2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -4 \\ 6 \end{bmatrix} = \lambda \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

No such Lambda, therefore not eigenvector, but

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

So $\lambda = -2$ is the EigenValue for the EigenVector that is $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$

4.3 Eigenvectors and Eigenvalues

Definition 4.3.1: Eigenvector and Eigenvalue

Let A be an $n \times n$ matrix. A non-zero vector $\vec{x}$ is an eigenvector of A if there exists a scalar λ (called an eigenvalue) such that:

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

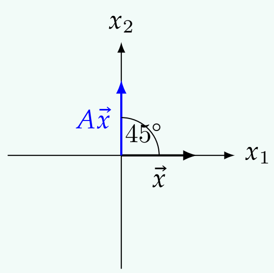
Note:-

The geometric interpretation of $A\vec{x} = \lambda\vec{x}$ is that applying the transformation represented by A to the eigenvector $\vec{x}$ results in a vector that is simply a scaled version of $\vec{x}$. The direction is either preserved (if $\lambda > 0$), reversed (if $\lambda < 0$), or the vector becomes the zero vector (if $\lambda = 0$). Crucially, $\vec{x}$ *cannot* be the zero vector, but λ can be.

Example 4.3.1 (Rotation Matrix)

Consider a rotation matrix A in 2D that rotates vectors by 45 degrees counterclockwise:

$$A = \begin{bmatrix} \cos(45°) & -\sin(45°) \\ \sin(45°) & \cos(45°) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$



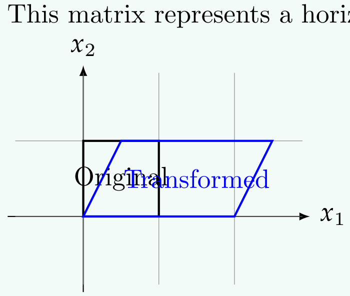
This matrix A has *no* real eigenvectors (and thus no real eigenvalues). Applying A to any non-zero vector $\vec{x}$ will *always* rotate it; there's no non-zero vector that will simply be scaled by A. The resulting vector $A\vec{x}$ will never have the same (or opposite) direction as the original $\vec{x}$.

Example 4.3.2 (Horizontal Shear and Stretch)

Consider the matrix:

$$A = \begin{bmatrix} 2 & 0.5 \\ 0 & 1 \end{bmatrix}$$

This matrix represents a horizontal shear and stretch. Let's visualize its effect on a unit square:



- Vectors along the x_1-axis (of the form $\alpha \begin{bmatrix} 1 \\ 0 \end{bmatrix}$) are stretched by a factor of 2, but their direction remains unchanged.
- Vectors along the x_2-axis (of the form $\alpha \begin{bmatrix} 0 \\ 1 \end{bmatrix}$) are sheared horizontally by 0.5 units, but their vertical

component is unchanged.

- Consider a general vector $\begin{bmatrix}\alpha\\\beta\end{bmatrix}$. Applying the transformation yields $A\begin{bmatrix}\alpha\\\beta\end{bmatrix} = \begin{bmatrix}2\alpha+0.5\beta\\\beta\end{bmatrix}$.

Let's find the eigenvectors and eigenvalues for this matrix.

- For vectors along the x_1 axis:

$$A\begin{bmatrix}\alpha\\0\end{bmatrix} = \begin{bmatrix}2 & 0.5\\0 & 1\end{bmatrix}\begin{bmatrix}\alpha\\0\end{bmatrix} = \begin{bmatrix}2\alpha\\0\end{bmatrix} = 2\begin{bmatrix}\alpha\\0\end{bmatrix}$$

  So, $\begin{bmatrix}\alpha\\0\end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 2$.

- For vectors along the x_2 axis, we might *think* they are eigenvectors, but let's check:

$$A\begin{bmatrix}0\\\alpha\end{bmatrix} = \begin{bmatrix}2 & 0.5\\0 & 1\end{bmatrix}\begin{bmatrix}0\\\alpha\end{bmatrix} = \begin{bmatrix}0.5\alpha\\\alpha\end{bmatrix} = \alpha\begin{bmatrix}0.5\\1\end{bmatrix}$$

  This is *not* a scalar multiple of the original vector $\begin{bmatrix}0\\\alpha\end{bmatrix}$. So, vectors along the x_2 axis are *not* eigenvectors.

- Now consider the case where the transformed vector is α times the original:

$$\begin{bmatrix}2 & 0.5\\0 & 1\end{bmatrix}\begin{bmatrix}-0.5\alpha\\\alpha\end{bmatrix} = \begin{bmatrix}-1\alpha+0.5\alpha\\\alpha\end{bmatrix} = \begin{bmatrix}-0.5\alpha\\\alpha\end{bmatrix} = 1 * \begin{bmatrix}-0.5\alpha\\\alpha\end{bmatrix}$$

  So $\begin{bmatrix}-0.5\alpha\\\alpha\end{bmatrix}$ is an eigenvector corresponding to eigenvalue 1.

4.3.1 Finding Eigenvalues and Eigenvectors

Theorem 4.3.1 Finding Eigenvalues

To find the eigenvalues of a matrix A, we solve the characteristic equation:

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

where I is the identity matrix of the same size as A.

Proof. We start with the definition $A\vec{x} = \lambda\vec{x}$. Rearranging, we get:

$$A\vec{x} - \lambda\vec{x} = \vec{0}$$
$$A\vec{x} - \lambda I\vec{x} = \vec{0}$$
$$(A - \lambda I)\vec{x} = \vec{0}$$

Let $B = A - \lambda I$. We have $B\vec{x} = \vec{0}$. We are looking for non-zero vectors $\vec{x}$ that satisfy this equation.

- If B is invertible (i.e., $\det(B) \neq 0$), then the only solution is $\vec{x} = \vec{0}$. But eigenvectors are *non-zero* by definition. Therefore, we must have $\det(B) = 0$.
- If $\det(B) = 0$, then B is *not* invertible, and the system $B\vec{x} = \vec{0}$ has infinitely many solutions (i.e., a non-trivial null space). These non-zero solutions are the eigenvectors.

Therefore, to find the eigenvalues λ, we must solve $\det(A - \lambda I) = 0$. ☺

Example 4.3.3 (Finding Eigenvalues (Continued))

Let's go back to our example:

$$A = \begin{bmatrix} 2 & 0.5 \\ 0 & 1 \end{bmatrix}$$

We form $B = A - \lambda I$:

$$B = \begin{bmatrix} 2 & 0.5 \\ 0 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2-\lambda & 0.5 \\ 0 & 1-\lambda \end{bmatrix}$$

Now we find the determinant of B:

$$\det(B) = (2-\lambda)(1-\lambda) - (0.5)(0) = (2-\lambda)(1-\lambda)$$

Setting $\det(B) = 0$, we get:

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

This gives us the eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 1$.

4.3.2 Finding Eigenvectors

Algorithm 5: Finding Eigenvalues and Eigenvectors

Input: A matrix A

Output: Eigenvalues and corresponding eigenvectors of A

```
Calculate the characteristic polynomial: det(A − λI);
Solve the characteristic equation det(A − λI) = 0 to find the eigenvalues λ_i;
for each eigenvalue λ_i do
    Form the matrix B_i = A − λ_i I;
    Solve the homogeneous system B_i x⃗ = 0⃗ to find the eigenvectors corresponding to λ_i;
    The solution set will typically involve free variables. Express the eigenvectors in terms of these free
      variables.;
end
return The eigenvalues λ_i and their corresponding eigenvectors.
```

Example 4.3.4 (Finding Eigenvectors (Continued))

We found the eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 1$ for the matrix $A = \begin{bmatrix} 2 & 0.5 \\ 0 & 1 \end{bmatrix}$. Now let's find the eigenvectors.

