1 Vectors

Definition

$\mathbb{R}^2$ is the set of vectors in the plane.
$\mathbb{R}^3$ is the set of vectors in space.
$\mathbb{R}^n = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ with $a_1, \ldots, a_n \in \mathbb{R}$ is the set of vectors with n components.

Binary vectors and modular arithmetic: The set of all m-ary vectors of length n is denoted by $\mathbb{Z}_m^n$.

Definition

Velocity, speed and acceleration in $\mathbb{R}^2$:

- Displacement at time t: $\mathbf{s}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$
- Velocity at time t: $\mathbf{v}(t) = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix}$
- Speed at time t: $||\mathbf{v}(t)|| = \sqrt{(x'(t))^2 + (y'(t))^2}$
- Acceleration at time t: $\mathbf{a}(t) = \mathbf{v}'(t) = \begin{pmatrix} x''(t) \\ y''(t) \end{pmatrix}$

1.1 The Dot Product

Definition

The **dot product** of vectors $\mathbf{u}, \mathbf{v}$ is a similarity measure:

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n \tag{1}$$
$$= ||\mathbf{u}|| ||\mathbf{v}|| \cos(\theta) \tag{2}$$

$\cos(\theta)$ is largest when the two vectors point in the same direction, because $\cos(0) = 1$.

Theorem

Rules of calculation: Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n, c \in \mathbb{R}$. Then

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ (commutativity)
2. $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ (distributivity)
3. $(c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$
4. $\mathbf{u} \cdot \mathbf{u} = ||\mathbf{u}||^2 \geq 0$

Theorem

Suppose $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$. Then

1. $\mathbf{u} \cdot \mathbf{v} > 0 \iff \theta \in [0, \frac{\pi}{2}]$ (acute angle)
2. $\mathbf{u} \cdot \mathbf{v} = 0 \iff \theta = \frac{\pi}{2}$ (**orthogonal**)
3. $\mathbf{u} \cdot \mathbf{v} < 0 \iff \theta \in [\frac{\pi}{2}, \pi]$ (obtuse angle)
4. The zero vector is orthogonal to all vectors: $\mathbf{0} \cdot \mathbf{u} = 0$
5. $\mathbf{u} \cdot \mathbf{u} = ||\mathbf{u}|| ||\mathbf{u}|| \cos 0 = ||\mathbf{u}||^2$

Definition

The **norm** of a vector is given by:

$$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}} \geq 0$$

If the norm of a vector is equal to one, it is **normalised**.

If a set of vectors are mutually orthogonal and normalised, they are **orthonormal**.

Theorem

The angle formula: For non-zero vectors $\mathbf{u}$ and $\mathbf{v}$:

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}|| ||\mathbf{v}||}$$

Theorem

The **distance** between two vectors:

$$d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}|| = \sqrt{(u_1 - v_1)^2 + \cdots + (u_n - v_n)^2}$$

Theorem

Let P be a point on a circle with centre C. The line tangent to the circle at P is orthogonal to the vector $\vec{CP}$.

The Pythagorean Theorem

$||\mathbf{v} - \mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2$ if and only if $\mathbf{v}, \mathbf{w}$ are orthogonal.

Proof:

$$||\mathbf{v} - \mathbf{w}||^2 = (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w}) \tag{3}$$
$$= \mathbf{v} \cdot \mathbf{v} + \mathbf{v} \cdot (-\mathbf{w}) + (-\mathbf{w}) \cdot \mathbf{v} + (-\mathbf{w}) \cdot (-\mathbf{w}) \quad \text{by rule 2} \tag{4}$$
$$= \mathbf{v} \cdot \mathbf{v} - 2\mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} \quad \text{by rules 1 and 2} \tag{5}$$
$$= ||\mathbf{v}||^2 - 2\mathbf{v} \cdot \mathbf{w} + ||\mathbf{w}||^2 \tag{6}$$
$$= ||\mathbf{v}||^2 + ||\mathbf{w}||^2 \quad \text{if and only if } \mathbf{v} \cdot \mathbf{w} = 0 \tag{7}$$

Cauchy-Schwarz Inequality

$$|\mathbf{u} \cdot \mathbf{v}| \leq ||\mathbf{u}|| ||\mathbf{v}||$$

Proof: TODO

Triangle Inequality

$$|\mathbf{u} + \mathbf{v}| \leq ||\mathbf{u}|| + ||\mathbf{v}||$$

Proof: TODO

1.2 The Cross Product

Definition

For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, the **cross product** of $\mathbf{u}$ and $\mathbf{v}$ is:

$$\mathbf{v} \times \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}$$

Definition

Rules of calculation:

1. **The right-hand rule**: $\mathbf{v} \times \mathbf{u} = -(\mathbf{u} \times \mathbf{v})$
2. It follows that $\mathbf{u} \times \mathbf{u} = \mathbf{0}$; and $\mathbf{u} \times \mathbf{0} = \mathbf{0}$.
3. $\mathbf{u} \times k\mathbf{v} = k(\mathbf{u} \times \mathbf{v})$
4. $\mathbf{u} \times k\mathbf{u} = \mathbf{0}$
5. $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$
6. $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$
7. $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{w})\mathbf{w}$
8. $||\mathbf{u} \times \mathbf{v}||^2 = ||\mathbf{u}||^2||\mathbf{v}||^2 - (\mathbf{u} \cdot \mathbf{v})^2$

Theorem

The cross product of two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^3$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$: $(\mathbf{u} \times \mathbf{v}) \perp \mathbf{u}, \qquad (\mathbf{u} \times \mathbf{v}) \perp \mathbf{v}$

Theorem

Two non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ are **parallel** iff $\mathbf{u} \times \mathbf{v} = \mathbf{0}$.

Theorem

Let $\mathbf{a}$ and $\mathbf{b}$ be vectors in $\mathbb{R}^3$. Then the parallelogram spanned by these vectors has area equal to the determinant of their cross product:

$$\text{Area} = ||\mathbf{a} \times \mathbf{b}|| = ||\mathbf{a}||||\mathbf{b}|| \sin(\theta)$$

Theorem

The volume of a parallelepiped spanned by vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ in $\mathbb{R}^3$ is given by (the absolute value of) the vectors' **triple product**:

$$\text{Volume} = |(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}| = ||\mathbf{a} \times \mathbf{b}||||\mathbf{c}|| \cos(\theta)|$$

1.3 Lines and Planes

Theorem

In $\mathbb{R}^2$, $n_1x + n_2y = c$ describes a line orthogonal to $\begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$.

If $n_2 \neq 0$, there exists a unique representation of the line:

$$y = ax + b \text{ where } a = -\frac{n_1}{n_2} \text{ and } b = \frac{c}{n_2}$$

In $\mathbb{R}^3$, $n_1x + n_2y + n_3z = c$ describes a plane orthogonal to $\begin{bmatrix} n_1 & n_2 & n_3 \end{bmatrix}^T$.

In $\mathbb{R}^k$, $n_1x + n_2x_2 + \cdots + n_kx_k = c$ describes a hyperplane orthogonal to $\begin{bmatrix} n_1 & n_2 & \dots & n_k \end{bmatrix}^T$.

The Balancing Formula

object dimension + number of general form equations
= dimension of the space

Theorem

The distance to the centre of a circle is equal to its norm:

$$r = \left|\left| \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} a \\ b \end{pmatrix} \right|\right| = \sqrt{(x-a)^2 + (y-b)^2}$$

This gives the equation of a **circle**:

$$(x-a)^2 + (y-b)^2 = r^2$$

Theorem

Similarly in $\mathbb{R}^3$, the distance to the centre of a sphere is:

$$r = \left|\left| \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right|\right| = \sqrt{(x-a)^2 + (y-b)^2 + (z-c)^2}$$

This gives the equation of a **sphere**:

$$(x-a)^2 + (y-b)^2 + (z-c)^2 = r^2$$

	Normal Form	General Form	Vector Form	Parametric Form
Lines in $\mathbb{R}^2$	$\mathbf{n} \cdot (\mathbf{x} - \mathbf{x}) = \mathbf{0} \iff \mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$	$n_1x + n_2y = c$	$\mathbf{x} = \mathbf{p} + t\mathbf{v}$	$\begin{cases} x = p_1 + tv1 \\ y = p_2 + tv2 \end{cases}$
Lines in $\mathbb{R}^3$	$\begin{cases} \mathbf{n}_1 \cdot \mathbf{x} = \mathbf{n}_1 \cdot \mathbf{p}_1 \\ \mathbf{n}_2 \cdot \mathbf{x} = \mathbf{n}_2 \cdot \mathbf{p}_2 \end{cases}$	$\begin{cases} n_{11}x + n_{12}y + n_{13}z = c_1 \\ n_{21}x + n_{22}y + n_{23}z = c_2 \end{cases}$	$\mathbf{x} = \mathbf{p} + t\mathbf{v}$	$\begin{cases} x = p_1 + tv1 \\ y = p_2 + tv2 \\ z = p_3 + tv3 \end{cases}$
Planes in $\mathbb{R}^3$	$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = \mathbf{0} \iff \mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$	$n_1x + n_2y + n_3z = c$	$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$	$\begin{cases} x = p_1 + su_1 + tv_1 \\ y = p_2 + su_2 + tv_2 \\ z = p_3 + su_3 + tv_3 \end{cases}$

$\mathbf{n} \neq \mathbf{0}$ is a normal vector. $\mathbf{p}$ is a given point on the line / in the plane. $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ are direction vectors of the line / plane.

Theorem

The distance from a point B to a line l with a given point A (/ equation $ax + by + \cdots = d$) is given by

$$d(B, l) = ||\vec{AB} - \text{Proj}_{\mathbf{d}}(\vec{AB})|| = \frac{|ax_0 + by_0 + \cdots - d|}{\sqrt{a^2 + b^2 + \dots}}$$

Theorem

The distance from a point B to a plane $\mathcal{P}$ with a given point A and normal vector $\mathbf{n}$ is given by

$$d(B, \mathcal{P}) = ||\text{Proj}_{\mathbf{n}}(\vec{AB})|| = \frac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2 + b^2 + c^2}}$$

2 Systems of Linear Equations

A system of linear equations can be solved by converting it to an augmented matrix $A\mathbf{x} = \mathbf{b}$, reducing it to echelon form, and finally to reduced echelon form:

System of equations:

$$\begin{aligned} x_1 + 2x_2 &= 5 \\ x_1 &= 3 \\ x_1 + 2x_2 &= 5 \end{aligned}$$

Vector equation:

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 5 \end{bmatrix}$$

Matrix-vector equation:

$$\begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 5 \end{bmatrix}$$

Augmented matrix:

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

Echelon form:

$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & -2 & -2 \\ 0 & 0 & 0 \end{array}\right]$$

RREF:

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

Definition

If an augmented matrix in echelon form has a row of the form $\left[\begin{array}{ccc|c} 0 & \dots & 0 & c \end{array}\right], c \neq 0$, then the corresponding linear system is **inconsistent**, i.e. it has **no solution**.

Else, the linear system is **consistent**, i.e. it has ***at least* one solution**.

A solution of a linear system is a vector that is simultaneously a solution of each equation in the system. The **solution set** of is the set of all solutions of a system.

Example: Is $A\mathbf{x} = \mathbf{0}$ consistent? Yes, $\mathbf{x} = \mathbf{0}$ is always a solution, regardless of A. Therefore, $A\mathbf{x} = \mathbf{0}$ has either one or infinitely many solutions.

Theorem

If an augmented matrix in echelon form has a pivot position in every column, then the corresponding linear system has exactly one solution.

If an augmented matrix in echelon form does not have a pivot position in every column, then the corresponding linear system has either zero solutions (if it is consistent), where the variables that correspond to the pivot-less columns can be chosen as free variables, or infinitely many solutions (if it is inconsistent).

Example: Suppose the augmented matrix corresponding to a consistent system is in echelon form. Then the variables corresponding to a pivot-less column can be chosen as free variables, in **parametric vector form**:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix} + s \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, s \in \mathbb{R}$$

Example: This augmented matrix in echelon form does not have a pivot position in every row, i.e. it is inconsistent for some β; but a pivot position in every column, so it has **either** 0 - if $\beta_4 \neq 0$ - **or infinitely many** - if $\beta_4 = 0$ - **solutions**:

$$\left[\begin{array}{ccc|c} 1 & 2 & 4 & \beta_1 \\ 0 & 2 & 2 & \beta_2 \\ 0 & 0 & 4 & \beta_3 \\ 0 & 0 & 0 & \beta_4 \end{array}\right]$$

Definition

A **linear equation** is an equation of the form

$$a_1x_1 + \dots + a_nx_n = b$$

with coefficients a_i, variables x_i and scalar b.

Theorem

Let $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ be an $m \times n$ matrix.
The following statements are equivalent:

1. A has a pivot position in every row.
2. $A\mathbf{x} = \mathbf{b}$ is consistent for all $b \in \mathbb{R}^m$.
3. $\mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_n) = \mathbb{R}^m$

Conversely, the following statements are equivalent:

1. Not all rows of A have a pivot position.
2. $A\mathbf{x} = \mathbf{b}$ is inconsistent for some, not necessarily all, $b \in \mathbb{R}^m$.
3. $\mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_n) \neq \mathbb{R}^m$

Theorem

Let $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ be an $m \times n$ matrix.
The following statements are equivalent:

1. A has a pivot position in every column.
2. $A\mathbf{x} = \mathbf{b}$ has *at most* one solution, i.e. 0 or 1.
3. $A\mathbf{x} = \mathbf{0}$ has only the *trivial solution* $\mathbf{x} = \mathbf{0}$.
4. $\mathbf{a}_1, \dots, \mathbf{a}_n$ are *linearly independent.*

Conversely, the following statements are equivalent:

1. Not all columns of A have a pivot position.
2. $A\mathbf{x} = \mathbf{b}$ has either 0 or infinitely many solutions.
3. $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions.
4. $\mathbf{a}_1, \dots, \mathbf{a}_n$ are *linearly dependent.*

Corollary

An **over-determined linear system** has more equations than variables, and may have any number of solutions.

Corollary

An **under-determined linear system** has fewer equations than variables, and 0 or infinitely many solutions.

Since each row has at most one pivot position and each column has at most one pivot position, there must be pivot-less columns. This means that if the system is consistent, there must be free variables. One solution is not possible.

2.1 Geometric Interpretation of Linear Systems

Systems of linear equations in two variables correspond to lines in $\mathbb{R}^2$. It is possible that they intersect in a point, are parallel, or do neither. Non-parallel lines that do not intersect are called skew lines.

Systems of linear equations in three variables correspond to planes in $\mathbb{R}^3$. It is possible that they intersect in a point, in a line, or neither (picture a water wheel).