- For $\lambda_1 = 2$:

  $$B_1 = A - 2I = \begin{bmatrix} 2-2 & 0.5 \\ 0 & 1-2 \end{bmatrix} = \begin{bmatrix} 0 & 0.5 \\ 0 & -1 \end{bmatrix}$$

  We solve $B_1\vec{x} = \vec{0}$:

  $$\begin{bmatrix} 0 & 0.5 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

  Row reducing (although not strictly necessary here, we do it for demonstration):

  $$\begin{bmatrix} 0 & 0.5 \\ 0 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

  This gives us $x_2 = 0$, and x_1 is free. Let $x_1 = t_1$. Then the eigenvector is:

  $$\vec{x} = \begin{bmatrix} t_1 \\ 0 \end{bmatrix} = t_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- For $\lambda_2 = 1$:

$$B_2 = A - I = \begin{bmatrix} 2-1 & 0.5 \\ 0 & 1-1 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0 & 0 \end{bmatrix}$$

We solve $B_2\vec{x} = \vec{0}$:

$$\begin{bmatrix} 1 & 0.5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This system is already in row-echelon form. We have $x_1 + 0.5x_2 = 0$, so $x_1 = -0.5x_2$. x_2 is free. Let $x_2 = t_2$. Then $x_1 = -0.5t_2$, and the eigenvector is:

$$\vec{x} = \begin{bmatrix} -0.5t_2 \\ t_2 \end{bmatrix} = t_2 \begin{bmatrix} -0.5 \\ 1 \end{bmatrix}$$

4.4 Properties of Eigenvectors

Definition 4.4.1: Dimension of eigenspace

When an eigenvalue repeats k times,

$$1 \leqslant dim(\text{eigenspace for that eigenvalue}) \leqslant k$$

Example 4.4.1 (Example 1 of this property)

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

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

$$\det(A - \lambda I) = (I1 - \lambda)^3$$

$$\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 1, k = 3$$

$$\lambda_1 \text{repeats} 3 (= k) \text{times}$$

And the eigenspace for$\lambda_1 = 1$ is

$$A - \lambda_1 I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

therefore the eigenspace is

$$\left\{ t_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + t_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} : t_1, t_2, t_3 \in \mathbb{R} \right\}$$

Example 4.4.2 (Example 2 of this property)

$$A = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}$$

$$\det(A - \lambda I) = (I1 - \lambda)^3$$

$$(1 - \lambda)^2 - (0.5)(0) = 0$$

$$\lambda_1 = 1, \lambda_2 = 1, k = 2$$

Therefore the eigenspace is

$$A - \lambda_1 I = \begin{bmatrix} 0 & 0.5 \\ 0 & 0 \end{bmatrix} \xrightarrow{R_1 = 2R_1} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Therefore the eigenspace for $\lambda = 1$ is

$$\left\{ t_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} : t_1 \in \mathbb{R} \right\}$$

$$\therefore dim(\text{eigenspace for } \lambda = 1) = 1 \leqslant k = 2$$

Definition 4.4.2: Dominant Eigenvector

An eigenvector whose eigenvalue (among all eigenvalues for some $A : m \times m$) has the largest magnitude ($|\lambda|$) gives the direction along which an initial blob of points is stretched the most.

Example 4.4.3 (Dominant Eigenvector Example)

A=

$$\begin{bmatrix} 2 & 4 \\ 3 & 3 \end{bmatrix}$$

$$\det(A - \lambda I) = (2 - \lambda)(3 - \lambda) - 4 * 3 = 0$$

$$\therefore \lambda_1 = 6, \lambda_2 = -1$$

Eigenspace for largest magnitude of eigenvalue $\lambda_1 = 6$

$$A = \begin{bmatrix} -4 & 4 \\ 3 & -3 \end{bmatrix} \xrightarrow{R_2 = R_2 + \frac{3}{4} R_1} \begin{bmatrix} -4 & 4 \\ 0 & 0 \end{bmatrix} \xrightarrow{R_1 = -\frac{1}{4} R_1} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$

$$\therefore \vec{v}_1 = t_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

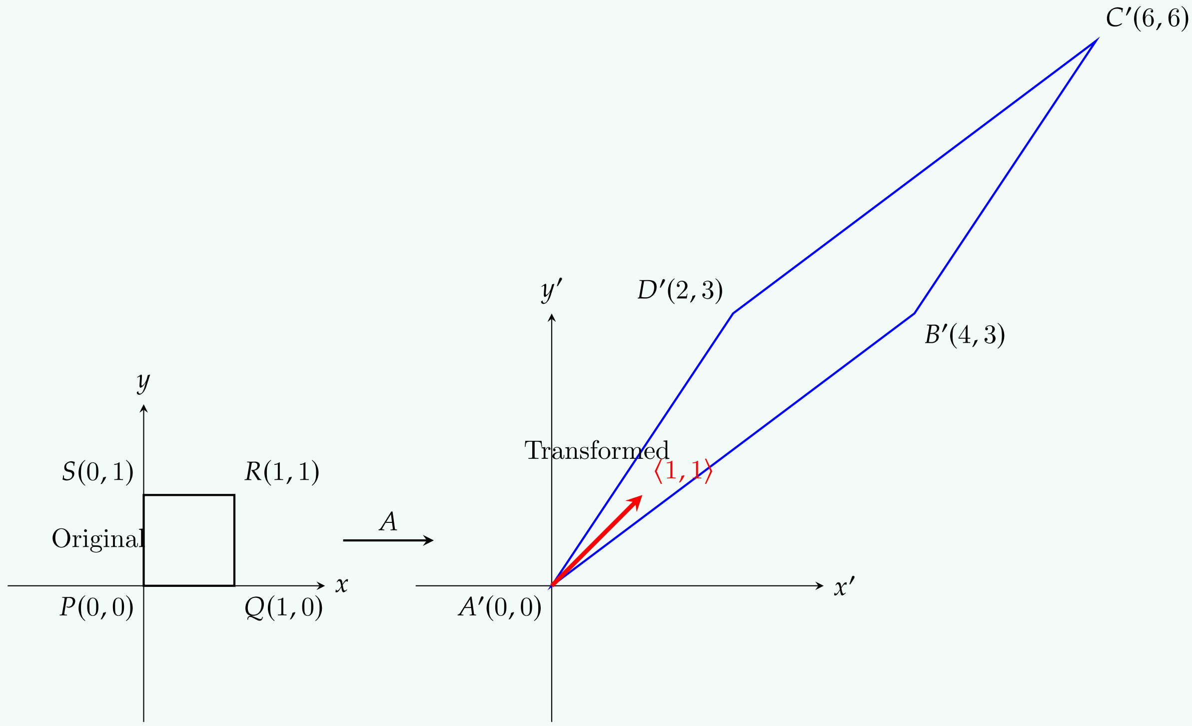
and the eigenspace for $\lambda_1 = 6$ is

$$\left\{ t \begin{bmatrix} 1 \\ 1 \end{bmatrix} : t \in \mathbb{R} \right\}$$

Now apply the transformation to a set of points (unit square)

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

Visually represented this looks like this



Example 4.4.4 (Dominant Eigenvector Example 2)
Consider

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

The eigenvalues come from

$$\det(A - \lambda I) = (3 - \lambda)(2 - \lambda) = 0,$$

so $\lambda_1 = 3$ and $\lambda_2 = 2$. The dominant eigenvalue is 3. To find its eigenvector, solve

$$(A - 3I)\vec{v} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \vec{0}.$$

The first row gives $v_2 = 0$; choosing $v_1 = 1$ yields

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

This indicates that the largest stretch occurs in the horizontal direction.

Transformation on a set of points:
Consider the original set of points

$$\mathcal{P} = \{(x, y) \mid -1 \le x \le 1,\ -0.5 \le y \le 0.5\}.$$

Under the transformation A, each point (x, y) is mapped to

$$(x', y') = (3x + y,\ 2y).$$

The four corners transform as follows:

- $(-1, -0.5)$ maps to $(3(-1) + (-0.5),\ 2(-0.5)) = (-3.5,\ -1)$.
- $(1, -0.5)$ maps to $(3(1) + (-0.5),\ 2(-0.5)) = (2.5,\ -1)$.

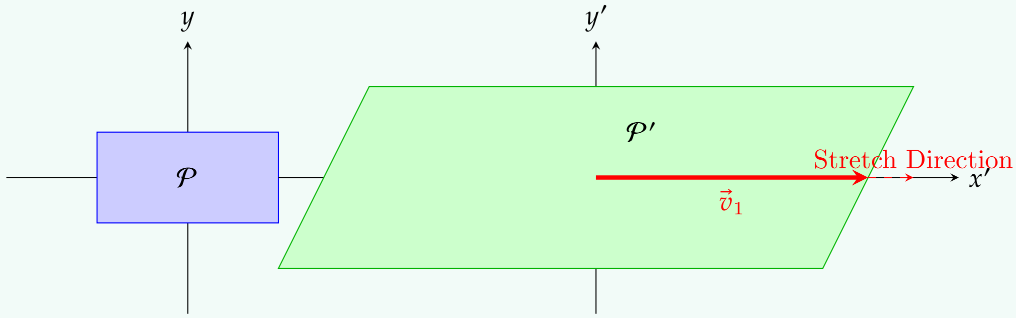
- $(1, 0.5)$ maps to $(3(1) + (0.5),\ 2(0.5)) = (3.5,\ 1)$.
- $(-1, 0.5)$ maps to $(3(-1) + (0.5),\ 2(0.5)) = (-2.5,\ 1)$.

Thus, the image of $\mathcal{P}$ is the quadrilateral

$$\mathcal{P}' = \{(x', y')\}$$

with vertices at $(-3.5, -1)$, $(2.5, -1)$, $(3.5, 1)$, and $(-2.5, 1)$.

Diagram:



Definition 4.4.3: Eigenvalues of Triangular Matrices

The elements on the main diagonal of a triangular matrix are the eigenvalues of that matrix.

Example 4.4.5 (EigenValues of Triangular matricies example)

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

$$\det(A - \lambda I) = (1 - \lambda)(4 - \lambda)(6 - \lambda)$$

$$\lambda_1 = 1, \lambda_2 = 4, \lambda_3 = 6$$

$$\text{Eigenvalues of } A = \{1, 4, 6\}$$

Definition 4.4.4: Number of Distinct EigenValues

The number of distinct eigenvalues of a matrix $A : m \times m$ is $\leqslant m$

Definition 4.4.5: Eigenvalues and Determinant

For a matrix $A : m \times m$

$$\det(A) = \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_m$$

Proof:

$$dat(A - \lambda I) = \pm\lambda^m + a_{m-1}\lambda^{m-1} + \dots + a_1\lambda + a_0$$

where the $\pm$ follows

$$\begin{cases} +1, & m\%2 = 0 \\ -1, & m\%2 = 1 \end{cases}$$

$$\therefore |A - \lambda I| = \begin{cases} \lambda^m - a_{m-1}\lambda^{m-1} + \dots - a_1\lambda + a_0, & m\%2 = 0 \\ -\lambda^m + a_{m-1}\lambda^{m-1} - \dots + a_1\lambda - a_0, & m\%2 = 1 \end{cases}$$

Set $\lambda = 0$

$$|A| = \begin{cases} \lambda_1, \lambda_2, \dots, \lambda_m, & m\%2 = 0 \\ \lambda_1, \lambda_2, \dots, \lambda_m, & m\%2 = 0 \end{cases}$$

Definition 4.4.6: Non-invertible Square matrix

For a non-invertible (singular) square matrix $A : m \times m$, the determinant is 0, therefore at least one eigenvalue is 0.

Definition 4.4.7: Spanning Eigenvectors

The eigenvectors of A span all of $\mathbb{R}^m$ if and only if A has m distinct eigenvalues. If an eigenvalue repeats k times, the dimension of the corresponding space is k.

Theorem 4.4.1

If $A : m \times m$ has m linearly independent eigenvectors $\iff$ Every eigenvalue that repeats $k : k \in \mathbb{Z}^{++}$ times has an eigenspace of dimension $k \iff A$ is diagonalizable.

Definition 4.4.8: Diagonalizable Matrices

A square matrix A is said to be diagonalizable if there exists an invertible matrix P and a diagonal matrix D such that:

$$A = PDP^{-1}.$$

In this case, the diagonal entries of D are the eigenvalues of A, and the columns of P are the corresponding eigenvectors of A.

Key Conditions:

- A is diagonalizable if and only if A has m linearly independent eigenvectors, where m is the size of A.
- If A has m distinct eigenvalues, then A is guaranteed to be diagonalizable.
- If an eigenvalue λ of A has algebraic multiplicity k, the dimension of its eigenspace (geometric multiplicity) must also be k for A to be diagonalizable.

Diagonalizable Matrices: Suppose A has m independent eigenvectors, $\overline{v_1}, \overline{v_2}, \dots, \overline{v_m}$.

Let

$$P = [\overline{v_1}\ \overline{v_2}\ \dots\ \overline{v_m}]$$

Clearly, P is invertible.

$$AP = [A\overline{v_1}\ A\overline{v_2}\ ...\ A\overline{v_m}]$$

$$= [\lambda_1\overline{v_1}\ \lambda_2\overline{v_2}\ ...\ \lambda_m\overline{v_m}]$$

because

$$\overline{v_1}...\overline{v_m}$$

are eigenvectors

$$= [\overline{v_1}\ \overline{v_2}\ ...\ \overline{v_m}]\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m\end{bmatrix}$$

$$\Rightarrow AP = PD$$

$$\Rightarrow P^{-1}AP = D$$

$$P^{-1}AP \text{ is Diagonal}$$

Suppose

$$P^{-1}AP = \begin{bmatrix}d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m\end{bmatrix}$$

Multiply both sides by P

$$AP = PD$$

suppose

$$P = \begin{bmatrix}\overline{\vec{w}_1} & \overline{\vec{w}_2} & \dots & \overline{\vec{w}_m}\end{bmatrix}$$

Then

$$\therefore \begin{bmatrix}A\overline{\vec{w}_1} & A\overline{\vec{w}_2} & \dots & A\overline{\vec{w}_m}\end{bmatrix} = \begin{bmatrix}d_1\overline{\vec{w}_1} & d_2\overline{\vec{w}_2} & \dots & d_m\overline{\vec{w}_m}\end{bmatrix}$$

$$\begin{cases}A\overline{\vec{w}_1} & = d_1\overline{\vec{w}_1} \\ A\overline{\vec{w}_2} & = d_2\overline{\vec{w}_2} \\ & \vdots \\ A\overline{\vec{w}_m} & = d_m\overline{\vec{w}_m}\end{cases}$$

😄

Example 4.4.6 (Diagonalizable Matricies)

$$A = \begin{bmatrix}1 & 1 \\ 0 & 0\end{bmatrix} \Rightarrow \text{Entries are not linearly independent} \iff \text{Non invertible}$$

for $\lambda_1 = 1$ solve for nullspace of $A - \lambda_1 I$

$$A - \lambda_1 I = \begin{bmatrix}0 & 1 \\ 0 & -1\end{bmatrix} \to \begin{bmatrix}0 & 1 \\ 0 & 0\end{bmatrix} \Rightarrow Null(A - \lambda_1 I) = \left\{t_1\begin{bmatrix}1 \\ 0\end{bmatrix} : t_1 \in \mathbb{R}\right\}$$

for $\lambda_2 = 0$ solve for nullspace of $A - \lambda_2 I$

$$A - \lambda_1 I = \begin{bmatrix}1 & 1 \\ 0 & 0\end{bmatrix} \Rightarrow Null(A - \lambda_1 I) = \left\{t_2\begin{bmatrix}-1 \\ 1\end{bmatrix} : t_2 \in \mathbb{R}\right\}$$

$$\therefore \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} \text{are linearly independent}$$

$$\text{Eigenvectors are linearly independent} \iff \text{Eigenvalues are distinct} \iff \text{Diagonalizable}$$

$$B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \Rightarrow \text{Entries are linearly independent} \iff \text{Invertible}$$

$$\lambda_1 = 1, \lambda_2 = 1 \Rightarrow \text{Repeting } k = 2 \text{ times}$$

$$B - \lambda_1 I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow Null(B - \lambda_1 I) = \left\{ t_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} : t_1 \in \mathbb{R} \right\} \Rightarrow Dim(Eigenspace) = 1 < k = 2$$

$$\therefore B \text{ is not diagonalizable}$$

4.5 Consequence of Diagonalizability

Definition 4.5.1: Diagonalizable

A matrix A is diagonalizable if there exists an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

Note:-

If A is diagonalizable, then $A = PDP^{-1}$, where D is a diagonal matrix of eigenvalues and P is a matrix whose columns are the corresponding eigenvectors.

Given $A = PDP^{-1}$, then $A^k = (PDP^{-1})^k$ for $k \in \mathbb{Z}^+$.

$$A^k = (PDP^{-1})(PDP^{-1}) \cdots (PDP^{-1}) = PD^kP^{-1}$$

Since D is a diagonal matrix, D^k is easily computed by raising each diagonal element (eigenvalue) to the power of k.

Theorem 4.5.1 Eigenvalues of A^k

If λ is an eigenvalue of A, then λ^k is an eigenvalue of A^k.

Note:-

Eigenvectors of A and A^k are the same.

4.5.1 Markov Chain Problems

Definition 4.5.2: Markov Chain

A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

Note:-

Population members can fall into states as a function of time that is indexed by steps.

Example 4.5.1 (People in Boston)

Consider a population of people who either live in the city of Boston or in the suburbs. Let's denote:

- State 1: City (C)
- State 2: Suburbs (S)

Suppose that initially, the distribution is:

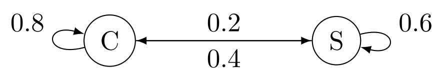
- 10% move to the city (C)
- 90% move to the suburbs (S)

Note:-

Markov Assumption: The fraction of population that changes from one state to another is always the same.

Note:-

Markov Diagram:



Question 16: Question

What fraction of people in Boston will be living in the city after 20 years?

Solution:

Note:-

State vector for year k: $\vec{x}_k = \begin{bmatrix} c \\ s \end{bmatrix}$, where c and s are the fractions of people in the city and suburbs, respectively. These elements are probabilities, so it is a probability vector.