Intersection in $\mathbf{R}^n$:	**Number of solutions:**
A line.	∞
A point.	1
None.	0

Definition

A system of linear equations is called **homogeneous** if the constant term in each equation is zero, i.e. $A\mathbf{x} = \mathbf{0}$.

Since it cannot have no solution, it will either have a unique solution or infinitely many solutions.

Definition

An **inhomogeneous system** $A\mathbf{x} = \mathbf{b}$ is one where the origin does not lie in the (hyper-)planes represented by the system - they are shifted by $\mathbf{b}$.

Theorem

If a homogeneous system $A\mathbf{x} = \mathbf{0}$ has a solution set W, then $A\mathbf{x} = \mathbf{b}$ either is inconsistent or has solution set $\vec{p} + W$ where $\mathbf{p}$ is such that $A\mathbf{x} = \mathbf{p}$, i.e. $\mathbf{p}$ is a particular solution.

$$\vec{p} + W = \{\vec{p} + \vec{w} \quad \forall \quad \vec{w} \in W\}$$

That is, the solution set of a *consistent* inhomogeneous linear system is parallel to the solution set of its corresponding homogeneous linear system.

Example: $A\mathbf{x} = \mathbf{b}$ has one solution, i.e. $|\{\mathbf{p} + W\}| = 1$. Then the size of the solution set of $A\mathbf{x} = \mathbf{0}$ is also 1.

2.2 Linear Independence

Definition

Given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$, a vector $\mathbf{w}$ is a **linear combination** of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ if, for some non-zero scalars $c_1, \ldots, c_k$:

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$

Definition

The set of all possible linear combinations of these vectors is called their **span**.

That is, if vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ span $\mathbb{R}^m$, every vector in $\mathbb{R}^m$ can be written as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

Example: $\text{Span}(\{\mathbf{u}, \ldots, \mathbf{v}\})$ is the plane through $\mathbf{u}, \ldots, \mathbf{v}$.

A set of vectors is **linearly dependent** if at least one of the vectors can be written as a linear combination of the others.

Definition

The **dependence relation**: A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly dependent if there are coefficients $c_1, \ldots, c_k \in \mathbb{R}$ such that

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

Definition

A homogeneous system always has a solution where all the variables are 0. This is the **trivial solution**.

A **non-trivial solution** is a solution where at least one variable is not 0.

Theorem

The following two statements are equivalent:
1. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
2. The set of vectors made up of the columns of A, $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly independent.

Theorem

The column vectors of a $m \times n$-matrix A are linearly dependent if and only if the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.

Theorem

The columns of a matrix A are linearly independent if and only if $\det(A) \neq 0$.

Elementary row operations have a specific effect on the determinant of the matrix:

- Swapping rows: This operation flips the sign of the determinant.
- Multiplying a row by a non-zero scalar: This operation multiplies the determinant by the same scalar.
- Adding a multiple of one row to another: This operation doesn't change the determinant itself.

If $\det(A) = 0$, then there exists a non-trivial solution to the system of linear equations represented by the rows of A.

Theorem

The row vectors of a $m \times n$-matrix A are linearly dependent if and only if $\text{rank}(A) < m$.

Thus, the rows of a matrix will be linearly dependent if elementary row operations can be used to crate a zero row.

Theorem

Any set of m vectors in $\mathbb{R}^n$ is linearly dependent if $m > n$, i.e. if there are "too many" vectors to be independent.

Theorem

Any set of vectors containing the zero vector is linearly dependent.

3 Linear Subspaces

Definition

A **linear subspace** of $\mathbb{R}^n$ is a set H such that:
1. $\mathbf{0} \in H \iff$ solution set H of $A\mathbf{x} = \mathbf{0}$ is never empty.
2. If $\mathbf{u} \in H$ and $\mathbf{v} \in H$, then $\mathbf{u} + \mathbf{v} \in H$.
3. If $\mathbf{v} \in H$, and $c \in \mathbb{R}$, then $c\mathbf{v} \in H$.

Corollary

Any linear combination of vectors in H is also in H.

Definition

That is, linear subspaces are closed under
- vector addition
- scalar multiplication
- linear combinations.

That is, you cannot get out of a linear subspace by means of any of these operations. **Example**: If vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are in a subspace and $c_1, \ldots, c_k$ are scalars, then the linear combination $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$ is in the subspace.

Examples: What is a subspace in $\mathbb{R}^n$?

- The origin $\mathbf{0}$ by itself is the smallest possible linear subspace.
- Every line through the origin is a subspace.
- Every plane through the origin is a subspace - it is $\mathbb{R}^2$. (And every higher-dimensional hyperplane in $\mathbb{R}^n$.)
- Let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$. Then $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ is a subspace of $\mathbb{R}^n$.
- Any line, plane or hyperplane that does not pass through the origin is not a linear subspace.
- The first quadrant in $\mathbb{R}^2$, is not a linear subspace because $-\mathbf{v}$ would not be in the space, i.e. property 3 does not hold.

Two subspaces of a vector space may never be disjoint because by definition they must contain the zero vector.

Definition

Let A be an $m \times n$-matrix. The **null space** of A is the set of all $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$:

$$\mathrm{Nul}(A) = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \mathbf{0}\}$$

That is, $\mathrm{Nul}(A)$ is the solution set of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Example: Is $\mathbf{p} \in \mathrm{Nul}(A)$? If $A\mathbf{p} = \mathbf{0}$, yes; else, no.

Theorem

Let A be an $m \times n$-matrix and $\mathbf{b}$ a vector in $\mathbb{R}^m$. The linear equation $A\mathbf{x} = \mathbf{b}$ is either inconsistent, or the solution set equals $\mathbf{p} + \mathrm{Nul}(A)$, where $\mathbf{p}$ is one particular solution of the equation.

Theorem

Let A be an $m \times n$-matrix. Then $\mathrm{Nul}(A)$ is a linear subspace of $\mathbb{R}^n$:
1. $A\mathbf{0} = \mathbf{0} \quad \Rightarrow \quad \mathbf{0} \in \mathrm{Nul}(A)$
2. $A(\mathbf{x} + \mathbf{y}) = \mathbf{0} \quad \Rightarrow \quad A\mathbf{x} + A\mathbf{y} = \mathbf{0} \quad \Rightarrow \quad \mathbf{x}, \mathbf{y} \in \mathrm{Nul}(A) \quad \Rightarrow \quad A\mathbf{x} = \mathbf{0}, A\mathbf{y} = \mathbf{0} \quad \Rightarrow \quad \mathbf{0} + \mathbf{0} = \mathbf{0}$
3. $A(c\mathbf{x}) = c(A\mathbf{x}) = c(\mathbf{0}) = \mathbf{0}$

Definition

Let A be an $m \times n$ matrix. The **row space** of A is the subspace $\mathrm{Row}(A)$ of $\mathbb{R}^n$ spanned by the rows of A.

Theorem

Let B be any matrix that is row-equivalent to a matrix A. Then $\mathrm{Row}(B) = \mathrm{Row}(A)$.

Definition

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

That is, the column space is the set of all possible outcomes of the multiplication $A\mathbf{x}$ - the set of all linear combinations of the columns of A:

$$\mathrm{Col}(A) = \{x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \text{ for } x_1, \ldots, x_n \in \mathbb{R}\}$$

Corollary

It follows that $\mathbf{b} \in \mathrm{Col}(A)$ iff $A\mathbf{x} = \mathbf{b}$ is consistent:

$$\mathrm{Col}(A) = \{\mathbf{b} \in \mathbb{R}^m \text{ for which } A\mathbf{x} = \mathbf{b} \text{ is consistent}\}$$

Example: Null Space

$$\left[\begin{array}{cccc|c} 1 & 2 & -5 & -1 & 0 \\ 2 & -1 & 0 & 3 & 0 \end{array}\right] \quad \rightarrow \quad \mathbf{x} = s\begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R} \quad \rightarrow \quad \mathrm{Nul}(A) = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Theorem

The span of any set of vectors is a linear subspace. That is, given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^m$, $\mathrm{Span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$ is a linear subspace of $\mathbb{R}^m$:

1. $0\mathbf{v}_1 + \cdots + 0\mathbf{v}_k = \mathbf{0} \Rightarrow \mathbf{0} \in \mathrm{Span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$
2. $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \in \mathrm{Span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}), \mathbf{y} = d_1\mathbf{v}_1 + \cdots + d_k\mathbf{v}_k \in \mathrm{Span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}) \Rightarrow \mathbf{x} + \mathbf{y} = (c_1 + d_1)\mathbf{v}_1 + \cdots + (c_k + d_k)\mathbf{v}_k \Rightarrow \mathbf{x} + \mathbf{y} \in \mathrm{Span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$
3. $c\mathbf{x} = (c \cdot c_1)\mathbf{v}_1 + \cdots + (c\cdot)\mathbf{v}_k \Rightarrow c\mathbf{x} \in \mathrm{Span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$

3.1 Basis

In a linear subspace, you can add vectors and multiply them by scalars, just as in $\mathbb{R}^n$. This connection between a subspace and $\mathbb{R}^n$ can be made precise using the definition of a basis.

Definition

A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \in \mathbb{R}^n$ is a **basis** for subspace $H \in \mathbb{R}^n$ if:

- the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent, and
- $H = \text{Span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$.

- A set of vectors is linearly independent if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$, i.e. pivot in every col.
- A set of vectors spans a subspace if $A\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b} \in \mathbb{R}^n$, i.e. if every row has a pivot position.

Theorem

Let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ be a basis for subspace $H \in \mathbb{R}^n$. Then, for each vector $\mathbf{x} \in H$ there exists a **unique** set of scalars $c_1, \ldots, c_k$ such that $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_k\mathbf{b}_k$.

That is, the B-coordinates of $\mathbf{x}$ are:

$$[\mathbf{x}]_B = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

Definition

The k^{th} standard unit vector $\mathbf{e}_k \in \mathbb{R}^n$ is the vector with component $k = 1$ and the other components equal to zero.

Definition

The set $\epsilon = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of the standard unit vectors in $\mathbb{R}^n$ is a basis for $\mathbb{R}^n$ - the **standard basis**:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$$

The pivot columns in $RREF(A)$ are the standard unit vectors. They are linearly independent.

Theorem

Let A be an $m \times n$-matrix. The pivot columns of A form a basis for $\text{Col}(A)$.

Theorem

Row operations preserve dependence relations between the columns of a matrix.

Theorem

Let A be an $m \times n$-matrix. Suppose that the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has k free variables. Let $\mathbf{b}_i$ be the solution for which the i^{th} free variable is 1. Then $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ is a **basis for the null space** $\text{Nul}(A)$.

That is, the solution vectors in parametric form span the null space.

This theorem can be used to find a basis for any linear subspace given by homogeneous linear equations.

3.2 Dimension

Dimension is the maximum number of independent directions.

That is, the dimension is equal to the number of vectors in a basis. But bases are not unique. However, all bases of a given subspace contain the same number of vectors:

Theorem

All bases for a linear subspace of $\mathbb{R}^n$ contain the same number of vectors.

Therefore, the idea of dimension works:

Definition

Let H be a linear subspace of $\mathbb{R}^n$. Let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ be a basis for H. The **dimension** of H is the number of vectors in B:

$$\dim(H) = k$$

Theorem

Let H be a linear subspace of $\mathbb{R}^n$ with $\dim(H) = k$. Then the following statements hold:

1. There are at most k distinct independent vectors in H.
2. At least k vectors are needed to span H.
3. Vectors in H can be described with k coordinates.

The Basis Theorem

Let H be a linear subspace of $\mathbb{R}^n$ with $\dim(H) = k$. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a set of vectors in H.

If set S is linearly independent, then S spans H.

Hence it is a basis.

NB: The vectors *must* be in H for the theorem to hold!

3.3 Rank

Definition

The **column space** of a matrix A is made up of the columns that have a pivot position.

Theorem

The **rank** of a matrix A is the dimension of the column space of A:

$$\text{rank}(A) = \dim(\text{Col}(A) = \dim(\text{Row}(A))$$

Definition

The **null space** of a matrix A is made up of the columns that do not have a pivot position, i.e. the number of free variables of equation $A\mathbf{x} = \mathbf{0}$.

Theorem

The **nullity** of a matrix A is the dimension of the null space of A:

$$\text{nullity}(A) = \dim(\text{Nul}(A))$$

Theorem

For a consistent system of linear equations with coefficient matrix A:
- $\text{rank}(A)$ is the *effective* number of equations.
- $\text{nullity}(A)$ is the number of free variables.

The number of free variables (/columns) equals the total number of variables (/columns) minus the effective number of equations (/rows):

Rank Theorem

Let A be a matrix with n columns. Then

$$\text{rank}(A) + \text{nullity}(A) = n$$

The effective set of equations for any linear system can be found by putting the system in echelon form.

Theorem

Let A be an $m \times n$-matrix with k pivot positions. Then

$$\text{rank}(A) = \dim(\text{Col}(A)) = k \quad (k \text{ pivot positions})$$
$$\text{nullity}(A) = \dim(\text{Nul}(A)) = n - k \quad (n - k \text{ free variables})$$

Theorem

For any matrix A,

$$\text{rank}(A^T) = \text{rank}(A)$$

4 Orthogonality

4.1 Orthogonal Sets

Definition

A set $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is an **orthogonal set** if all distinct vectors in the set are pairwise orthogonal:

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \text{ for } i \neq j$$

Definition

A set of vectors in $\mathbb{R}^n$ is an **orthonormal set** if it is an orthogonal set of unit vectors:

$$||\mathbf{v}_i|| = 1 \text{ for all } i$$

Example: Orthonormal set from an orthogonal set

$$S = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\} \Rightarrow \tilde{S} = \{\frac{1}{||\mathbf{u}||}\mathbf{u}, \frac{1}{||\mathbf{v}||}\mathbf{v}, \frac{1}{||\mathbf{w}||}\mathbf{w}\}$$

Theorem

If $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is an orthogonal set of non-zero vectors in $\mathbb{R}^n$, then S is linearly independent.

Theorem

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then the unique scalars $c_1, c_2, \dots, c_k$ such that $\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k$ are given by

$$c_i = \frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$$

and this representation is unique.