Note:-

Initial state vector: $\vec{x}_0 = \begin{bmatrix} c \\ s \end{bmatrix}$

Note:-

Year 1: $\vec{x}_1 = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix} \begin{bmatrix} c \\ s \end{bmatrix}$

Definition 4.5.3: State Transition Matrix

$A = [a_{ij}]$ is the fraction of people in state i at time (k) that end up in state j at $(k+1)$.

Note:-

Year k: $\vec{x}_k = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix}^k \vec{x}_0$. How to find A^k?

Note:-

If A is known to be diagonalizable with real eigenvalues and eigenvectors, the matrix A is a stochastic matrix.

Theorem 4.5.2 Properties of Stochastic Matrices

1. $\lambda = 1$ is always an eigenvalue.
2. If there are no 0 values, all other eigenvalues will have $|\lambda| < 1$.

Theorem 4.5.3 Consequence

$$\lim_{k\to\infty} \vec{x}_k$$

is the probability vector that is an eigenvector corresponding to $\lambda = 1$.

Example 4.5.2 (Back to Example)

$A = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix}$

$\det(A - \lambda I) = (0.8 - \lambda)(0.6 - \lambda) - 0.2 \cdot 0.4 = 0$

$(0.8 - \lambda)(0.6 - \lambda) - 0.08 = \lambda^2 - 1.4\lambda + 0.4 = 0$ $(\lambda - 1)(\lambda - 0.4) = 0$

Eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 0.4$.

For $\lambda = 1$:

$$A - I = \begin{bmatrix} -0.2 & 0.4 \\ 0.2 & -0.4 \end{bmatrix} \to \begin{bmatrix} -0.2 & 0.4 \\ 0 & 0 \end{bmatrix}$$

$-0.2x_1 + 0.4x_2 = 0$, so $x_1 = 2x_2$.

$\text{Null}(A - I) = span\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\}$.

Probability vector: $\begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$

Note:-

Therefore, in the long run, $\frac{2}{3}$ will live in the city and $\frac{1}{3}$ will live in the suburbs.

4.6 Projections

Chapter 5

Projections

Definition 5.0.1: Projection of a vector v onto a line S

Let $S = \{t\mathbf{a} \mid t \in \mathbb{R}\}$ be a line spanned by a vector $\mathbf{a}$. The projection of a vector $\mathbf{v}$ onto S is the vector $\mathbf{x}$ such that $\mathbf{x} \in S$ and $\mathbf{v} - \mathbf{x}$ is orthogonal to S.

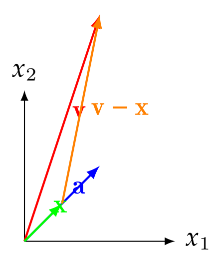
5.0.1 Finding the Projection

Question 17: Question

Find the projection of $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ onto $\mathbf{a}$.

Note:-

We have that $\mathbf{x} = p\mathbf{a}$, where $p \in \mathbb{R}$. Also, $\mathbf{a} \perp (\mathbf{v} - \mathbf{x})$.



Solution: Since $\mathbf{a} \cdot (\mathbf{v} - \mathbf{x}) = 0$, we have $\mathbf{a} \cdot (\mathbf{v} - p\mathbf{a}) = 0$, which implies $\mathbf{a} \cdot \mathbf{v} = p\mathbf{a} \cdot \mathbf{a}$. Therefore, $p = \frac{\mathbf{a}^T\mathbf{v}}{\mathbf{a}^T\mathbf{a}}$, and $\mathbf{x} = p\mathbf{a} = \frac{\mathbf{a}^T\mathbf{v}}{\mathbf{a}^T\mathbf{a}}\mathbf{a}$. We can rewrite $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{a}\frac{\mathbf{a}^T\mathbf{v}}{\mathbf{a}^T\mathbf{a}} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}\mathbf{v} = P\mathbf{v}$$

where $P = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}$ is the projection matrix. In this case, P is a 2×2 matrix.

Example 5.0.1 (Example)

Let $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Then $\mathbf{a}^T\mathbf{a} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 + 1 = 2$. Also,

$$P = \frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix}}{2} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Then,
$$P\mathbf{v} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} + \frac{3}{2} \\ \frac{1}{2} + \frac{3}{2} \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \mathbf{x}.$$

5.1 Generalization to $\mathbb{R}^m$ for $m > 2$

Let $S = \{\mathbf{a}_1 t_1 + \mathbf{a}_2 t_2 + \cdots + \mathbf{a}_k t_k \mid t_i \in \mathbb{R}, k < m\}$.

Note:-

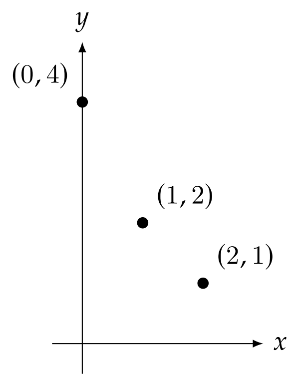
If $\sum_i \mathbf{a}_i t_i = 0$, then $t_i = 0$ for all i. This implies that $A = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\}$ and $\operatorname{rank}(A) = k$.

The projection matrix is given by
$$P = A(A^T A)^{-1} A^T.$$

Chapter 6

Statistics

6.1 Fitting a Straight Line to Data Points



Question 18: *

Can we draw a straight line through the given three points?

Solution: No, the points do not lie on a straight line.

Question 19: *

How can we phrase the previous question in linear algebra terms, solving for m and b in $y = mx + b$ by forming a system of linear equations? Figure out the values of a_{ij}.

Solution: We can represent the problem as a system of linear equations:

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

where

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} m \\ b \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$$

6.1.1 Solving the System

The augmented matrix A' is:

$$A' = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix}$$

We perform row reduction to find the row echelon form (REF):

1. Swap rows 1 and 2:

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

2. Subtract 2 times row 1 from row 3:

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

3. Add row 2 to row 3:

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

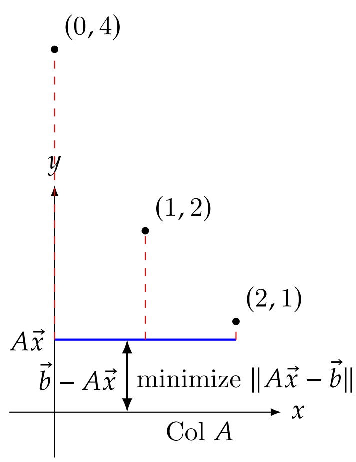
The REF of A' has a pivot in the last column. This indicates that the system $A\vec{x} = \vec{b}$ has no solution, which confirms that the three points do not lie on a straight line.

6.1.2 Best Fit Straight Line

Since we cannot find a line that passes through all three points, we seek the "best fit" straight line through the given data points.

Note:-

Instead of requiring $A\vec{x} - \vec{b} = 0$, we require $\|A\vec{x} - \vec{b}\|^2$ to be as small as possible.



Note:-

Find an $\vec{x}$ such that $A\vec{x} - \vec{b}$ is as small as possible in its length.

6.1.3 Finding the Optimal $\vec{x}$

We can find this $\vec{x}$ by requiring $(A\vec{x} - \vec{b})$ to be orthogonal to $A\vec{x}$. This means the error vector is orthogonal to the column space of A.

$$(A\vec{x} - \vec{b}) \text{ is orthogonal to } A\vec{x}$$
$$(A\vec{x} - \vec{b})^T (A\vec{x}) = 0$$
$$(A\vec{x})^T (A\vec{x} - \vec{b}) = 0$$
$$\vec{x}^T A^T (A\vec{x} - \vec{b}) = 0$$
$$\vec{x}^T A^T A\vec{x} = \vec{x}^T A^T \vec{b}$$

$$A^T A\vec{x} = A^T\vec{b}$$

If $A^T A$ is invertible, then

Definition 6.1.1: Least Squares Solution

The least squares solution to the system $A\vec{x} = \vec{b}$ is given by

$$\boxed{\vec{x} = (A^T A)^{-1} A^T \vec{b}}$$

Example 6.1.1 (Least Squares Fit Example)

Consider the matrix

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

Then the transpose of A is

$$A^T = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$$

Now we compute $A^T A$:

$$A^T A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix}$$

The determinant of $A^T A$ is $|A^T A| = (5)(3) - (3)(3) = 15 - 9 = 6$. Since the determinant is non-zero, $A^T A$ is invertible. The inverse of $A^T A$ is

$$(A^T A)^{-1} = \frac{1}{6} \begin{bmatrix} 3 & -3 \\ -3 & 5 \end{bmatrix}$$

Now suppose we want to find the best least squares fit to the data points $(0, 4)$, $(1, 2)$, and $(2, 1)$. We can represent this as $y = mx + b$, where we want to find $x = \begin{bmatrix} m \\ b \end{bmatrix}$. Let $\vec{b} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$. Then the least squares solution $\bar{x}$ is given by

$$\bar{x} = (A^T A)^{-1} A^T \vec{b} = \frac{1}{6} \begin{bmatrix} 3 & -3 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$$

This represents the line $y = mx + b$ that is the best least squares fit to the 3 data points given.