Example: How to use orthogonality to find coordinates:

$$S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \{\begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 5 \\ 6 \end{bmatrix}\}, \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$$

$$\begin{cases} \mathbf{x} \cdot \mathbf{v}_1 = c_1\mathbf{v}_1 \cdot \mathbf{v}_1 + c_2\mathbf{v}_1 \cdot \mathbf{v}_2 + c_3\mathbf{v}_1 \cdot \mathbf{v}_3 \\ \mathbf{x} \cdot \mathbf{v}_2 = c_1\mathbf{v}_2 \cdot \mathbf{v}_1 + c_2\mathbf{v}_2 \cdot \mathbf{v}_2 + c_3\mathbf{v}_2 \cdot \mathbf{v}_3 \\ \mathbf{x} \cdot \mathbf{v}_3 = c_1\mathbf{v}_3 \cdot \mathbf{v}_1 + c_2\mathbf{v}_3 \cdot \mathbf{v}_2 + c_3\mathbf{v}_3 \cdot \mathbf{v}_3 \end{cases}$$

$$\begin{cases} c_1 = \frac{\mathbf{x} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} & (\text{since } \mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot \mathbf{v}_3 = 0) \\ c_2 = \frac{\mathbf{x} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} & (\text{since } \mathbf{v}_2 \cdot \mathbf{v}_1 = \mathbf{v}_2 \cdot \mathbf{v}_3 = 0) \\ c_3 = \frac{\mathbf{x} \cdot \mathbf{v}_3}{\mathbf{v}_3 \cdot \mathbf{v}_3} & (\text{since } \mathbf{v}_3 \cdot \mathbf{v}_1 = \mathbf{v}_3 \cdot \mathbf{v}_2 = 0) \end{cases}$$

Definition

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthogonal set.

Theorem

Let H be a linear subspace of $\mathbb{R}^n$, and $B = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ an orthogonal basis for H. Let $\mathbf{x}$ be a vector in H.

Then the coordinates of $\mathbf{x}$ with respect to B are given by the orthogonal projection of $\mathbf{x}$ onto each $\mathbf{b}_i$:

$$\mathbf{x} = (\frac{\mathbf{x} \cdot \mathbf{b}_1}{\mathbf{b}_1 \cdot \mathbf{b}_1})\mathbf{b}_1 + \dots + (\frac{\mathbf{x} \cdot \mathbf{b}_k}{\mathbf{b}_k \cdot \mathbf{b}_k})\mathbf{b}_k$$

Theorem

If B is an **orthonormal basis**, then the coordinates are given by the dot product:

$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{b}_1)\mathbf{b}_1 + \dots + (\mathbf{x} \cdot \mathbf{b}_k)\mathbf{b}_k$$

Example: A property of the standard basis in $\mathbb{R}^n$ is that each standard basis vector is a unit vector.

Theorem

The columns of an $m \times n$ matrix Q form an orthonormal set if and only if $Q^TQ = I_n$.

Definition

A square matrix whose rows and columns each form an orthonormal set is called an **orthogonal matrix**.

Theorem

A square matrix is orthogonal if and only if $Q^{-1} = Q^T$.

Theorem

If two square matrices Q_1, Q_2 are orthogonal, then so is Q_1Q_2.

Theorem

Let Q be an orthogonal matrix. Then

1. $||Q\mathbf{x}|| = ||\mathbf{x}||$ for every $\mathbf{x} \in \mathbb{R}^n$
2. $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

Theorem

Let Q be an orthogonal matrix. Then
1. Q^{-1} is orthogonal
2. $\det(Q) = \pm 1$
3. If λ is an eigenvalue of Q, then $|\lambda| = 1$.

4.2 Orthogonal Complements

Orthogonally decomposing a vector with respect to a linear subspace H comes down to "splitting" the vector into a component in H and a component in the orthogonal of H.

Definition

A vector $\mathbf{x}$ is orthogonal to a linear subspace H or $\mathbb{R}^n$ if it is orthogonal to all vectors in H. That is: $\mathbf{v} \cdot \mathbf{x} = 0$ for all $\mathbf{x}$ in H.

Definition

The set of all vectors orthogonal to H is the **orthogonal complement** of H, denoted $H^\perp$. That is,
$H^\perp = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \cdot \mathbf{x} = 0 \quad \forall \quad \mathbf{x} \in H\}$

Theorem

Let W be a subspace of $\mathbb{R}^n$. If $W = \text{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_k)$, then $\mathbf{v} \in W^\perp$ if and only if $\mathbf{v} \cdot \mathbf{w}_i = 0$ for all $i = 1, \ldots, k$.

Example: The orthogonal complement of $\mathbb{R}^n$ is the zero space - $\mathbf{0}$ is the only vector orthogonal to every vector in $\mathbb{R}^n$.

Theorem

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

$$(\text{Row}(A))^\perp = \text{Nul}(A), \qquad (\text{Nul}(A))^\perp = \text{Row}(A)$$

$$(\text{Col}(A))^\perp = \text{Nul}(A^T), \qquad (\text{Nul}(A))^\perp = \text{Col}(A^T)$$

Theorem

The $m \times n$ matrix A defines a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$
- whose range is $\text{Col}(A)$, and
- which sends $\text{Nul}(A)$ to 0 in $\mathbb{R}^n$.

Proof: Suppose that $\mathbf{x} \in \text{Nul}(A^T)$, that is, $A^T\mathbf{x} = \mathbf{0}$. This means that $\mathbf{x}$ is orthogonal to each row of A^T and hence to the span of the rows of A^T. But the rows of A^T are the columns of A. This means that $\mathbf{x}$ is orthogonal to $\text{Col}(A)$.

Theorem

Let $H = Span\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a linear subspace of $\mathbb{R}^n$. Then

$$H^\perp = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{v}_1 \cdot \mathbf{x} = 0, \mathbf{v}_2 \cdot \mathbf{x} = 0, \ldots, \mathbf{v}_k \cdot \mathbf{x} = 0\}$$

That is,
1. $H^\perp$ is the solution set, i.e. the **null space**, of the homogeneous system of equations $A\mathbf{x} = \mathbf{0}$, and
2. the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ make up the row space of the matrix A.

Theorem

Let H be a linear subspace of $\mathbb{R}^n$. Then

1. $(H^\perp)^\perp = H$.
2. $dim(H) + dim(H^\perp) = n$
3. $H^\perp$ is a linear subspace of $\mathbb{R}^n$. (This follows from the fact that it is the solution set of a homogeneous system of equations.)
4. The only vector that is both in H and $H^\perp$ is the zero vector: $H \cap H^\perp = \{\mathbf{0}\}$

Example:

Let $W = \text{Span}\left\{\begin{pmatrix} -2, \\ 6 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -6 \\ -2 \\ 10 \end{pmatrix}, \begin{pmatrix} -4 \\ 12 \\ 4 \\ -20 \end{pmatrix}\right\}$. The basis $\left\{\begin{pmatrix} 3 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 3 \\ 1 \end{pmatrix}\right\}$ for $W^\perp$ can be found by solving the homogeneous system

$$\begin{bmatrix} -2 & 6 & -1 & -1 & | & 0 \\ 2 & -6 & -2 & 10 & | & 0 \\ -4 & 12 & 4 & -20 & | & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 0 & 2 & | & 0 \\ 0 & 0 & 1 & -3 & | & 0 \\ 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$$

4.3 Orthogonal Projections

Definition

If $\mathbf{u}, \mathbf{v}$ are vectors in $\mathbb{R}^n$ and $\mathbf{u} \neq \mathbf{0}$, then the projection of $\mathbf{v}$ onto $\mathbf{u}$ is a scalar multiple of $\mathbf{u}$ - the vector

$$\text{Proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\right)\mathbf{u}$$

If $\mathbf{u}$ is a unit vector, then $\text{Proj}_{\mathbf{u}}(\mathbf{v}) = (\mathbf{u} \cdot \mathbf{v})\mathbf{u}$.

Definition

Let W be a subspace of $\mathbb{R}^n$ and let $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ be an *orthogonal* basis for W. For any vector $\mathbf{v} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{v}$ onto W is

$$\begin{aligned}\text{Proj}_W(\mathbf{v}) &= \text{Proj}_W(\mathbf{u}_1) + \cdots + \text{Proj}_W(\mathbf{u}_k)\\ &= \left(\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)\mathbf{u}_1 + \cdots + \left(\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\right)\mathbf{u}_k\end{aligned}$$

The component of $\mathbf{v}$ orthogonal to W is the vector $\text{perp}_W(\mathbf{v}) = \mathbf{v} - \text{Proj}_W(\mathbf{v})$

Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Then there are unique vectors $\mathbf{w}$ in W and $\mathbf{w}^\perp$ in $W^\perp$ such that

$$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$$

Theorem

If W is a subspace of $\mathbf{R}^n$, then $(W^\perp)^\perp = W$.

Theorem

If W is a subspace of $\mathbb{R}^n$, then $\dim(W) + \dim(W^\perp) = n$.

Rank Theorem

If A is an $m \times n$ matrix, then

- $\text{rank}(A) + \text{nullity}(A) = n$
- $\text{rank}(A) + \text{nullity}(A^T) = m$

Often, you want to decompose a vector into two vectors, with one component in a certain subspace and the other component orthogonal to it.

Theorem

Let H be a subspace of $\mathbb{R}^n$ with an *orthogonal* basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$. Then the orthogonal projection of a vector $\mathbf{x}$ onto H is given by:

$$\begin{aligned}\text{proj}_H(\mathbf{x}) &= \text{proj}_{\mathbf{b}_1}(\mathbf{x}) + \cdots + \text{proj}_{\mathbf{b}_k}(\mathbf{x})\\ &= \left(\frac{\mathbf{x} \cdot \mathbf{b}_1}{\mathbf{b}_1 \cdot \mathbf{b}_1}\right)\mathbf{b}_1 + \cdots + \left(\frac{\mathbf{x} \cdot \mathbf{b}_k}{\mathbf{b}_k \cdot \mathbf{b}_k}\right)\mathbf{b}_k\end{aligned}$$

NB: B must be an *orthogonal* basis! If if it is not, use the Gram-Schmidt process.

Orthogonal decompositions can be defined for any vector and any linear subspace:

Theorem

Let $\mathbf{x}$ be a vector in $\mathbb{R}^n$, and H a subspace of $\mathbb{R}^n$. Suppose that we can find y in H and z orthogonal to H such that $\mathbf{x} = y + z$. This sum is called an **orthogonal decomposition** of $\mathbf{x}$ with respect to H.

Theorem

Let H be a linear subspace of $\mathbb{R}^n$, and $\mathbf{x}$ a vector in $\mathbb{R}^n$. Then $\mathbf{x}$ has a **unique** orthogonal decomposition with respect to $H : \mathbf{x} = \mathbf{y} + \mathbf{z}$, where

- $\mathbf{y} = \text{proj}_H \mathbf{x}$
- $\mathbf{z} = \mathbf{x} - \text{proj}_H \mathbf{x}$

Example: Given is the unit vector $\mathbf{y}$ and the space $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. The vector $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto W. What is $\mathbf{u}_1 \cdot (\mathbf{y} - \hat{\mathbf{y}})$?

By definition, the difference between a vector and its orthogonal projection is orthogonal to the space it is being projected onto. This means it is orthogonal to each vector in this space.

Definition

The **distance** between vectors $\mathbf{v}$ and $\mathbf{w}$ is $||\mathbf{v} - \mathbf{w}||$.

Orthogonal projections give us a way to find the shortest distance between a vector and a subspace.

Theorem

Let H be a linear subspace of $\mathbb{R}^n$, $\mathbf{x}$ a vector in $\mathbb{R}^n$, and $\mathbf{p} = \text{proj}_H \mathbf{x}$. Then $\mathbf{p}$ is the closest point in H to $\mathbf{x}$.

That is, for any other point $\mathbf{q}$ in H, $||\mathbf{x} - \mathbf{p}|| \le ||\mathbf{x} - \mathbf{q}||$.

Example: Suppose $A\mathbf{x} = \mathbf{b}$ is inconsistent, i.e. $\mathbf{b} \notin \text{Col}A$, e.g. due to measurement errors. However, replacing $\mathbf{b}$ with $\hat{\mathbf{b}} = \text{proj}_{\text{Col}A}\mathbf{b}$ gives a consistent system $A\mathbf{x} = \hat{\mathbf{b}}$.

4.4 The Gram-Schmidt Orthogonalisation Process

Theorem

Let $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ be a basis for a subspace H of $\mathbb{R}^n$. Then the Gram-Schmidt process produces an orthogonal basis $\{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ for H:

- $\mathbf{b}_1 = \mathbf{a}_1$
- $\mathbf{b}_2 = \mathbf{a}_2 - \text{proj}_{\mathbf{b}_1}\mathbf{a}_2$
- $\mathbf{b}_3 = \mathbf{a}_3 - \text{proj}_{\mathbf{b}_1}\mathbf{a}_3 - \text{proj}_{\mathbf{b}_2}\mathbf{a}_3$
- $\dots$
- $\mathbf{b}_m = \mathbf{a}_m - \text{proj}_{\mathbf{b}_1}\mathbf{a}_m - \cdots - \text{proj}_{\mathbf{b}_{m-1}}\mathbf{a}_m$

Define B as the set of **non-zero** vectors $\mathbf{b}_i$: $B = \{\mathbf{b}_1, \dots, \mathbf{b}_m\} \backslash \mathbf{0}$. Then B is an **orthogonal basis** for H.

4.5 The QR Factorisation

Definition

Let A be an $m \times n$ matrix with *linearly independent* columns. Then A can be factored as $A = QR$, where Q is an $m \times n$ matrix with orthonormal columns and R is an invertible upper triangular matrix.

We can also arrange for the diagonal entries of R to be positive. If any $r_{ii} < 0$, replace $\mathbf{q}_i$ by $-\mathbf{q}_i$ and r_{ii} by $-r_{ii}$.

5 The Least-Squares Method

An inconsistent linear system is a system without solutions. However, such a system does have a so-called least-squares solution.

Definition

The **least-squares solution** $\hat{\mathbf{x}}$ of an inconsistent linear system $A\mathbf{x} = \mathbf{b}$ is the vector $\hat{\mathbf{x}}$ that minimizes the least-squares error.

Definition

The least-squares error is the distance of the vector $\mathbf{b}$ to the column space of A:

$$E = \|A\hat{\mathbf{x}} - \mathbf{b}\| = ||\hat{\mathbf{b}} - \mathbf{b}||$$

Given an inconsistent linear system $A\mathbf{x} = \mathbf{b}$,

1. (If the basis of the column space is not orthogonal, you have to apply the Gram-Schmidt process first.)
2. Find $\hat{\mathbf{b}} \in Col(A)$ such that $\|\mathbf{b} - \hat{\mathbf{b}}\|$ is minimized: $\hat{\mathbf{b}} = \text{proj}_{Col(A)}\mathbf{b}$.
3. Solve the consistent linear system $A\mathbf{x} = \hat{b}$. The solution is the least-squares solution $\hat{\mathbf{x}}$.
4. The least-squares error gives an indication of how far the original equation is from being consistent.

Note: If $A\mathbf{x} = \mathbf{b}$ is consistent, then $\hat{\mathbf{b}} = \mathbf{b}$, the least squares error is 0, and the least squares solution is the ordinary solution.

5.1 The Normal Equation

Calculation of the projection can be a lot of work. Luckily, it can be circumvented by using the normal equation.