Theorem 6.1.1 Invertibility of $A^T A$

If A is an $m \times k$ matrix and has k independent columns, then $A^T A$ is invertible.

6.2 Proof of Invertibility of $A^T A$

Question 20: Why is $A^T A$ invertible when A has independent columns?

We need to show that the nullspace of $A^T A$ contains only the zero vector.

Solution: Suppose $\vec{x} \in \mathrm{Nul}(A)$. Then $A\vec{x} = \vec{0}$. Multiplying by A^T, we get $A^T A\vec{x} = A^T\vec{0} = \vec{0}$. Thus, $\vec{x} \in \mathrm{Nul}(A^T A)$.

Now suppose $\vec{x} \in \mathrm{Nul}(A^T A)$. Then $A^T A\vec{x} = \vec{0}$. Multiplying by $\vec{x}^T$ on the left, we get $\vec{x}^T A^T A\vec{x} = \vec{x}^T\vec{0} = 0$. This can be rewritten as $(A\vec{x})^T(A\vec{x}) = 0$, which is equivalent to $\|A\vec{x}\|^2 = 0$. This implies that $A\vec{x} = \vec{0}$, so $\vec{x} \in \mathrm{Nul}(A)$.

Therefore, $\text{Nul}(A^T A) = \text{Nul}(A)$.

Note:-

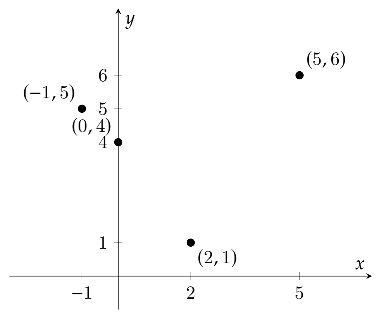
Since A has k independent columns, the nullspace of A contains only the zero vector.

Note:-

Since the nullspace of $A^T A$ contains only the zero vector, $A^T A$ is invertible.

6.3 Parabola Fitting

Consider the problem of finding a parabola that passes through the points $(-1, 5)$, $(0, 4)$, and $(5, 6)$. We are given that the general form of a parabola is $y = ax^2 + bx + c$. Our goal is to determine the coefficients a, b, and c.



For each point (x, y), we substitute the x and y values into the general equation of the parabola.

1. Point $(-1, 5)$: Substituting $x = -1$ and $y = 5$ gives us: $5 = a(-1)^2 + b(-1) + c$, which simplifies to $a - b + c = 5$.
2. Point $(0, 4)$: Substituting $x = 0$ and $y = 4$ gives us: $4 = a(0)^2 + b(0) + c$, which simplifies to $c = 4$.
3. Point $(5, 6)$: Substituting $x = 5$ and $y = 6$ gives us: $6 = a(5)^2 + b(5) + c$, which simplifies to $25a + 5b + c = 6$.

We now have a system of three linear equations:

$$\begin{aligned} a - b + c &= 5 \\ c &= 4 \\ 25a + 5b + c &= 6 \end{aligned}$$

6.3.1 Matrix Representation

This system of equations can be represented in matrix form as $A\vec{x} = \vec{b}$, where:

- A is the coefficient matrix:

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 25 & 5 & 1 \end{bmatrix}$$

- $\vec{x}$ is the vector of unknowns (the coefficients a, b, and c):

$$\vec{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

- $\vec{b}$ is the vector of constants (the y values):

$$\vec{b} = \begin{bmatrix} 5 \\ 4 \\ 6 \end{bmatrix}$$

Thus, we have:

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 25 & 5 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 6 \end{bmatrix}$$

Note:-

The matrix representation is a compact way to express the system of linear equations. It allows us to use linear algebra techniques to solve for the unknowns a, b, and c.

6.3.2 Overdetermined System

The original image also includes a point (2,1) and creates a 4x3 system, which means more equations than unknowns. This is known as an overdetermined system. In such cases, a unique solution might not exist. We typically seek a "best fit" solution, often obtained using techniques like least squares. The system would then look like:

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 4 & 2 & 1 \\ 25 & 5 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 1 \\ 6 \end{bmatrix}$$

Note:-

When dealing with overdetermined systems, the goal shifts from finding an exact solution to finding the best possible approximation. The least squares method minimizes the sum of the squares of the errors between the predicted values (from the equation) and the actual values.

6.3.3 Solving the System

Least Squares (for overdetermined systems): If the system is overdetermined, you can find the least squares solution by solving the normal equations: $A^T A\vec{x} = A^T\vec{b}$. This gives you the coefficients that minimize the error.

Once you find the values of a, b, and c, you can substitute them back into the equation $y = ax^2 + bx + c$ to obtain the equation of the parabola.

6.4 Orthogonality and Orthonormality

6.4.1 Orthogonal Matrix

Definition 6.4.1: Orthogonal Matrix

An orthogonal matrix is a square matrix A of size $m \times m$ such that all of its columns are orthogonal and non-zero.

Note:-

Let A be an orthogonal matrix. We can represent A as a matrix whose columns are vectors:

$$A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \end{bmatrix}$$

Note:-

The columns of an orthogonal matrix satisfy the following properties:

1. $\vec{v}_i \cdot \vec{v}_j = 0$ whenever $i \neq j$ (orthogonality).
2. $||\vec{v}_j|| \neq 0$ for all j (non-zero vectors).

Question 21: Is an Orthogonal Matrix Invertible?

Is an orthogonal matrix always an invertible matrix?

Solution: Yes.

Note:-

To demonstrate this, let's consider the matrix-vector product:

$$A\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m$$

Note:-

Now, let's take the dot product of the first column vector $\vec{v}_1$ with the result of this matrix-vector product:

$$\vec{v}_1^T(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m) = x_1||\vec{v}_1||^2 + x_2(\vec{v}_1 \cdot \vec{v}_2) + \cdots + x_m(\vec{v}_1 \cdot \vec{v}_m)$$

Since the vectors are orthogonal, all dot products $\vec{v}_1 \cdot \vec{v}_j$ for $j \neq 1$ are zero. Therefore, we get:

$$\vec{v}_1^T(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m) = x_1||\vec{v}_1||^2 = 0$$

Similarly, for other columns:

$$\vec{v}_2^T(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m) = x_2||\vec{v}_2||^2 = 0$$

$$\vdots$$

$$\vec{v}_m^T(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m) = x_m||\vec{v}_m||^2 = 0$$

Note:-

Since $||\vec{v}_j|| \neq 0$, it follows that $x_1, x_2, \ldots, x_m$ must all be zero. This implies that the columns $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ are linearly independent.

Note:-

A matrix with linearly independent columns is invertible. Therefore, A must be invertible.

6.4.2 Orthonormal Matrix

Definition 6.4.2: Orthonormal Matrix

An orthonormal matrix A of size $m \times m$ is an orthogonal matrix where each of its columns has a norm (magnitude) of 1.

Note:-

In other words, the columns of an orthonormal matrix are not only orthogonal to each other but also are unit vectors.

Example 6.4.1 (Example)

Let's examine some matrices and determine if they are orthogonal and/or orthonormal.

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

- $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
- $C = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

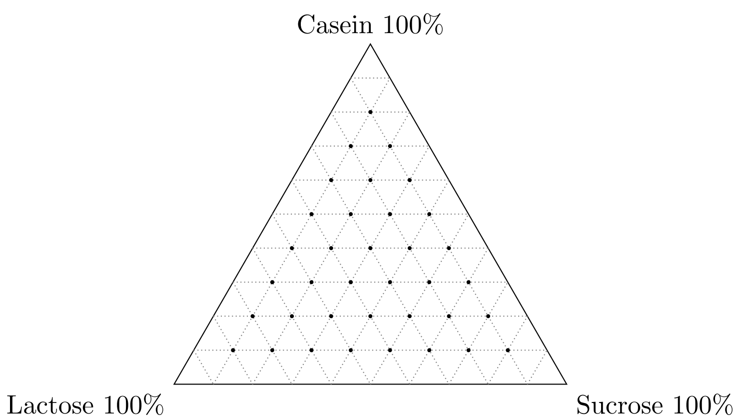
Note:-

- Matrix A is orthogonal, but not orthonormal because the columns are orthogonal, but their norms are not 1.
- Matrix B is both orthogonal and orthonormal because the columns are orthogonal and have a norm of 1.
- Matrix C is both orthogonal and orthonormal because the columns are orthogonal and have a norm of 1. This matrix represents a rotation in the plane.

6.5 Application: Near-Infrared (NIR) Spectroscopy for Quality Control of Mixtures

6.5.1 Context: Analyzing Mixtures

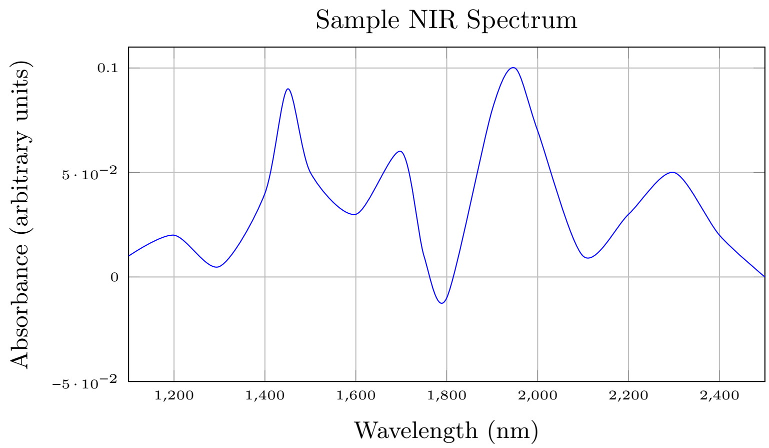
Near-Infrared (NIR) spectroscopy is a technique used to analyze the composition of materials. In this example, it's applied to determine the relative amounts of three components (Casein, Lactose, Sucrose) in various mixture samples. A total of 231 different mixture samples were prepared and analyzed.



Ternary diagram illustrating the composition space of the three components. Each point inside the triangle represents a unique mixture composition. A total of 231 mixture samples were analyzed.

6.5.2 Data Representation: Spectra as Vectors

Each mixture sample yields a spectrum when analyzed by NIR spectroscopy. This spectrum measures the absorbance (or reflectance) of light at various wavelengths. In this specific case, measurements are taken at 700 distinct wavelengths.



Example NIR spectrum for one mixture sample. The horizontal axis represents the wavelength of light, and the vertical axis represents the measured absorbance. Data is collected at 700 discrete wavelengths.

Mathematically, we can represent the spectrum of a single sample as a vector. Since there are 700 measurements (one for each wavelength), each spectrum corresponds to a vector in a 700-dimensional space, $\mathbb{R}^{700}$.

Definition 6.5.1: Spectrum Vector

Let $\vec{s}$ be the vector representing a single NIR spectrum. The components of $\vec{s}$ are the absorbance values at the 700 measured wavelengths:

$$\vec{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{700} \end{bmatrix} \in \mathbb{R}^{700}$$

where s_i is the absorbance at the i-th wavelength.

Note:-

Often, raw spectral data is pre-processed. A common step is mean-centering, where the average spectrum (calculated across all samples) is subtracted from each individual spectrum. This results in data where the mean of each feature (wavelength) across all samples is zero. The notes mention replacing data with a mean of zero, implying mean-centering might have been performed.

The entire dataset consists of 231 such vectors, which can be organized into a matrix.

Definition 6.5.2: Data Matrix

Let A be the matrix containing all 231 sample spectra. Each row of A corresponds to the spectrum vector of one sample.

$$A = \begin{bmatrix} \longleftarrow & \vec{s}_1^T & \longrightarrow \\ \longleftarrow & \vec{s}_2^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{s}_{231}^T & \longrightarrow \end{bmatrix}$$

Here, $\vec{s}_i^T$ denotes the row vector representing the spectrum of the i-th sample. The matrix A has dimensions 231×700.

6.5.3 Goal: Principal Components Analysis (PCA)

The primary goal is to analyze this high-dimensional dataset (231 points in $\mathbb{R}^{700}$) using Principal Components Analysis (PCA). PCA aims to find the directions (principal components) in the 700-dimensional space along which the data varies the most. These directions can reveal underlying patterns, reduce dimensionality, and help visualize the relationships between the different mixture samples.

Note:-

PCA finds a new set of orthogonal basis vectors (the principal components) for the data space. The first principal component (PC1) is the direction of maximum variance. The second principal component (PC2) is the direction of maximum variance among all directions orthogonal to PC1, and so on.

We will build up the concept of PCA, starting with a simple geometric interpretation.

6.6 Part One: Geometric Interpretation (Single Spectrum)

Question 22: Maximizing Projection

Given a single spectrum vector $\vec{s} \in \mathbb{R}^{700}$, how can we find a direction, represented by a unit vector $\vec{v} \in \mathbb{R}^{700}$ (meaning $||\vec{v}||^2 = \vec{v}^T\vec{v} = 1$), such that the projection of $\vec{s}$ onto $\vec{v}$ is maximized?

Solution: Let's recall the definition of the projection of a vector $\vec{s}$ onto a vector $\vec{v}$.

Definition 6.6.1: Projection of $\vec{s}$ onto $\vec{v}$

The scalar projection of $\vec{s}$ onto $\vec{v}$ is given by $\text{comp}_{\vec{v}}\vec{s} = \frac{\vec{s}\cdot\vec{v}}{||\vec{v}||}$. The vector projection of $\vec{s}$ onto $\vec{v}$ is given by $\text{proj}_{\vec{v}}\vec{s} = \left(\frac{\vec{s}\cdot\vec{v}}{||\vec{v}||^2}\right)\vec{v}$. The length (norm) of the vector projection is $||\text{proj}_{\vec{v}}\vec{s}|| = \left|\frac{\vec{s}\cdot\vec{v}}{||\vec{v}||^2}\right| ||\vec{v}|| = \frac{|\vec{s}\cdot\vec{v}|}{||\vec{v}||}$.

We are looking for a unit vector $\vec{v}$ ($||\vec{v}|| = 1$) that maximizes the length (or magnitude) of the projection, which simplifies to maximizing $|\vec{s}\cdot\vec{v}|$.

Recall the geometric definition of the dot product:

$$\vec{s}\cdot\vec{v} = ||\vec{s}||\,||\vec{v}||\cos\theta$$

where θ is the angle between the vectors $\vec{s}$ and $\vec{v}$.

Since we require $\vec{v}$ to be a unit vector, $||\vec{v}|| = 1$. Therefore:

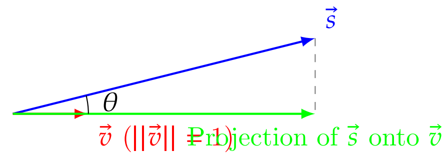
$$\vec{s}\cdot\vec{v} = ||\vec{s}||\cos\theta$$

To maximize the projection length $|\vec{s}\cdot\vec{v}| = ||\vec{s}||\,|\cos\theta|$, we need to maximize $|\cos\theta|$. The maximum value of $|\cos\theta|$ is 1, which occurs when $\theta = 0$ or $\theta = \pi$.

- If $\theta = 0$, $\vec{v}$ points in the same direction as $\vec{s}$. Then $\vec{v} = \frac{\vec{s}}{||\vec{s}||}$, and the projection length is $||\vec{s}||\cos 0 = ||\vec{s}||$.

- If $\theta = \pi$, $\vec{v}$ points in the opposite direction of $\vec{s}$. Then $\vec{v} = -\frac{\vec{s}}{||\vec{s}||}$, and the projection length is $||\vec{s}|||\cos\pi| = ||\vec{s}||$.

In both cases, the maximum projection length is $||\vec{s}||$. The direction $\vec{v}$ that maximizes this projection is aligned with the vector $\vec{s}$ itself.



Goal: Maximize length of green vector by choosing direction of red vector $\vec{v}$
Maximum occurs when $\theta = 0$, i.e., $\vec{v}$ is parallel to $\vec{s}$.

Note:-

For a single data point (vector) $\vec{s}$, the direction that "best represents" it or captures its magnitude is simply its own direction. PCA becomes more interesting when we consider multiple data points.

6.7 Part Two: Generalization to Two Spectra

Now, let's consider two spectrum vectors, $\vec{s}_1$ and $\vec{s}_2$, both in $\mathbb{R}^{700}$.

Question 23: Maximizing Sum of Squared Projections

Find a unit vector $\vec{v} \in \mathbb{R}^{700}$ ($||\vec{v}|| = 1$) such that the sum of the squared lengths of the projections of $\vec{s}_1$ and $\vec{s}_2$ onto $\vec{v}$ is maximized.

Solution: The length of the projection of $\vec{s}_1$ onto the unit vector $\vec{v}$ is $|\vec{s}_1 \cdot \vec{v}|$. The length of the projection of $\vec{s}_2$ onto the unit vector $\vec{v}$ is $|\vec{s}_2 \cdot \vec{v}|$.