Definition

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

Proof: Because $\hat{\mathbf{b}}$ is the projection of $\mathbf{b}$ onto $Col(A)$, it is orthogonal to the error vector $\mathbf{b} - \hat{\mathbf{b}}$. Therefore, $(\mathbf{b} - \hat{\mathbf{b}})$ is orthogonal to $Col(A)$. Given $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, $A\hat{\mathbf{x}} - \mathbf{b}$ is orthogonal to $Col(A)$. That is,

$$\begin{cases} \mathbf{v}_1 \cdot (A\hat{\mathbf{x}} - \mathbf{b}) = 0 \\ \mathbf{v}_2 \cdot (A\hat{\mathbf{x}} - \mathbf{b}) = 0 \end{cases} \Rightarrow \begin{cases} \mathbf{v}_1^T (A\hat{\mathbf{x}} - \mathbf{b}) = 0 \\ \mathbf{v}_2^T (A\hat{\mathbf{x}} - \mathbf{b}) = 0 \end{cases} \Rightarrow \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{bmatrix} (A\hat{\mathbf{x}} - \mathbf{b}) = 0 \Rightarrow A^T(A\hat{\mathbf{x}} - \mathbf{b}) = 0 \Rightarrow A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$$

Theorem

For a system $A\mathbf{x} = \mathbf{b}$, the least-squares solutions of the normal equation $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ are precisely the least-squares solutions of the original system. That is, $A\hat{\mathbf{x}}$ is the vector in $\text{Col}(A)$ closest to $\mathbf{b}$, and $\hat{\mathbf{x}}$ solves both:

- $A\hat{\mathbf{x}} = \hat{\mathbf{b}} = \text{proj}_{Col(A)}\mathbf{b}$, and
- the normal equation $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

5.2 Regression

Definition

$$X\beta = y$$

where X is the **design matrix**, β is the **parameter vector**, and y is the **response or observation vector**.

Definition

This gives the **Normal Equation**:

$$X^T X\beta = X^T y$$

and the **Least-Squares Solution**:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

that minimises the **Least-Squares Error**:

$$E = \|y - X\hat{\beta}\| = \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Definition

Suppose that two quantities x and y are related in the following way: $y = \beta_0 f_0(x) + \cdots + \beta_k f_k(x)$, for some coefficients $\beta_0, \ldots, \beta_k$ and (possibly non-linear) functions $f_0, \ldots, f_k$.

$$\begin{bmatrix} f_0(x_1) & \ldots & f_k(x_1) \\ \vdots & \ddots & \vdots \\ f_0(x_N) & \ldots & f_k(x_N) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

Especially for large N this system will almost certainly be inconsistent. The least-squares line provides the best estimate of the coefficients by minimising the sum of the squares of the vertical distances between the regression line and the data points.

6 Matrix Algebra

Theorem

Let A be an $m \times n$ matrix, $\mathbf{e}_i$ a $1 \times m$ standard unit vector, and $\mathbf{e}_j$ an $n \times 1$ standard unit vector. Then

- $\mathbf{e}_i A$ is the i^{th} row of A,
- $A\mathbf{e}_j$ is the j^{th} column of A.

Thus, multiplication by a standard unit vector can be used to "pick out" a particular column or row.

Properties of Matrix Addition and Scalar Multiplication

Let A, B, C be matrices whose sizes are such that the indicated operations can be performed, and let c, d be scalars.

1. Commutativity: $A + B = B + A$
2. Associativity: $(A + B) + C = A + (B + C)$
3. $A + 0 = A$
4. $A + (-A) = 0$
5. Distributivity: $c(A + B) = cA + cB$
6. Distributivity: $(c + d)A = cA + dA$
7. $c(dA) = (cd)A$
8. $1A = A$

Properties of Matrix Multiplication

Let A, B, C be matrices whose sizes are such that the indicated operations can be performed, and let k be a scalar.

1. Associativity: $A(BC) = (AB)C$
2. Left Distributivity: $A(B + C) = AB + AC$
3. Right Distributivity: $(A + B)C = AC + BC$
4. $k(AB) = (kA)B = A(kB)$
5. Multiplicative Identity: $I_m A = A = AI_n$ if A is $m \times n$
6. In general, no commutativity: $AB \neq BA$.
7. Therefore, $(A + B)^2 = A^2 + 2AB + B^2$ if and only if A and B do commute, i.e. $AB + BA = 2AB$.
8. Moreover, $A^2 = 0$ does *not* imply that $A = 0$.

Properties of Matrix Powers

If A is a square matrix, and k, r, s are non-negative integers, then

1. $A^0 = I_n$
2. $A^2 = AA$
3. $A^k = A \ldots A$
4. $A^r A^s = A^{r+s}$
5. $(A^r)^s = A^{rs}$

Definition

A square matrix A is called **idempotent** if $A^2 = A$; and **nilpotent** if $A^m = 0$ for some $m > 1$.

Properties of Matrix Transpositions

Let A, B be matrices whose sizes are such that the indicated operations can be performed, and let k be a scalar.

1. $(A^T)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(kA)^T = k(A^T)$
4. $(AB)^T = B^T A^T$
5. $(A^r)^T = (A^T)^r$ for all non-negative integers r

Definition

The **trace** of an $n \times n$ matrix is the sum of the n values on its diagonal:

$$\mathrm{Tr} : \mathbb{R}^{n \times n} \to \mathbb{R}, \quad \mathrm{Tr}[A] \equiv \sum_{i=1}^{n} a_{ii}$$

Properties of Traces

1. $\mathrm{Tr}[\alpha A + \beta B] = \alpha \mathrm{Tr}[A] + \beta \mathrm{Tr}[B]$ (linear)
2. $\mathrm{Tr}[AB] = \mathrm{Tr}[BA]$
3. $\mathrm{Tr}[ABC] = \mathrm{Tr}[CAB] = \mathrm{Tr}[BCA]$ (cyclic)
4. $\mathrm{Tr}[A^T] = \mathrm{Tr}[A]$
5. $\mathrm{Tr}[A] = \sum_{i=1}^{n} \lambda_i$, where $\{\lambda_i\}$ are its eigenvalues.

6.1 Matrix Multiplication - TODO: remove this section, merge into previous page

Definition

Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. The product of A and B is the $m \times p$ matrix

$$AB = [A\mathbf{b}_1, A\mathbf{b}_2, \ldots, A\mathbf{b}_p]$$

where $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p$ are the columns of B.

Each column of AB is a linear combination of the columns of A, which all have m elements. So just like A, the matrix AB has m rows. On the other hand, the column-by-column computation of AB shows that AB and B have the same number of columns p.

Definition

Element (i, j) of AB is the dot product of the i-th row of A and the j-th column of B:

$$(AB)_{ij} = \mathbf{a}_i^T \mathbf{b}_j = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj}$$

Theorem

Properties of Matrix Multiplication:

1. Left-Distributive: $A(B + C) = AB + AC$
2. Right-Distributive: $(A + B)C = AC + BC$
3. Associative: $A(BC) = (AB)C$
4. Identity: $I_m A = A = AI_n$
5. Zero: $0_{k\times m}A_{m\times n} = 0_{k\times n}$ and $A_{m\times n}0_{n\times l} = 0_{m\times l}$

Theorem

Suppose A and B are matrices such that AB is defined. Then

$$(AB)^T = B^T A^T$$

Theorem

Given the associative property of matrix multiplication, for any integer $k \geq 1$, the k-th power of a square matrix

$$A^k = AA^{k-1} = A \ldots A$$

6.2 Special Types of Matrices - TODO: refactor

- **Elementary matrices** can be used to perform elementary row operations by matrix multiplication.
- **Triangular matrices** have special properties when it comes to the calculation of determinants and eigenvalues.
- **Diagonal matrices** are used to perform element-wise operations on vectors, since calculation with such matrices is the closest to calculation with scalars.

6.2.1 Elementary Matrices

An elementary matrix E is any matrix that can be obtained by performing an elementary row operation on an identity matrix I_n.

If the same elementary row operation is performed on an $n\times r$-matrix A, the result is the same as the matrix EA. Thus, elementary matrices make it possible to describe row operations in the language of matrix multiplication.

Permutation matrices are orthogonal matrices permuting one or more rows, e.g. $PA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$, or columns, e.g. $AP = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} b & a \\ d & c \end{bmatrix}$. All permutation matrices are elementary matrices, because swapping rows is a type of row operation. However, not all elementary matrices are permutation matrices. The other types of row operations (scaling a row, adding a multiple of one row to another) cannot be achieved with a simple rearrangement.

6.2.2 Triangular Matrices

Definition

Let A be an $n \times n$ matrix. The **main diagonal** of A consists of the entries A_{ii} (with $1 \leq i \leq n$) with equal row and column number.

Definition

A is an **upper triangular matrix** if all entries below the main diagonal are zero, i.e. $A_{ij} = 0$ for $n \geq i > j \geq 1$.

A is a **lower triangular matrix** if all entries above the main diagonal are zero, i.e. $A_{ij} = 0$ for $n \geq j > i \geq 1$.

Triangular matrices are useful because

- it is easy to calculate the *determinant* of a triangular matrix.
- it is easy to find the *eigenvalues* of a triangular matrix.
- triangularity of matrices behaves nicely with respect to matrix multiplication:

Theorem

Let A and B be two upper triangular $n \times n$ matrices. Then the product AB is also an upper triangular matrix.

Let A and B be two lower triangular $n\times n$ matrices. Then the product AB is also a lower triangular matrix.

- many square matrices can be written as the product of a *lower matrix* and an *upper diagonal matrix* - which is faster to solve many non-homogeneous systems $A\mathbf{x} = \mathbf{b}$:

Theorem

Let A be a matrix such that some elementary matrices E_i row-reduce A to an upper-triangular matrix U; that is, $E_3E_2E_1A = U$. Then the inverse of these elementary operations gives a lower-triangular matrix $L = E_1^{-1}E_2^{-1}E_3^{-1}$.

The **LU decomposition** of matrix A is given by:

$$A = LU(= E_1^{-1}E_2^{-1}E_3^{-1}E_3E_2E_1A)$$

6.2.3 Diagonal Matrices

Definition

A square matrix whose non-diagonal entries are all zero is called a **diagonal matrix**: $A_{ij} = 0$ for $i \neq j$(where $1 \leq i, j \leq n$).
A diagonal matrix all of whose diagonal entries are the same is called a **scalar matrix**.
If the scalar on the diagonal is 1, the scalar matrix is called an identity matrix.

Theorem

Let A be an $n \times n$ diagonal matrix and B be an $n \times n$ matrix. Then the following holds:

1. AB is also a diagonal matrix with diagonal elements $(AB)_{ii} = A_{ii}B_{ii}$.
2. Multiplication of diagonal matrices is commutative: $AB = BA$.
3. For any integer $k \geq 0$, A^k is also a diagonal matrix, with diagonal elements $(A^k)_{ii} = A_{ii}^k$.

6.2.4 Orthogonal Matrices

Definition

An **orthogonal matrix** is a square matrix whose columns (and rows) form a set of orthonormal vectors such that

$$A^{-1} = A^T \iff AA^T = I, A^T A = I$$

$$\textbf{Proof}: \quad A^T A = \begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix} \begin{bmatrix} a_1 & a_2 \end{bmatrix} = \begin{bmatrix} a_1^T a_1 & a_1^T a_2 \\ a_2^T a_1 & a_2^T a_2 \end{bmatrix} = I$$

$$\iff a_1^T a_1 = a_2^T a_2 = 1, a_1^T a_2 = a_2^T a_1 = 0$$

Theorem

Let A be an $n \times n$ orthogonal matrix, and $\mathbf{x}$ an $n \times 1$ column vector.

An orthogonal matrix preserves lengths:

$$||A\mathbf{x}||^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T I\mathbf{x} = \mathbf{x}^T I\mathbf{x} = ||\mathbf{x}||^2$$

6.3 Matrix Inversion

Definition

Let A be an $n \times n$ matrix. The **inverse** of A is the matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$.

Definition

An $n \times n$ matrix A is called **invertible** if A has an inverse, and **singular** otherwise.

Theorem

If A is invertible,

1. its RREF of is the identity matrix I_n, i.e. it has a pivot position in every row and column
2. the system $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$
3. the inverse of A is **unique**:

$$C = CI_n = C(AD) = (CA)D = I_n D = D$$

Theorem

Let A and B be $n \times n$ matrices. Then the following holds:

1. If A is invertible, then A^{-1} is invertible and $(A^{-1})^{-1} = A$.

2. If A and B are invertible matrices of the same size, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.
(If A and B are invertible *diagonal* matrices of the same size, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1} = A^{-1}B^{-1}$.)
3. If A and / or B is / are singular, then AB is singular.

4. If A is invertible, then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.
5. If A is singular, then A^T is singular.
6. Let A be an $m \times n$ matrix. Then $\text{Rank}(A^T A) = \text{Rank}(A)$. The $n \times n$ matrix $A^T A$ is invertible if and only if $\text{Rank}(A) = n$.

7. If A is invertible and c is a non-zero scalar, then cA is invertible and $(cA)^{-1} = \frac{1}{c}A^{-1}$.
8. If A is invertible, then A^n is invertible for all non-negative integers n and $(A^n)^{-1} = (A^{-1})^n$.
9. If A is invertible and n is a positive integer, then $A^{-n} = (A^n)^{-1} = (A^{-1})^n$.

Proofs:

1. Assume that A is invertible. Then there is a matrix C such that $A^{-1}C = CA^{-1} = I_n$, namely $C = A$. So we find that A^{-1} is invertible, and $A^{-1} = A$.

2. Assume that A and B are invertible, and define $C = B^{-1}A^{-1}$. Then $(AB)C = ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I_n$. Similarly, $C(AB) = B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I_n$. So we find that AB is invertible, and $(AB)^{-1} = C = B^{-1}A^{-1}$.

3. Assume that A is invertible. Let $C = (A^{-1})^T$. Then $A^T C = A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$. Similarly, $CA^T = I$. So A^T is invertible, and $(A^T)^{-1} = C = (A^{-1})^T$.

4. Since $(A^T)^T = A$, the contrapositive of the previous statement is equivalent to this statement. That is, if A^T is invertible, then A is invertible. And so if A is singular, then A^T is singular.

Theorem

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then A is invertible if and only if $ad - bc \neq 0$ s.t. $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

If $ad - bc = 0$, then A is singular.

Example: The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is singular.

Theorem

Every diagonal matrix is invertible, unless one of its diagonal elements is zero.

Theorem

Every elementary matrix is invertible, and its inverse is an elementary matrix of the same type.

Theorem

Let A be a square matrix. If a sequence of elementary row operations reduces A to I, then the same sequence of elementary row operations transforms I into A^{-1}.