We want to maximize the sum of the squares of these lengths:

$$L = (|\vec{s}_1 \cdot \vec{v}|)^2 + (|\vec{s}_2 \cdot \vec{v}|)^2 = (\vec{s}_1 \cdot \vec{v})^2 + (\vec{s}_2 \cdot \vec{v})^2$$

We can express the dot products using matrix notation. Let's define a matrix A where the rows are the (transposed) spectrum vectors:

$$A = \begin{bmatrix} \vec{s}_1^T \\ \vec{s}_2^T \end{bmatrix}$$

This is a 2×700 matrix. Now consider the product $A\vec{v}$:

$$A\vec{v} = \begin{bmatrix} \vec{s}_1^T \\ \vec{s}_2^T \end{bmatrix} \vec{v} = \begin{bmatrix} \vec{s}_1^T \vec{v} \\ \vec{s}_2^T \vec{v} \end{bmatrix} = \begin{bmatrix} \vec{s}_1 \cdot \vec{v} \\ \vec{s}_2 \cdot \vec{v} \end{bmatrix}$$

The resulting vector, let's call it $\vec{q} = A\vec{v}$, is a 2×1 vector (in $\mathbb{R}^2$) whose components are the scalar projections of $\vec{s}_1$ and $\vec{s}_2$ onto $\vec{v}$.

Now, let's look at the squared norm (squared length) of $\vec{q}$:

$$||\vec{q}||^2 = ||A\vec{v}||^2 = \left\| \begin{bmatrix} \vec{s}_1 \cdot \vec{v} \\ \vec{s}_2 \cdot \vec{v} \end{bmatrix} \right\|^2 = (\vec{s}_1 \cdot \vec{v})^2 + (\vec{s}_2 \cdot \vec{v})^2$$

This is exactly the quantity L we want to maximize!

So, the problem is equivalent to maximizing $||A\vec{v}||^2$ subject to the constraint $||\vec{v}|| = 1$.

Let's expand $||A\vec{v}||^2$ using matrix transpose properties:

$$||A\vec{v}||^2 = (A\vec{v})^T(A\vec{v}) = (\vec{v}^T A^T)(A\vec{v}) = \vec{v}^T(A^T A)\vec{v}$$

Let $C = A^T A$.

$$A^T = \begin{bmatrix} \vec{s}_1 & \vec{s}_2 \end{bmatrix} \quad (700 \times 2 \text{ matrix})$$

$$C = A^TA = \begin{bmatrix} \vec{s}_1 & \vec{s}_2 \end{bmatrix} \begin{bmatrix} \vec{s}_1^T \\ \vec{s}_2^T \end{bmatrix} \quad \text{(This is incorrect dimensions in notes, let's re-evaluate } A^TA\text{)}$$

Let's re-check the dimensions. A is 2×700. A^T is 700×2. $\vec{v}$ is 700×1. $A\vec{v}$ is 2×1. A^TA is $(700 \times 2) \times (2 \times 700) = 700 \times 700$. $\vec{v}^T$ is 1×700. $\vec{v}^T(A^TA)\vec{v}$ is $(1 \times 700) \times (700 \times 700) \times (700 \times 1) = 1 \times 1$ (a scalar), which is correct.

The matrix $C = A^TA$ is a 700×700 matrix. It is constructed as:

$$C = A^TA = \begin{bmatrix} \vec{s}_1 & \vec{s}_2 \end{bmatrix} \begin{bmatrix} \vec{s}_1^T \\ \vec{s}_2^T \end{bmatrix} = \vec{s}_1\vec{s}_1^T + \vec{s}_2\vec{s}_2^T \quad \text{(Still incorrect, this is outer product sum)}$$

Let's write A^T and A properly:

$$A^T = \begin{bmatrix} | & | \\ \vec{s}_1 & \vec{s}_2 \\ | & | \end{bmatrix} \quad (700 \times 2)$$

$$A = \begin{bmatrix} — & \vec{s}_1^T & — \\ — & \vec{s}_2^T & — \end{bmatrix} \quad (2 \times 700)$$

The product $C = A^TA$ is indeed a 700×700 matrix. The (i, j)-th element of C is the dot product of the i-th column of A^T and the j-th row of A. This doesn't seem right. Let's use the definition $C = A^TA$.

The problem is now: Maximize $\vec{v}^TC\vec{v}$ subject to $||\vec{v}|| = 1$, where $C = A^TA$.

This is a standard problem in linear algebra related to quadratic forms and eigenvalues.

6.8 Part Three: Generalization to Multiple Spectra

We now extend the idea to the full dataset of $m = 231$ spectra, $\{\vec{s}_1, \vec{s}_2, \dots, \vec{s}_{231}\}$, each in $\mathbb{R}^{700}$.

Question 24: Maximizing Total Sum of Squared Projections

Let A be the 231×700 data matrix where the i-th row is $\vec{s}_i^T$. Find a unit vector $\vec{v} \in \mathbb{R}^{700}$ ($||\vec{v}|| = 1$) that maximizes the sum of the squared lengths of the projections of all 231 spectra onto $\vec{v}$.

Solution: The quantity we want to maximize is:

$$L = \sum_{i=1}^{231} (\text{length of projection of } \vec{s}_i \text{ onto } \vec{v})^2 = \sum_{i=1}^{231} (\vec{s}_i \cdot \vec{v})^2$$

Following the same logic as in Part Two, we define the data matrix A:

$$A = \begin{bmatrix} \longleftarrow & \vec{s}_1^T & \longrightarrow \\ \longleftarrow & \vec{s}_2^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{s}_{231}^T & \longrightarrow \end{bmatrix} \quad (231 \times 700)$$

Consider the product $A\vec{v}$:

$$A\vec{v} = \begin{bmatrix} \vec{s}_1^T \\ \vec{s}_2^T \\ \vdots \\ \vec{s}_{231}^T \end{bmatrix} \vec{v} = \begin{bmatrix} \vec{s}_1^T\vec{v} \\ \vec{s}_2^T\vec{v} \\ \vdots \\ \vec{s}_{231}^T\vec{v} \end{bmatrix} = \begin{bmatrix} \vec{s}_1 \cdot \vec{v} \\ \vec{s}_2 \cdot \vec{v} \\ \vdots \\ \vec{s}_{231} \cdot \vec{v} \end{bmatrix} = \vec{q}$$

The vector $\vec{q} = A\vec{v}$ is a 231×1 vector containing the scalar projections of each spectrum onto $\vec{v}$.

The squared norm of $\vec{q}$ is:

$$||\vec{q}||^2 = ||A\vec{v}||^2 = \sum_{i=1}^{231} (\vec{s}_i \cdot \vec{v})^2 = L$$

Again, we need to maximize $||A\vec{v}||^2$ subject to $||\vec{v}|| = 1$. As before, we rewrite this using the matrix $C = A^TA$:

$$||A\vec{v}||^2 = (A\vec{v})^T(A\vec{v}) = \vec{v}^TA^TA\vec{v} = \vec{v}^TC\vec{v}$$

Here, A^T is 700×231, and A is 231×700. The matrix $C = A^T A$ is a 700×700 matrix.

The problem is: Maximize the quadratic form $\vec{v}^T C \vec{v}$ subject to the constraint $||\vec{v}||^2 = \vec{v}^T \vec{v} = 1$.

Note:-

The matrix $C = A^T A$ is closely related to the covariance matrix of the data, especially if the data matrix A is mean-centered. If A is mean-centered, then $C = A^T A$ is proportional to the sample covariance matrix (specifically, $C = (m-1) \times \text{Cov}$, where $m = 231$ is the number of samples). Maximizing $\vec{v}^T C \vec{v}$ is equivalent to finding the direction $\vec{v}$ along which the variance of the projected data points ($q_i = \vec{s}_i \cdot \vec{v}$) is maximized.

6.9 Mathematical Foundation: Eigenvalue Analysis of $\mathbf{C = A^T A}$

We need to find the unit vector $\vec{v}$ that maximizes $\vec{v}^T C \vec{v}$, where $C = A^T A$. Let's analyze the properties of the matrix C.

6.9.1 Properties of C

1. **Size:** C is a 700×700 matrix.
2. **Symmetry:** C is a symmetric matrix. Proof: $C^T = (A^T A)^T = (A)^T (A^T)^T = A^T A = C$.