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

1. If A has a pivot position in every row, $A\mathbf{x} = \mathbf{b}$ has *at least one* solution for each $\mathbf{b} \in \mathbb{R}^m$, the columns of A span $\mathbb{R}^m$, and A has rank m.
2. If A has a pivot position in every column, $A\mathbf{x} = \mathbf{b}$ has *at most one* solution for each $\mathbf{b} \in \mathbb{R}^m$, the columns of A are linearly independent, and A has rank n.

The Invertible Matrix Theorem

Let $A_{n\times n}$ be a square matrix. Then the following statements are equivalent:

1. A is invertible.

2. A has a pivot position in every row.
3. Equation $A\mathbf{x} = \mathbf{b}$ has *at least* one solution for each $\mathbf{b} \in \mathbb{R}^n$.
4. The columns of A span $\mathbb{R}^n$.

5. A has a pivot position in every column.
6. $A\mathbf{x} = \mathbf{b}$ has *at most* one solution for each $\mathbf{b} \in \mathbb{R}^n$.
7. The columns of A are linearly independent.
8. A has rank n.

Gauss-Jordan Method for Matrix Inversion

Let A be an $n \times n$ matrix. Perform row reduction on the super-augmented matrix $[A|I_n]$.

- If $A \sim I_n$ by performing row reduction, i.e. $[A|I_n] \sim [I_n|B]$, then A is invertible and $B = A^{-1}$.
- If $A \not\sim I_n$, then A is singular.

6.4 Inverses of Non-Square Matrices

Definition

Let A be an $m \times n$ matrix. A **right inverse** of A is an $n \times m$ matrix C such that $AC = I_m$. A **left inverse** of A is an $n \times m$ matrix D such that $DA = I_n$.

Theorem

Let A be an $m \times n$ matrix, and b be a vector in $\mathbb{R}^m$.

- If A has a right inverse C, then the system $A\mathbf{x} = \mathbf{b}$ has *at least* one solution $\mathbf{x} = C\mathbf{b}$.
- If A has a left inverse D, then the system $A\mathbf{x} = \mathbf{b}$ has *at most* one solution, which if it exists must be $\mathbf{x} = D\mathbf{b}$.

Proof:

Let A be an $m\times n$ matrix and b a vector in $\mathbb{R}^m$. Suppose that A has a right inverse C. Take $\mathbf{x} = C\mathbf{b}$. Then $A\mathbf{x} = AC\mathbf{b} = I_m\mathbf{b} = \mathbf{b}$. You see that $\mathbf{x} = C\mathbf{b}$ is a solution. Hence the system has at least one solution. Now suppose that A has a left inverse D and suppose that there is a vector $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{b}$. Then $DA\mathbf{x} = D\mathbf{b}$. Since $DA = I_n$, it follows that $\mathbf{x} = D\mathbf{b}$. Hence the system has at most one solution, $\mathbf{x} = D\mathbf{b}$.

Theorem

Let A be an $m\times n$ matrix. Then the following statements are equivalent:

1. There exists a matrix C such that $AC = I_m$; that is, a **right inverse**.
2. For each $\mathbf{b} \in \mathbb{R}^m$, the system $A\mathbf{x} = \mathbf{b}$ has *at least* one solution.
3. The columns of A span $\mathbb{R}^m$.
4. A has a pivot position in every row.
5. A has rank m.

Theorem

Let A be an $m\times n$ matrix. Then the following statements are equivalent:

1. There exists a matrix D such that $DA = I_n$; that is, a **right inverse**.
2. For each $\mathbf{b} \in \mathbb{R}^n$, the system $A\mathbf{x} = \mathbf{b}$ has *at most* one solution.
3. The columns of A are linearly independent.
4. A has a pivot position in every column.
5. A has rank n.

Theorem

Existence of an Inverse: Let A be an $m \times n$ matrix.

1. If A has a right inverse C, then A is invertible, and $A^{-1} = C$.
2. If A has a left inverse D, then A is invertible, and $A^{-1} = D$.
3. $C \neq D$

Consequently, if $A_{n\times n}$ and $B_{n\times n}$ are sqare matrices, if the product AB is invertible, then A and B are invertible.

7 Linear Transformations

Linear transformations are functions or mappings that preserve the structure of the vector space. In other words, they preserve the operations of addition and scalar multiplication.

Definition

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** if it satisfies the following properties:

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

For every vector $\mathbf{v}$ in the domain of T, the vector $T(\mathbf{v})$ in the codomain is called the **image** of $\mathbf{v}$ under (the action of) T. The set of all possible images $T(\mathbf{v})$ (as $\mathbf{v}$ varies throughout the domain of T) is called the **range** of T.

Theorem

Any matrix transformation is a linear transformation.

Proof: Both properties hold as a direct consequence of the rules of calculation of matrix-vector multiplication:

- $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$
- $A(c\mathbf{u}) = cA\mathbf{u}$

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T has the following properties:

1. T maps the zero vector to the zero vector: $T(\mathbf{0}) = \mathbf{0}$.
2. T maps each line in $\mathbb{R}^n$ to a line or a point in $\mathbb{R}^m$.
3. T maps *parallel* lines in $\mathbb{R}^n$ either to (possibly coinciding) *parallel* lines in $\mathbb{R}^m$ or to (possibly coinciding) points in $\mathbb{R}^m$.
4. Let W be a linear subspace of $\mathbb{R}^n$ with dimension k. Then $T(W)$ is a linear subspace of $\mathbb{R}^m$ with dimension $\leq k$.

Examples of linear transformations in geometry:

- Rotations around the origin
- Orthogonal projections on a line through the origin (or any other linear subspace)
- Reflections through a line through the origin (or any other linear subspace)

Example: $T : \mathbb{R}^2 \to \mathbb{R}^3$:

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \\ a_{31}x_1 + a_{32}x_2 \end{bmatrix}$$

Example: Rotation of a vector $\mathbf{r} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} r\cos(\phi) \\ r\sin(\phi) \end{bmatrix}$ by an angle θ can be derived using trigonometry and addition formula for cosine and sine:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} r\cos(\phi+\theta) \\ r\sin(\phi+\theta) \end{bmatrix} = \begin{bmatrix} r(\cos\theta\cos\phi - \sin\theta\sin\phi) \\ r(\sin\theta\cos\phi + \cos\theta\sin\phi) \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Thus, the rotation matrix by an angle θ is the orthogonal matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

7.1 The Standard Matrix of a Linear Transformation

Any matrix transformation is a linear transformation. The converse is also true:

Theorem

For any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, there exists a matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

This has far-reaching consequences: any linear transformation, including rotations, projections and reflections, can be evaluated using matrix-vector multiplication. It makes it easy to implement linear transformations on a computer, which is very convenient in, for example, Computer Graphics.

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and $e_1, \dots, e_n$ be the standard basis of $\mathbb{R}^n$. Then the standard matrix of T is the matrix with columns where the transformation T is applied to each standard basis vector:

$$A = \begin{bmatrix} T(e_1) & \dots & T(e_n) \end{bmatrix}$$

Example: Rotations:

Let $T_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation around the origin by angle α. Then $T_\alpha(\mathbf{x}) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}^2$.

$\mathrm{Ker}(T_A) \equiv \mathcal{N}(A), \quad \mathrm{Image}(T_A) \equiv \mathcal{C}(A).$

Theorem

Let L be the line in $\mathbb{R}^2$ spanned by vector $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. Let $\mathrm{Proj}_L : \mathbb{R}^2 \to \mathbb{R}^2$ be the orthogonal projection on L. Then $\mathrm{Proj}_L(\mathbf{x}) = \frac{1}{a^2+b^2}\begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}\mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}^2$. If L makes an angle α with the x_1-axis, then this can be rewritten to:

$$\mathrm{Proj}_L(\mathbf{x}) = \begin{bmatrix} \cos^2\alpha & \cos\alpha\sin\alpha \\ \cos\alpha\sin\alpha & \sin^2\alpha \end{bmatrix}\mathbf{x}$$

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let $B = \{b_1, \dots, b_n\}$ be a basis for $\mathbb{R}^n$. Then there is a unique $n \times n$ matrix M_B such that for each vector $\mathbf{x}$ in $\mathbb{R}^n$ the following holds:

$$[T(\mathbf{x})]_B = M_B[\mathbf{x}]_B$$

The k-th column of M_B is the coordinate vector of $T(b_k)$ with respect to the basis B.
Furthermore, if A is the standard matrix to T, then $A = PM_BP^{-1}$, i.e. A and M_B are similar.

7.2 Composition and Inversion of Linear Transformations, and their Relation with Matrix Algebra

Theorem

Composition of Linear Transformations:
Let $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $T_2 : \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations, with standard matrices A_1 and A_2. Then the composition $T_3 = T_2 \circ T_1 : \mathbb{R}^n \to \mathbb{R}^k$ defined by $T_3(\mathbf{x}) = T_2(T_1(\mathbf{x}))$ is a linear transformation, with standard matrix $A_3 = A_2A_1$.

Theorem

Linear transformations are **associative**: $T_x(\mathbf{x}) = T_2(T_1(\mathbf{x})) = A_2(A_1\mathbf{x}) = (A_2A_1)\mathbf{x}$

That is, you can either first apply A_1 and then A_2, or directly apply $A_3 = A_2A_1$.

Example: Rotations

$R_\alpha(\mathbf{x}) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\mathbf{x}$ is the standard matrix of the rotation $T_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ around the origin by angle α.

$R_{2\alpha} = \begin{bmatrix} \cos 2\alpha & -\sin 2\alpha \\ \sin 2\alpha & \cos 2\alpha \end{bmatrix}\mathbf{x}$ is the standard matrix of the rotation $T_{2\alpha} : \mathbb{R}^2 \to \mathbb{R}^2$ around the origin by angle 2α.

$$R_\alpha(\mathbf{x})R_\alpha(\mathbf{x}) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\mathbf{x} = \begin{bmatrix} \cos^2\alpha - \sin^2\alpha & -2\sin\alpha\cos\alpha \\ 2\sin\alpha\cos\alpha & \cos^2\alpha - \sin^2\alpha \end{bmatrix}\mathbf{x} = \begin{bmatrix} \cos 2\alpha & -\sin 2\alpha \\ \sin 2\alpha & \cos 2\alpha \end{bmatrix}\mathbf{x} = R_{2\alpha}\mathbf{x}.$$

Definition

Inversion of Linear Transformations:
Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. A function $S : \mathbb{R}^n \to \mathbb{R}^n$ is an **inverse** of T if

1. $S(T(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$,
2. $T(S(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Example: Rotations

$R_\alpha(\mathbf{x}) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\mathbf{x}$ is the standard matrix of the rotation $T_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ around the origin by angle α.

$R_{-\alpha}(\mathbf{x}) = \begin{bmatrix} \cos -\alpha & -\sin -\alpha \\ \sin -\alpha & \cos -\alpha \end{bmatrix}\mathbf{x} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}\mathbf{x}$ is the standard matrix of the rotation $T_{-\alpha} : \mathbb{R}^2 \to \mathbb{R}^2$ around the origin by angle $-\alpha$.

$$R_\alpha^{-1}(\mathbf{x}) = \frac{1}{\cos^2\alpha + \sin^2\alpha}\begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}\mathbf{x} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}\mathbf{x} = R_{-\alpha}(\mathbf{x}).$$

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation, with inverse $S : \mathbb{R}^n \to \mathbb{R}^n$. Then the following statements hold:
1. S is unique,
2. S is a linear transformation,
3. If T has a standard matrix A, then S has a standard matrix A^{-1}.

7.3 Injectivity and Surjectivity of Linear Transformations

The properties of a linear transformation are related to the properties of its standard matrix.

Definition

Let X and Y be sets and $f : X \to Y$ be a function.
f is **injective** if for each $y \in Y$ there is *at most* one $x \in X$ such that $f(x) = y$.

Example: e^x is injective, x^2 is not.

Definition

Let X and Y be sets and $f : X \to Y$ be a function.
f is **surjective** if for each $y \in Y$ there is *at least* one $x \in X$ such that $f(x) = y$.

Example: $x^3 - 3x$ is surjective, e^x and x^2 are not because the range of e^x is $(0, \infty)$ and the range of x^2 is $[0, \infty)$.

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with standard matrix A.
- T is **injective** if and only if for each $\mathbf{y} \in \mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{y}$ has *at most* one solution, i.e. a pivot position in every column.
- T is **surjective** if and only if for each $\mathbf{y} \in \mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{y}$ has *at least* one solution, i.e. a pivot position in every row.

Theorem

Characterisation of Injectivity:
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with standard matrix A.
Then the following statements are equivalent:
1. T is injective.
2. For any vector $\mathbf{y} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{y}$ has *at most* one solution.
3. The columns of A are linearly independent.
4. A has a left inverse, i.e. there exists a matrix B such that $BA = I_n$.
5. A has a pivot position in every column.
6. The rank of A is n.

Theorem

Characterisation of Surjectivity:
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with standard matrix A.
Then the following statements are equivalent:
1. T is surjective.
2. For any vector $\mathbf{y} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{y}$ has *at least* one solution.
3. The columns of A span $\mathbb{R}^m$.
4. A has a right inverse, i.e. there exists a matrix B such that $AB = I_m$.
5. A has a pivot position in every row.
6. The rank of A is m.

7.4 Orthogonal Transformations

Definition

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is **orthogonal** if for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ the following holds:

$$T(\mathbf{u}) \cdot T(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$$

Theorem

An orthogonal linear transformation preserves the norms of all vectors and angles between them.

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with standard matrix A.
Then the following statements are equivalent:
1. T preserves the dot product, i.e. $T(\mathbf{u}) \cdot T(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.
2. T preservces the norms of vectors, i.e. $\|T(\mathbf{u})\| = \|\mathbf{u}\|$ for all $\mathbf{u} \in \mathbb{R}^n$, and angles between them, i.e. $\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{T(\mathbf{u})\cdot T(\mathbf{v})}{\|T(\mathbf{u})\|\|T(\mathbf{v})\|}$.
3. The columns of A are mutually orthogonal and have norm 1.
4. The standard matrix A is an orthogonal matrix, i.e. $A^T A = I_n$; that is, $A^T = A^{-1}$.

Example: The orthogonal projection $\mathrm{Proj}_V(\mathbf{x})$ on a linear subspace V of $\mathbb{R}^n$ is not an orthogonal linear transformation:

If $V \neq \mathbb{R}^n$, we can take a vector $\mathbf{x} \neq \mathbf{0}$ that is orthogonal to V. Then $\mathrm{Proj}_V(\mathbf{x}) = \mathbf{0}$, so Proj_V does not preserve the norm of $\mathbf{x}$.

However, matrices with orthonormal columns play a role in the context of projections:

Theorem

Let $V \neq \mathbf{0}$ be a linear subspace of $\mathbb{R}^n$ and $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be an *orthonormal basis* for V. Let U be the $n \times k$ matrix with columns $\mathbf{u}_1, \ldots, \mathbf{u}_k$.

Then UU^T is the standard matrix of the orthogonal projection on V:

$$\text{Proj}_V(\mathbf{x}) = UU^T\mathbf{x}$$

1. It is important that the basis is orthonormal. If it is not, apply the Gram-Schmidt process.
2. It does not matter which orthonormal basis you use. The product UU^T is the same for all orthonormal bases of V; the standard matrix is unique.
3. Note that $(UU^T)^T = (U^T)^T U^T = UU^T$. *The standard matrix of an orthogonal projection is symmetric.*

Reflections and rotations are orthogonal linear transformations.

Projections, and translations by a non-zero vector are not.

Definition

Let V be a linear subspace of $\mathbb{R}^n$. Let $\mathbf{x}$ be a vector in $\mathbb{R}^n$ with orthogonal decomposition $\mathbf{x} = \mathbf{x}_{||} + \mathbf{x}_{\perp}$, where $\mathbf{x}_{||} \in V$ and $\mathbf{x}_{\perp} \in V^{\perp}$. Then the **reflection** of $\mathbf{x}$ through V is

$$\text{Refl}_V(\mathbf{x}) = \mathbf{x}_{||} - \mathbf{x}_{\perp}$$

Theorem

Let V be a linear subspace of $\mathbb{R}^n$. Then the following statements hold:

1. The reflection Refl_V is an orthogonal linear transformation.

2. $\text{Refl}_V(\mathbf{x}) = 2\text{Proj}_V(\mathbf{x}) - \mathbf{x}$ for all vectors $\mathbf{x} \in \mathbb{R}^n$.

3. Suppose $V \notin \mathbf{0}$. Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be an orthonormal basis for V and let U be the $n \times k$ matrix with columns $\mathbf{u}_1, \ldots, \mathbf{u}_k$. Then the standard matrix of the reflection Refl_V is $2UU^T - I_n$. If $V = \mathbf{0}$, the standard matrix of the reflection Refl_V is $-I_n$.

8 Determinants

Theorem

Let A be an $n \times n$ matrix. Then the following statements are equivalent:

1. $\det(A) \neq 0$
2. The row vectors in A are linearly independent.
3. A is invertible.

Definition

The **determinant** of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is

$$\det(A) = |A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Definition

The **determinant** of a 3×3 matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ is

$$\det(A) = |A| = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13})$$
$$= \sum_{j=1}^{3} (-1)^{1+j} a_{ij} \det(A_{1j})$$

This leads to the definition of cofactors and the checkerboard-pattern of the Laplace Expansion Theorem - which is particularly useful when the matrix is upper or lower triangular.

Definition

The **cofactor** C_{ij} of element a_{ij} of an $n \times n$ matrix A is

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

where A_{ij} is the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and j-th column of A.

Laplace Expansion Theorem

For an $n \times n$ matrix $A = [a_{ij}]$, where $n \geq 2$:

- across row i:

$$\det(A) = \sum_{j=1}^{n} a_{ij}C_{ij} \qquad \left(= \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\det(A_{ij})\right)$$

- down column j:

$$\det(A) = \sum_{i=1}^{n} a_{ij}C_{ij} \qquad \left(= \sum_{i=1}^{n} a_{ij}(-1)^{i+j}\det(A_{ij})\right)$$

Theorem

The determinant of a triangular matrix is the product of the entries on its main diagonal.
Specifically, if $A = [a_{ij}]$ is an $n \times n$ **triangular matrix**, then

$$\det(A) = a_{11}a_{22}\dots a_{nn}$$

8.1 Matrix Algebra

Theorem

Let A, B, C be $n \times n$ matrices. Then:

1. If A has a zero row (or column), then $\det(A) = 0$.
2. If A has two identical rows (or columns), then $\det(A) = 0$.
3. $\det(A^T) = \det(A)$.
4. If A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$.
5. If B is obtained by interchanging two rows (or columns) of A, then $\det(B) = -\det(A)$.
6. If B is obtained by adding a multiple of one row (or column) of A to another row (or column) of A, then $\det(B) = \det(A)$.
7. If B is obtained by multiplying a row (or column) of A by c, then $\det(B) = c\det(A)$.
8. $\det(AB) = \det(A)\det(B)$.
9. If A, B, C are identical except that the i^{th} row (or column) of C is the sum of the i^{th} rows (or columns) of A and B, then $\det(C) = \det(A) + \det(B)$.

Theorem

For any scalar c, $\det(cA) = c^n\det(A)$
(one power of c for every row of the matrix).

Theorem

If A is an $n \times n$ matrix and E is an $n \times n$ elementary matrix, then $\det(EA) = \det(E)\det(A)$.

Theorem

Let E be an $n \times n$ elementary matrix.

1. If E results from interchanging two rows of I_n, then $\det(E) = -1$.
2. If E results from multiplying one row of I_n, by k, then $\det(E) = k$.
3. If E results from adding a multiple of one row of I_n to another row, then $\det(E) = 1$.

Theorem

Let A be an $n \times n$ matrix, and R its RREF. Then

$$|R| = |E_q||E_{q-1}|\dots|E_1||A|, \quad |E| \neq 0$$

where $E_1, \dots, E_q$ are the elementary matrices that row reduce A to R.
A is invertible if and only if $\det(A) \neq 0$.

- if $R = I$, then $|R| = |I| = 1$, $|A| \neq 0$ and A is invertible,
- if $R \neq I$, then R must have a zero row, so $|R| = 0$, $|A| = 0$ and A is singular.

Cramer's Rule

Let $A_i(\mathbf{b})$ denote the matrix obtained by replacing the i^{th} column of A by $\mathbf{b}$: $A_i(\mathbf{b}) = [\mathbf{a}_1 \dots \mathbf{b} \dots \mathbf{a}_n]$.
Let A be an invertible $n \times n$ matrix and let $\mathbf{b}$ be a vector in $\mathbb{R}^n$. Then the unique solution $\mathbf{x}$ of the system $A\mathbf{x} = \mathbf{b}$ is given by

$$x_i = \frac{\det(A_i(\mathbf{b}))}{\det(A)} \text{ for } i = 1, \dots, n$$

Definition

The determinant is the "volume" of the geometric shape whose edges are the rows of the matrix:

1. For 2×2 matrices, the determinant corresponds to the area of a parallelogram.
2. For 3×3 matrices, the determinant corresponds to the volume of a parallelepiped.
3. For dimensions $d > 3$, the determinant measures a d-dimensional hyper-volume.

Theorem

Let A be a 2×2 matrix with rows $\mathbf{r}_1, \mathbf{r}_2$ and columns $\mathbf{k}_1, \mathbf{k}_2$.

P_r is the parallelogram in $\mathbb{R}^2$ spanned by $\mathbf{r}_1, \mathbf{r}_2$. P_k is the parallelogram in $\mathbb{R}^2$ spanned by $\mathbf{k}_1, \mathbf{k}_2$.

$$\text{Area}(P_r) = \text{Area}(P_k) = |\det(A)| = \text{abs}(\det(A))$$

Theorem

Let A be a 3×3 matrix with rows $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ and columns $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3$.

P_r is the parallelopiped in $\mathbb{R}^3$ spanned by $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$. P_k is the parallelopiped in $\mathbb{R}^3$ spanned by $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3$.

$$\text{Volume}(P_r) = \text{Volume}(P_k) = |\det(A)| = \text{abs}(\det(A))$$

Definition

Consider a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ defined through the matrix-vector product with a matrix A^T.

After passing through T, the region is transformed to an area / volume / hyper-volume $\det(A^T)$.

$\det(A^T)$ is the scale-factor associated with the linear transformation T:

1. Linear transformations that "shrink" areas have $\det(A^T) < 1$.
2. Linear transformations that preserve areas have $\det(A^T) = 1$.
3. Linear transformations that "enlarge" areas have $\det(A^T) > 1$.

Theorem

If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation with $n \times n$ standard matrix A, and S is a region in $\mathbb{R}^n$, then:

$$\text{Volume}(T(S)) = |\det(A)| \text{Volume}(S)$$

Orthogonal transformations preserve volumes, orthogonal projections do not:

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an **orthogonal transformation**.

Then for any region R in $\mathbb{R}^n$, $\text{Volume}(T(R)) = \text{Volume}(R)$.

Theorem

Let V be a linear subspace of $\mathbb{R}^n$ not equal to $\mathbb{R}^n$. Let $P : \mathbb{R}^n \to \mathbb{R}^n$ be the **orthogonal projection** on V.

Then for any region R in $\mathbb{R}^n$, $\text{Volume}(P(R)) = 0$.

9 Eigenvalues and Eigenvectors

Definition

Let A be a $n \times n$ matrix. A scalar λ is called an eigenvalue of A if there is a corresponding non-zero **eigenvector** $\mathbf{x}$ of A such that
$$A\mathbf{x} = \lambda\mathbf{x}, \mathbf{x} \neq \mathbf{0}$$
The collection of all eigenvectors corresponding to an eigenvalue λ of A, together with the zero vector, is called the **eigenspace** of λ, denoted by E_λ.

The eigenspace is a subspace because it's closed under vector addition and scalar multiplication:

1. If $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors for λ, i.e. $A\mathbf{u} = \lambda\mathbf{u}, A\mathbf{v} = \lambda\mathbf{v}$, then $(\mathbf{u}+\mathbf{v})$ is also an eigenvector for λ, as $A(\mathbf{u}+\mathbf{v}) = \lambda(\mathbf{u}+\mathbf{v})$.
2. If $\mathbf{u}$ is an eigenvector for λ and c is a scalar, then $c\mathbf{u}$ is also an eigenvector for λ, since $A(c\mathbf{u}) = c(A\mathbf{u}) = c(\lambda\mathbf{u}) = \lambda(c\mathbf{u})$.

Theorem

Assume $\mathbf{v}$ is an eigenvector of A with eigenvalue λ. Then

1. $c\mathbf{v}$ is also an eigenvector of A with eigenvalue λ, for any $c \in \mathbb{R}, c \neq 0$.
2. $\mathbf{v}$ is an eigenvector of cA, with eigenvalue $c\lambda$, for every $c \in \mathbb{R}$.
3. $\mathbf{v}$ is an eigenvector of A^k, with eigenvalue λ^k, for every integer $k \geq 0$.

4. 0 is an eigenvalue of A if and only if A is singular (not invertible).
5. If A is invertible, $\mathbf{v}$ is an eigenvector of A^{-1}, with eigenvalue $\lambda^{-1} = 1/\lambda$.
6. If A is invertible, then $\mathbf{v}$ is an eigenvector of A^k, with eigenvalue λ^k, for any integer k.

Proofs:

1. $A\mathbf{v} = \lambda\mathbf{v} \Rightarrow c(A\mathbf{v}) = c(\lambda\mathbf{v}) \Rightarrow A(c\mathbf{v}) = \lambda(c\mathbf{v})$
2. $A\mathbf{v} = \lambda\mathbf{v} \Rightarrow c(A\mathbf{v}) = c(\lambda\mathbf{v}) \Rightarrow (cA)\mathbf{v} = (c\lambda)\mathbf{v}$
3. $A\mathbf{v} = \lambda\mathbf{v} \Rightarrow A^2\mathbf{v} = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda(A\mathbf{v}) = \lambda(\lambda\mathbf{v}) = \lambda^2\mathbf{v} \Rightarrow A^3\mathbf{v} = A(A^2\mathbf{v}) = A(\lambda^2\mathbf{v}) = \lambda^2(A\mathbf{v}) = \lambda^2(\lambda\mathbf{v}) = \lambda^3\mathbf{v}$
4. 0 eigenvalue $A \iff$ a vector $\mathbf{v} \neq 0$ exists such that $A\mathbf{v} = 0\mathbf{v} = \mathbf{0} \iff A$ is not invertible.
5. Suppose A is an eigenvalue of A assuming that A is invertible. From the previous point, we know that $\lambda \neq 0$. Now let $A\mathbf{v} = \lambda\mathbf{v}, \mathbf{v} \neq \mathbf{0}$ and consider the following sequence of implications: $A\mathbf{v} = \lambda\mathbf{v} \Rightarrow A^{-1}(A\mathbf{v}) = A^{-1}(\lambda\mathbf{v}) \Rightarrow (A^{-1}A)\mathbf{v} = \lambda(A^{-1}\mathbf{v}) \Rightarrow \mathbf{v} = \lambda(A^{-1}\mathbf{v}) \Rightarrow \lambda^{-1}\mathbf{v} = A^{-1}\mathbf{v}$

Definition

Let A be an $n \times n$-matrix and λ an eigenvalue of A.
The **eigenspace** of A corresponding to λ is the solution set of the homogeneous system $(A - \lambda I)\mathbf{x} = \mathbf{0}$:
$$E_\lambda = \mathrm{Nul}(A - \lambda I)$$

- $E_\lambda = \mathrm{Nul}(A - \lambda I)$ hence E_λ is a linear subspace of $\mathbb{R}^n$.
- E_λ contains all eigenvectors of A belonging to the eigenvalue λ **and the zero vector** (which is not an eigenvector).
- $\dim(E_\lambda) \geq 1$ (otherwise λ wouldn't be an eigenvalue).

Theorem

Let V be a linear subspace of $\mathbb{R}^n$. Let Proj_V be the **orthogonal projection** onto V and let P_V be its standard matrix; that is, $\mathrm{Proj}_V(\mathbf{x}) = P_V\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Then P_V can only have eigenvalues 0 and 1.

1. If $V \neq \{\mathbf{0}\}$, then V is the eigenspace for 1.
2. If $V \neq \mathbb{R}^n$, then $V^\perp$ is the eigenspace for 0.

Theorem

Let V be a linear subspace of $\mathbb{R}^n$. Let Refl_V be the **orthogonal reflection** through V and let R_V be its standard matrix; that is, $\mathrm{Refl}_V(\mathbf{x}) = R_V\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Then R_V can only have eigenvalues 1 and -1.

1. If $V \neq \{\mathbf{0}\}$, then V is the eigenspace for 1.
2. If $V \neq \mathbb{R}^n$, then $V^\perp$ is the eigenspace for -1.

Theorem

Let A be an $n \times n$ matrix and let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be distinct eigenvalues of A with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are linearly independent.

Theorem

Suppose the $n \times n$ matrix A has eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$.
If $\mathbf{x}$ is a vector in $\mathbb{R}^n$ that can be expressed as a linear combination of these eigenvectors, $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$ - in the ideal case a basis of $\mathbb{R}^n$ consisting of eigenvectors of A, then, for any integer k
$$A^k\mathbf{x} = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_m\lambda_m\mathbf{v}_m$$

9.1 The Characteristic Equation

Finding the eigenvalues of an $n \times n$ matrix amounts to finding the roots of a polynomial of degree n - the **characteristic polynomial** of the matrix.