Theorem 6.9.1 Spectral Theorem for Symmetric Matrices

Any real symmetric matrix C (like our $n \times n$ matrix $C = A^T A$, where $n = 700$) has the following properties:

- All eigenvalues of C are real numbers.
- C is orthogonally diagonalizable. This means there exists an orthogonal matrix P and a diagonal matrix D such that $C = PDP^T$.
- The columns of the matrix P are orthonormal eigenvectors of C. Let these eigenvectors be $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$.
- The diagonal entries of D are the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

$$D = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

- The relationship $C = PDP^T$ implies $C\vec{v}_i = \lambda_i \vec{v}_i$ for each $i = 1, \ldots, n$.
- Since P is orthogonal, $P^T P = PP^T = I$, and $P^{-1} = P^T$. The eigenvectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$ form an orthonormal basis for $\mathbb{R}^n$. That is, $\vec{v}_i \cdot \vec{v}_j = \vec{v}_i^T \vec{v}_j = \delta_{ij}$ (1 if $i = j$, 0 if $i \neq j$). In particular, $||\vec{v}_i||^2 = 1$ for all i.

6.9.2 Relationship to the Maximization Problem

We want to maximize $\vec{v}^T C \vec{v}$ subject to $||\vec{v}|| = 1$. Let the eigenvalues of C be ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{700}$. Let the corresponding orthonormal eigenvectors be $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{700}$.

Any unit vector $\vec{v} \in \mathbb{R}^{700}$ can be written as a linear combination of the orthonormal basis of eigenvectors:

$$\vec{v} = c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_{700} \vec{v}_{700}$$

Since $||\vec{v}|| = 1$ and the eigenvectors are orthonormal:

$$||\vec{v}||^2 = \vec{v}^T \vec{v} = (c_1 \vec{v}_1 + \cdots + c_{700} \vec{v}_{700})^T (c_1 \vec{v}_1 + \cdots + c_{700} \vec{v}_{700}) = \sum_{i=1}^{700} c_i^2 = 1$$

Now let's evaluate the quadratic form $\vec{v}^T C\vec{v}$:

$$\begin{aligned}
\vec{v}^T C\vec{v} &= \vec{v}^T C\left(\sum_{j=1}^{700} c_j\vec{v}_j\right) \\
&= \vec{v}^T\left(\sum_{j=1}^{700} c_j(C\vec{v}_j)\right) \\
&= \vec{v}^T\left(\sum_{j=1}^{700} c_j(\lambda_j\vec{v}_j)\right) \quad \text{(since } C\vec{v}_j = \lambda_j\vec{v}_j\text{)} \\
&= \left(\sum_{i=1}^{700} c_i\vec{v}_i\right)^T\left(\sum_{j=1}^{700} c_j\lambda_j\vec{v}_j\right) \\
&= \sum_{i=1}^{700}\sum_{j=1}^{700} c_i c_j\lambda_j(\vec{v}_i^T\vec{v}_j)
\end{aligned}$$

Since $\vec{v}_i^T\vec{v}_j = \delta_{ij}$ (it's 1 if $i = j$ and 0 otherwise), the double summation simplifies:

$$\vec{v}^T C\vec{v} = \sum_{i=1}^{700} c_i^2\lambda_i$$

So, we want to maximize $\sum_{i=1}^{700} c_i^2\lambda_i$ subject to the constraint $\sum_{i=1}^{700} c_i^2 = 1$.

Since $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{700}$, we have:

$$\sum_{i=1}^{700} c_i^2\lambda_i \leq \sum_{i=1}^{700} c_i^2\lambda_1 = \lambda_1\left(\sum_{i=1}^{700} c_i^2\right) = \lambda_1(1) = \lambda_1$$

The maximum value of $\vec{v}^T C\vec{v}$ is λ_1, the largest eigenvalue of C. This maximum is achieved when the expression becomes an equality. This happens if all the "weight" c_i^2 is on the term corresponding to λ_1. That is, when $c_1^2 = 1$ and $c_i^2 = 0$ for $i > 1$. This means $c_1 = \pm 1$ and $c_i = 0$ for $i > 1$. In this case, the vector $\vec{v}$ is $\vec{v} = (\pm 1)\vec{v}_1 + 0 + \cdots + 0 = \pm\vec{v}_1$.

Therefore, the unit vector $\vec{v}$ that maximizes $\vec{v}^T C\vec{v}$ is the eigenvector $\vec{v}_1$ corresponding to the largest eigenvalue λ_1 (or its negative, $-\vec{v}_1$). This vector $\vec{v}_1$ is the first principal component of the data.

6.9.3 Non-negativity of Eigenvalues

Theorem 6.9.2 Eigenvalues of $A^T A$

The eigenvalues of the matrix $C = A^T A$, where A is any real matrix, are always non-negative ($\lambda_i \geq 0$).

Justification: Let λ_i be an eigenvalue of C and $\vec{v}_i$ be the corresponding normalized eigenvector ($||\vec{v}_i|| = 1$). We know $C\vec{v}_i = \lambda_i\vec{v}_i$. Substitute $C = A^T A$:

$$A^T A\vec{v}_i = \lambda_i\vec{v}_i$$

Now, consider the squared norm of the vector $A\vec{v}_i$:

$$||A\vec{v}_i||^2 = (A\vec{v}_i)^T(A\vec{v}_i)$$

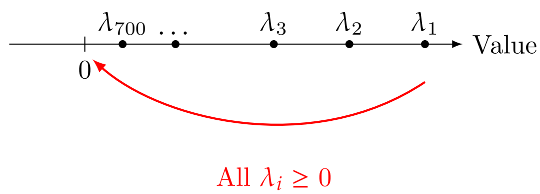
This squared norm must be non-negative, i.e., $||A\vec{v}_i||^2 \geq 0$. Let's expand the expression:

$$\begin{aligned} ||A\vec{v}_i||^2 &= (\vec{v}_i^T A^T)(A\vec{v}_i) \\ &= \vec{v}_i^T (A^T A)\vec{v}_i \\ &= \vec{v}_i^T (C\vec{v}_i) \\ &= \vec{v}_i^T (\lambda_i \vec{v}_i) \quad \text{(using the eigenvalue equation)} \\ &= \lambda_i (\vec{v}_i^T \vec{v}_i) \\ &= \lambda_i ||\vec{v}_i||^2 \end{aligned}$$

Since the eigenvector $\vec{v}_i$ is normalized, $||\vec{v}_i||^2 = 1$. So, we have:

$$||A\vec{v}_i||^2 = \lambda_i$$

Since $||A\vec{v}_i||^2 \geq 0$, it directly follows that $\lambda_i \geq 0$. All eigenvalues of $C = A^T A$ must be non-negative.



Schematic representation of the non-negative eigenvalues of $C = A^T A$.

6.10 Summary: Finding Principal Components

Principal Components Analysis (PCA) for a data matrix A (with m samples as rows and n features as columns, here $m = 231, n = 700$) involves the following steps, based on the analysis above:

1. **Mean-Center Data (Optional but Recommended):** Subtract the mean of each column (feature/wavelength) from the corresponding column entries in A. Let the resulting matrix still be denoted by A.

2. **Form the Covariance-Related Matrix:** Calculate $C = A^T A$. This is an $n \times n$ (700×700) symmetric matrix. (Alternatively, calculate the actual covariance matrix $\text{Cov} = \frac{1}{m-1} A^T A$).

3. **Find Eigenvalues and Eigenvectors:** Solve the eigenvalue problem for C: $C\vec{v} = \lambda\vec{v}$. This yields n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (which are all ≥ 0) and their corresponding eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$.

4. **Order Eigenpairs:** Sort the eigenvalues in descending order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Reorder the corresponding eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ accordingly. Ensure the eigenvectors are normalized ($||\vec{v}_i|| = 1$) and mutually orthogonal ($\vec{v}_i \cdot \vec{v}_j = 0$ for $i \neq j$).

5. **Identify Principal Components:**

   - The first principal component (PC1) is the eigenvector $\vec{v}_1$ corresponding to the largest eigenvalue λ_1. This is the direction in $\mathbb{R}^{700}$ that captures the maximum variance in the data.
   - The second principal component (PC2) is the eigenvector $\vec{v}_2$ corresponding to the second largest eigenvalue λ_2. This is the direction orthogonal to PC1 that captures the maximum remaining variance.
   - ...and so on. The k-th principal component (PCk) is the eigenvector $\vec{v}_k$.

6. **Project Data (Optional):** To visualize the data or perform dimensionality reduction, project the original data points (rows of A) onto the first few principal components. The coordinates of the i-th data point $\vec{s}_i$ in the new PC basis are given by the components of the vector $A\vec{v}_j$ or simply $\vec{s}_i \cdot \vec{v}_j$. For example, the coordinates in the PC1-PC2 plane are $(\vec{s}_i \cdot \vec{v}_1, \vec{s}_i \cdot \vec{v}_2)$.

The eigenvalue λ_k represents the variance of the data when projected onto the k-th principal component $\vec{v}_k$. Specifically, $\lambda_k = ||A\vec{v}_k||^2 = \sum_{i=1}^{m} (\vec{s}_i \cdot \vec{v}_k)^2$. If data was mean-centered and the actual covariance matrix was used, λ_k would be the variance directly.