Definition

The equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** of A.

Theorem

If A is $n \times n$, its **characteristic polynomial**, $\det(A - \lambda I)$, is of degree n.

Corollary

By the Fundamental Theorem of Algebra, it follows that every $n \times n$-matrix has n eigenvalues λ (real or complex, and multiplicities taken into account).

Theorem

The following statements are equivalent:

1. λ is an eigenvalue of a matrix A.
2. There exists a vector $\mathbf{x} \neq \mathbf{0}$ such that $(A - \lambda I)\mathbf{x} = \mathbf{0}$.
3. The matrix $A - \lambda I$ is *not* invertible.
4. $\det(A - \lambda I) = 0$

Theorem

The eigenvalues of a **triangular matrix** are the entries on its main diagonal: $(a_{11} - \lambda)(a_{22} - \lambda) \ldots (a_{nn} - \lambda) = 0$

Proof: Upper-triangular matrices

$$\det(A - \lambda I) = \det \begin{bmatrix} a_{11} - \lambda & a_{12} & \ldots & a_{1n} \\ 0 & a_{22} - \lambda & & \vdots \\ \vdots & 0 & \ddots & \\ & \vdots & \ddots & \\ 0 & 0 & \ldots & a_{nn} - \lambda \end{bmatrix}$$

$$= (a_{11} - \lambda)\det \begin{bmatrix} a_{12} & \ldots & a_{1n} \\ a_{22} - \lambda & & \vdots \\ 0 & \ddots & \\ \vdots & \ddots & \\ 0 & \ldots & a_{nn} - \lambda \end{bmatrix}$$

$$= (a_{11} - \lambda)(a_{22} - \lambda) \ldots (a_{nn} - \lambda) = 0$$

NB: Every matrix is row equivalent to an upper triangular matrix, i.e. echelon form. But **row operations change eigenvalues**!

Process for finding the Eigenspaces (Eigenvalues and Eigenvectors) of a Matrix:

1. Compute the characteristic polynomial $\det(A - \lambda I)$ of A.
2. Find the eigenvalues of A by solving the characteristic equation $\det(A - \lambda I) = 0$ for λ.
3. For each eigenvalue λ, find the null space of the matrix $A - \lambda I$. This is the eigenspace E_λ, the non-zero vectors of which are the eigenvectors of A corresponding to λ.
4. Find a basis for each eigenspace.

9.2 Multiplicities

There are two natural ways to define multiplicity: algebraically and geometrically. Interestingly, although these two multiplicities are related, they are not always equal.

Definition

The **algebraic** multiplicity of λ_i is its multiplicity as a root, i.e. the number of times the factor $(\lambda - \lambda_i)$ appears in, the characteristic polynomial

$$\text{a.m.}(\lambda_i) = \det(A - \lambda I), 1 \leq \text{a.m.}(\lambda_i) \leq n$$

Definition

The **geometric** multiplicity of λ_i is the dimension of its corresponding eigenspace

$$\text{g.m.}(\lambda_i) = E_{\lambda_i} = \mathbf{Nul}(A - \lambda_i I), 1 \leq \text{g.m.}(\lambda_i) \leq n$$

Theorem

Let A be an $n \times n$-matrix. For each of the eigenvalues λ_i of A:

$$1 \leq \text{g.m.}(\lambda_i) \leq \text{a.m.}(\lambda_i) \leq n$$

9.3 Complex Eigenvalues

Definition

$\mathbb{C}^n$ is the set of column vectors $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$ with $z_1, z_2, \ldots, z_n \in \mathbb{C}$.

Addition and scalar multiplication are defined in the same way as in $\mathbb{R}^n$.

Theorem

Let A be a real $n \times n$-matrix. Then the following holds:

If $A\mathbf{v} = \lambda\mathbf{v}$ for some $\lambda \in \mathbb{C}, \mathbf{v} \in \mathbb{C}^n$, then $A\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$.

That is, for every complex eigenvalue-eigenvector pair $\lambda, \mathbf{v}$ of a real matrix A the conjugate pair $\lambda, \mathbf{v}$ is also an eigenvalue-eigenvector pair of A.

This is generally not true for matrices that are not real!

Cayley-Hamilton Theorem

If $p(\mathbf{x}) = a_0 + a_1\mathbf{x} + \cdots + a_{n-1}\mathbf{x}_{n-1} + \mathbf{x}_n$ and A is a square matrix, then there is a square matrix

$$p(A) = a_0 I + a_1 A + \cdots + a_{n-1}A^{n-1} + A^n$$

If $c_A(\lambda)$ is the characteristic polynomial of the matrix A, then $c_A(A) = 0$. (Every matrix satisfies its polynomial equation!)

10 Similarity and Diagonalisation

A matrix A can in some cases be decomposed as $A = PBP^{-1}$ - with B a diagonal or a rotation-scaling matrix. A and B share many properties, and in that sense are similar.

Definition

Two $n \times n$-matrices A and B are **similar** if there exists an invertible $n \times n$-matrix P such that

$$A \sim B \iff A = PBP^{-1} \iff AP = PB$$

The transformation from A to $P^{-1}AP$ is called a **similarity transformation**.

The matrix P depends on A and B, and is not unique for a given pair of similar matrices A and B.

An example of a similarity transformation is diagonalisation: It gives a relation between a matrix and a diagonal matrix. That is convenient, because usually, a diagonal matrix is much easier to work with, e.g. to compute a high power of the matrix.

Theorem

Let A, B, C be $n \times n$ matrices. Then
1. $A \sim A$
2. Symmetry: If $A \sim B$, then $B \sim A$.
3. Transitivity: If $A \sim B$ and $B \sim C$, then $A \sim C$.

Theorem

Let $A = PBP^{-1}$ with change-of-basis matrix $P = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \ldots \mathbf{v}_n\end{bmatrix}$.

If T is the linear transformation with standard matrix A then B is the matrix of T with respect to the basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$:

$$BP^{-1}\mathbf{x} = P^{-1}A\mathbf{x}$$

Theorem

If $A \sim B$,

$$A\mathbf{v} = \lambda\mathbf{v} \iff B(P^{-1}\mathbf{v}) = \lambda(P^{-1}\mathbf{v})$$

That is,

1. A and B have the same eigenvalues.
2. $\mathbf{v}$ is eigenvector of A with eigenvalue $\lambda_i \iff P^{-1}\mathbf{v}$ is eigenvector of B with same eigenvalue λ_i (note: $P^{-1}\mathbf{v} \neq \mathbf{0}$ because P^{-1} is invertible).
3. A and B have the same geometric multiplicity:

$$\dim E_{\lambda_i}(A) = \dim E_{\lambda_i}(B)$$

Theorem

If $A \sim B$,

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

That is,

1. $\det(A) = \det(B)$
2. A is invertible if and only if B is invertible.
3. A and B have the same rank.
4. A and B have the same characteristic polynomial.
5. the multiplicity of each shared root of that polynomial, i.e. the algebraic multiplicity, is the same for A as for B.

Proof:

$$A - \lambda I = PBP^{-1} - \lambda I = PBP^{-1} - \lambda PIP^{-1} = P(B - \lambda I)P^{-1}$$

$$\Rightarrow \quad \det(A - \lambda I) = \det(P(B - \lambda I)P^{-1}) = \det(P)\det(B - \lambda I)\det(P^{-1}) = \det(B - \lambda I)$$

Theorem

If a matrix is subjected to a similarity transformation then the eigenvalues and their multiplicities (algebraic and geometric) are preserved.

Hence, two $n \times n$-matrices A and B cannot be similar if

- A and B do not have the same characteristic polynomial, or
- A and B have the same characteristic polynomial but for at least one of the shared eigenvalues the geometric multiplicities are not the same.

Similar matrices have the same eigenvalues with the same algebraic and geometric multiplicities.

Conversely, however, if two matrices have the same eigenvalues with the same algebraic and geometric multiplicities, they are not necessarily similar.

Example: Both A and B have one eigenvalue 0 with algebraic multiplicity equal to 4 and geometric multiplicity equal to 2.

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

But there is no invertible matrix P such that $A = PBP^{-1}$. ($AP = PB$ implies that the second column of P is equal to the zero column, so P cannot be invertible.)

10.1 Diagonalisation

Definition

An $n \times n$ matrix A is diagonalisable if there is a diagonal matrix D such that A is similar to D; that is, if there is an invertible $n \times n$ matrix P such that

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

Application: Calculating high powers of a matrix:

$$\begin{aligned} A^k &= (PDP^{-1})^k \\ &= PDP^{-1}PDP^{-1} \dots PDP^{-1} \\ &= PDIDI \dots DP^{-1} \\ &= PD^kP^{-1} \end{aligned}$$

Theorem

Let A be an $n \times n$-matrix. Then A is diagonalisable if and only if A has n linearly independent eigenvectors.

That is, there exist an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$ if and only if the columns of P are n *linearly independent eigenvectors* of A and the diagonal entries of D are the eigenvalues of A corresponding to the eigenvectors in P in the same order:

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}, D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Application: Finding the standard matrix of a reflection:

Consider the line $L = \text{Span}\left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\} \in \mathbb{R}^2$. Let R_L be the standard matrix of reflection through L. From geometry, we know that L is an eigenspace with eigenvalue 1, and $L^\perp$ is an eigenspace with eigenvalue -1. In particular, $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ are eigenvectors with eigenvalues 1 and -1, respectively. Since they are linearly independent:

$$\begin{aligned} R_L &= PDP^{-1} \\ &= \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix} \end{aligned}$$

Application: Discrete dynamical systems:

Consider two towns A and B. At a certain point in time, town A has 4,000 inhabitants, and town B has 5,000. Every year, 5% of the population of A moves to B, and 10% of the population of B moves to A. What is the population distribution after 10 years?

$$\begin{aligned} \begin{bmatrix} p_{A,k} \\ p_{B,k} \end{bmatrix} &= \begin{bmatrix} 0.95 & 0.1 \\ 0.05 & 0.9 \end{bmatrix} \begin{bmatrix} 4000 \\ 5000 \end{bmatrix} \\ &= PD^kP^{-1} \begin{bmatrix} 4000 \\ 5000 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.85 \end{bmatrix}^k \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 4000 \\ 5000 \end{bmatrix} \\ &= PD^kP^{-1} \begin{bmatrix} 4000 \\ 5000 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.85 \end{bmatrix}^k \frac{1}{3} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 4000 \\ 5000 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2 + 0.85^k & 2 - 2 \cdot 0.85^k \\ 1 - 0.85^k & 1 + 2 \cdot 0.85^k \end{bmatrix} \begin{bmatrix} 4000 \\ 5000 \end{bmatrix} \end{aligned}$$

In the long run, the solution stabilises:

$$\lim_{k \to \infty} \begin{bmatrix} p_{A,k} \\ p_{B,k} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} p_{A,0} \\ p_{B,0} \end{bmatrix} = \begin{bmatrix} \frac{2}{3}(p_{A,0} + p_{B,0}) \\ \frac{1}{3}(p_{A,0} + p_{B,0}) \end{bmatrix}$$

Theorem

Let A be an $n \times n$-matrix. If A is diagonalisable, then:

1. There exists an **eigenbasis**, i.e. a set of n linearly independent eigenvectors / n distinct eigenvalues of A.
2. The sum of the geometric multiplicities of the eigenvalues of A, i.e. the sum of the dimensions of the eigenspaces of A, is equal to n. (If it is less than n, there aren't enough eigenvectors for an n-dimensional eigenbasis.)
3. For each eigenvalue of A, the algebraic multiplicity is equal to the geometric multiplicity. ($n \geq a.m. \geq g.m. \geq 1$, so $g.m. = n \Rightarrow a.m. = n = g.m.$)

It follows from statement 4 that:

Corollary

If all eigenvalues have $a.m. = 1$ (for a total of n), then A is diagonalisable.

10.2 Rotation-Scaling Matrices

If $A = PDP^{-1}$ is a diagonalisable matrix with complex eigenvalues, then the matrices P and D necessarily have complex entries.

However, for 2×2-matrices you can also make the decomposition $A = PCP^{-1}$ where C is the so-called rotation-scaling matrix - the *composition of a rotation matrix and a scaling matrix.*

Theorem

Let A be a real 2×2-matrix with eigenvalues $a \pm ib$, where $b \neq 0$ and $a, b \in \mathbb{R}$. Then there exists an invertible real matrix P such that

$$A = P \begin{bmatrix} a & -b \\ b & a \end{bmatrix} P^{-1} \iff P = \begin{bmatrix} \text{Re } \mathbf{v} & \text{Im } \mathbf{v} \end{bmatrix}$$

where $\mathbf{v}$ is an eigenvector of A that belongs to the eigenvalue $a - ib$.

11 Symmetric Matrices

Definition

A square matrix A is called **symmetric** if $A^T = A$.

Definition

A square matrix A is **skew-symmetric** if $A^T = -A$.

Examples of symmetric matrices:

- A standard matrix P for an **orthogonal projection**
- A **reflection** about a line through the origin that makes angle θ with the x_1-axis:

$$A = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}$$

Theorem

If A is a symmetric matrix, then *any two eigenvectors corresponding to distinct eigenvalues* of A are **orthogonal**.

Proof:

$$\begin{aligned} \lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 &= (A\mathbf{v}_1) \cdot \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 = (\mathbf{v}_1^T A^T)\mathbf{v}_2 \\ &= \mathbf{v}_1^T (A\mathbf{v}_2) = \mathbf{v}_1^T(\lambda_2 \mathbf{v}_2) = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2 \end{aligned}$$

Hence $(\lambda_2 - \lambda_1)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$. Since $\lambda_1 \neq \lambda_2$, we find $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Theorem

If A is a square matrix, then $A + A^T$ is a symmetric matrix.

Theorem

If a matrix A is symmetric, it has **only real eigenvalues**.

Proof: Suppose $A\mathbf{x} = \lambda \mathbf{x}$ for some $0 \neq \mathbf{x} \in \mathcal{C}^n$. Then

$$\begin{aligned} \lambda \mathbf{x} \cdot \bar{\mathbf{x}} &= A\mathbf{x} \cdot \bar{\mathbf{x}} = (A\mathbf{x})^T \bar{\mathbf{x}} = \mathbf{x}^T A^T \bar{\mathbf{x}} = \mathbf{x}^T A \bar{\mathbf{x}} \\ &= \mathbf{x}^T \bar{A}\bar{\mathbf{x}} = \mathbf{x}^T \bar{\lambda}\bar{\mathbf{x}} = \bar{\lambda} \mathbf{x}^T \bar{\mathbf{x}} \\ &= \bar{\lambda} \mathbf{x} \cdot \bar{\mathbf{x}} \end{aligned}$$

Since $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x} \cdot \bar{\mathbf{x}} = x_1 \bar{x_1} + \cdots + x_n \bar{x_n} = |x_1|^2 + \cdots + |x_n|^2 > 0$$

So $\lambda = \bar{\lambda}$ and, hence, λ is real.

Only square matrices can be symmetric. However, from a non-square matrix we can easily build a symmetric square matrix, as can be seen from the following theorem (which plays an important role in the Singular Value Decomposition):

Theorem

For any matrix A (potentially non-square), the matrices AA^T and A^TA are symmetric.

Proof: Let A be an $n \times m$-matrix. Then A^TA is $m \times n$, AA^T is $n \times n$, and both are symmetric:

$$(A^TA)^T = A^T(A^T)^T = A^TA \qquad (AA^T)^T = (A^T)^TA^T = AA^T$$

11.1 Orthogonal Diagonalisation

Definition

An $n \times n$-matrix A is **orthogonally diagonalisable** if there exists an *orthogonal* matrix Q, i.e. $Q^{-1} = Q^T$, and a *diagonal* matrix D such that

$$A = QDQ^{-1} (= QDQ^T)$$

In other words, A must be diagonalisable with an orthogonal transformation matrix.

Any matrix that is symmetric and diagonalisable, is orthogonally diagonalisable.

In fact, every symmetric matrix is diagonalisable.

Theorem

If A is a symmetric matrix, then A is diagonalisable.

Spectral Theorem

Let A be a $n \times$ real matrix. Then A is **symmetric if and only if it is orthogonally diagonalisable**.

Algorithm for orthogonal diagonalisation:

1. Compute the eigenvalues of A.
2. For each eigenvalue, construct an *orthonormal* basis for the corresponding eigenspace:
   - Use the Gram-Schmidt process (if necessary).
   - Normalise the basis vectors.
3. Construct the matrix P using the basis vectors constructed in step 2 as columns.
4. Construct the matrix D using the corresponding eigenvalues on the diagonal.

Spectral Decomposition

If $A = QDQ^T = [\mathbf{q}_1 \dots \mathbf{q}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \Longleftrightarrow \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix}$,

then

$$A = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T = \sum_{i=1}^n \lambda_i \mathbf{q}_i \mathbf{q}_i^T$$

Definition

A quadratic form on $\mathbb{R}^n$ is a function of the form

$$Q(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$$

Examples:

- Ellipses: $2x_1^2 + x_2^2 = 1$
- Hyperbolas: $x_1^2 - x_2^2 = 1$
- Lines: $4x_1^2 - 3x_2^2 = 0$

Definition

If A is an $n \times n$-matrix, then $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a quadratic form on $\mathbb{R}^n$.

$$Q\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$$

Example: If A is the identity matrix, then

$$Q(\mathbf{x}) = \mathbf{x}^T I_n \mathbf{x} = x_1^2 + \dots + x_n^2 = ||\mathbf{x}||^2$$

Theorem

For every quadratic form $Q(\mathbf{x})$ on $\mathbb{R}^n$ there exists a unique *symmetric* $n \times n$-matrix A such that

$$A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

for every vector $\mathbf{x}$ in $\mathbb{R}^n$.

Principal-Axis Theorem

Consider a quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ for a symmetric $n \times n$-matrix A.

There exists an orthogonal matrix P such that the change of variable $\mathbf{x} = P\mathbf{y}$ transforms Q into a quadratic form $\sim Q(\mathbf{y}) = \mathbf{y}^T D\mathbf{y}$ with no cross terms.

Definition

The **principal axes** of the quadratic form Q are the lines that are generated by the columns of P.

Algorithm to remove cross terms: Let $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be the quadratic form with symmetric matrix A.

1. Orthogonal diagonalisation $A = PDP^T$ with orthonormal basis $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for the columns of P.
2. Define the new variable $\mathbf{y} = [\mathbf{x}]_{\mathcal{B}} = P^T\mathbf{x}$. Note that $\mathbf{x} = P\mathbf{y}$, since $P^T = P^{-1}$.
3. Then

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T A\mathbf{x} = \mathbf{x}^T PDP^T\mathbf{x} = (P^T\mathbf{x})^T D(P^T\mathbf{x}) = \mathbf{y}^T D\mathbf{y} \\ &= \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 \end{aligned}$$

Definition

The quadratic form $Q(x) = \mathbf{x}^T A\mathbf{x}$ is

positive semi-definite	$\iff$	$Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$
positive definite	$\iff$	$Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$
indefinite	$\iff$	$Q(\mathbf{x})$ assumes both positive and negative values
negative semi-definite	$\iff$	$Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$
negative definite	$\iff$	$Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$

Application: The second-derivative test:

- Local maximum $\iff$ negative definite Hessian matrix
- Local minimum $\iff$ positive definite Hessian matrix
- Saddle point $\iff$ indefinite Hessian matrix

Classification Theorem

Let A be an $n \times n$ symmetric matrix. Then the quadratic form $\mathbf{x}^T A\mathbf{x}$ is defined by the eigenvalues of matrix A:

quadratic form $\mathbf{x}^T A\mathbf{x}$:		eigenvalues of A:
positive semi-definite	$\iff$	all ≥ 0
positive definite	$\iff$	all > 0
indefinite	$\iff$	both > 0 and < 0
negative semi-definite	$\iff$	all ≤ 0
negative definite	$\iff$	all < 0

Theorem

Let A be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

The **maximal value** of $\mathbf{x}^T A\mathbf{x}$ under the constraint $||\mathbf{x}|| = 1$ is λ_1. This value is attained if $\mathbf{x}$ is any unit-eigenvector of A corresponding to λ_1.

The **minimal value** of $\mathbf{x}^T A\mathbf{x}$ under the constraint $||\mathbf{x}|| = 1$ is λ_n. This value is attained if $\mathbf{x}$ is any unit-eigenvector of A corresponding to λ_n.

Theorem

Let A be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$ and let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthonormal set of corresponding eigenvectors. Then the maximal value of $\mathbf{x}^T A\mathbf{x}$ subject to the constraints

$$\begin{aligned} ||\mathbf{x}|| &= 1 \\ \mathbf{x} \cdot \mathbf{v} &= 0 \\ &\vdots \\ \mathbf{x} \cdot \mathbf{v}_{k-1} &= 0 \end{aligned}$$

is equal to the eigenvalue λ_k of A. This maximal value is at least attained at $x = \pm\mathbf{v}_k$.

Theorem

The eigenvalues of the symmetric matrix A^TA are **non-negative** for any $m \times n$ matrix A.

Proof: Let λ be the eigenvalue of A^TA corresponding to a vector $\mathbf{v} s.t. \mathbf{v}^T\mathbf{v} = 1$. Then

$$\lambda = \mathbf{v}^T \lambda \mathbf{v} = \mathbf{v}^T A^T A \mathbf{v} = (A\mathbf{v}) \cdot (A\mathbf{v}) = ||A\mathbf{v}||^2 \geq 0$$

Definition

The **singular values** of an $m \times n$ matrix A are the square roots of the eigenvalues of A^TA, usually denoted by

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$$

including multiplicities and arranged in decreasing order.

Definition

Let A be an $m \times n$ matrix of rank r. A set of **right-singular** vectors is an orthonormal set of eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of A^TA such that the corresponding eigenvalues are in descending order.

$$(A\mathbf{v}_i) \cdot (A\mathbf{v}_j) = \mathbf{v}_i^T A^T A \mathbf{v}_j = \sigma_j^2 \mathbf{v}_i^T \mathbf{v}_j = \begin{cases} \sigma_j^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$||A\mathbf{v}_i|| = \sqrt{(A\mathbf{v}_i) \cdot (A\mathbf{v}_i)} = \sigma_i$$

Since the lengths of these vectors are given exactly by the singular values and they follow the same ordering as the singular values, the first r vectors are non-zero and the rest are zero by virtue of the rank of the matrix A:

$$\{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\} = \{A\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{0}, \ldots, \mathbf{0}\}$$

Theorem

If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a set of right-singular vectors for the matrix A then the set $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n\}$ is orthogonal. Moreover, the set $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ is an orthogonal basis for $\mathrm{Col}(A)$.

Definition

Let A be an $m \times n$ matrix of rank r. A set of **left-singular** vectors is an orthonormal set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ where $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is obtained by normalising the set $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ for some set of right-singular vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$.

Definition

Let A be an $m \times n$ matrix of rank r. The **matrix of singular values** Σ is the $m \times n$ matrix obtained by adding rows and columns of zeros where necessary to the $r \times r$ matrix

$$D = \begin{bmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_r \end{bmatrix}$$

Denote the matrix whose columns are a set of right-singular vectors by V.

Denote the matrix whose columns are a set of left-singular vectors by U.

Singular Value Decomposition

Any $m \times n$ matrix A may be decomposed as a product

$$A = U\Sigma V^T$$

where U is an orthogonal $m \times m$ matrix, V is an orthogonal $n \times n$ matrix, and Σ is the matrix of singular values.

Proof: For $i = 1, \ldots, n$, $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$, and so

$$\begin{aligned} AV = \begin{pmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \ldots & A\mathbf{v}_n \end{pmatrix} &= \begin{pmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \ldots & \sigma_n\mathbf{u}_n \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \ldots & \sigma_r\mathbf{u}_r & \mathbf{0} \end{pmatrix} \end{aligned}$$

$$U\Sigma = \begin{pmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \ldots & \sigma_r\mathbf{u}_r & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix} \Rightarrow A = U\Sigma V^T$$

Applications:

1. reliable numerical computations
2. pseudo-inverses / least-squares solutions / linear regression
3. bases of fundamental spaces
4. principal component analysis (PCA)

Definition

Description: Text

$$\mathbf{v} \cdot \mathbf{w} = 0 \iff \alpha = \frac{\pi}{2} \iff \mathbf{v} \perp \mathbf{w}$$

Principal component analysis uses the singular value decomposition to determine the dimensionality of the data, and provides the data in a new frame of reference that removes redundancy and orders the data according to variance.

12 Vector Spaces

Definition

Let V be a set on which two operations, addition and scalar multiplication, are defined. If the following axioms hold for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and scalars c, d, then V is called a vector space and its elements are called vectors.

- Closure under addition: $\mathbf{u} + \mathbf{v} \in V$
- Commutativity: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- Associativity: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
- There exists an element $\mathbf{0} \in V$, called the zero vector, such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
- For each $\mathbf{u} \in V$, there exists an element $-\mathbf{u} \in V$ s.t. $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
- Closure under scalar multiplication: $c\mathbf{u} \in V$
- Distributivity: $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
- Distributivity: $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{v}$
- $c(d\mathbf{u}) = (cd)\mathbf{u}$
- $1\mathbf{u} = \mathbf{u}$

Theorem

Let V be a vector space, $\mathbf{u}$ a vector in V, and c a scalar. Then

- $0\mathbf{u} = \mathbf{0}$
- $c\mathbf{0} = \mathbf{0}$
- $(-1)\mathbf{u} = -\mathbf{u}$
- If $c\mathbf{u} = \mathbf{0}$, then $c = 0$ or $\mathbf{u} = \mathbf{0}$.

A vector space consists of a set of vectors and all linear combinations of these vectors, and is equipped with operations of addition and scalar multiplication, satisfying certain properties like closure under addition and scalar multiplication, associativity, commutativity, and distributivity.

A linear subspace is a subset of a vector space operating that is itself a vector space. It retains all the properties of a vector space - within that subset, and it operates within the context of a larger vector space.

12.1 Subspaces

A subset of a vector space, $S \subseteq V$, can be described in the form $S = \{\mathbf{v} \in V | \langle \text{ some condition} \rangle\}$.

A subspace of a vector space, $W \subseteq V$, is a *subset with vector space structure*, meaning it is closed under addition (for all $\mathbf{w}_1, \mathbf{w}_2 \in W, \mathbf{w}_1 + \mathbf{w}_2 \in W$) and scalar multiplication (for all $\mathbf{w} \in W, \lambda \in \mathbb{R}, \lambda\mathbf{w} \in W$). You can take arbitrary elements in the set, scale or add them together, and obtain an element of the same set.

Definition

A subset W of a vector space V is called a subspace of V if W is itself a vector space with the same scalars, closure under addition and scalar multiplication as V.

Checking whether a subset W of a vector space V is a subspace of V involves testing only two of the ten vector space axioms.

Theorem

Let V be a vector space, and let W be a non-empty subset of V. Then W is a subspace of V if and only if:

1. If $\mathbf{u}, \mathbf{v} \in W$, then $\mathbf{u} + \mathbf{v} \in W$.
2. If $\mathbf{u} \in W, c \in \mathbb{R}$, then $c\mathbf{u} \in W$.

Definition

If $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a set of vectors in a vector space V, then the set of linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_k$ is called the span of $\mathbf{v}_1, \dots, \mathbf{v}_k$, denoted by $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ or $\text{Span}(S)$.

If $V = \text{Span}(S)$, then S is called a **spanning set** for V, and V is said to be spanned by S.

Theorem

Let $\mathbf{v}_1, \dots, \mathbf{v}_k$ be vectors in a vector space V. Then
- $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ is a subspace of V.
- $\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ is the smallest subspace of V that contains $\mathbf{v}_1, \dots, \mathbf{v}_k$.

A matrix $M \in \mathbb{R}^{m \times n}$ has four fundamental vector spaces: the column space, row space, null space, and left null space.

Definition

The **column space** $\mathcal{C}(M)$ is the span of the columns of the matrix. It consists of all possible output vectors the matrix can produce when multiplied by a vector:

$$\mathcal{C}(M) \equiv \{\mathbf{w} \in \mathbb{R}^m | \mathbf{w} = M\mathbf{v} \text{ for some } \mathbf{v} \in \mathbb{R}^n\}$$

Definition

The **row space** $\mathcal{R}(M)$ is the span of the rows of the matrix. It consists of all possible output vectors the matrix can produce when multiplied by a vector from the left:

$$\mathcal{R}(M) \equiv \{\mathbf{v} \in \mathbb{R}^n | \mathbf{v} = \mathbf{w}^T M \text{ for some } \mathbf{w} \in \mathbb{R}^m\}$$

Definition

The **null space** $\mathcal{N}(M)$ or **kernel** of a matrix $M \in \mathbb{R}^{m \times n}$ consists of all vectors the matrix M sends to the zero vector:

$$\mathcal{N}(M) \equiv \{\mathbf{v} \in \mathbb{R}^n | M\mathbf{v} = \mathbf{0}\}$$

Definition

The **left null space** $\mathcal{N}(M^T)$ of a matrix $M \in \mathbb{R}^{m \times n}$ consists of all vectors the matrix M sends to the zero vector when multiplied from the left:

$$\mathcal{N}(M^T) \equiv \{\mathbf{w} \in \mathbb{R}^m | \mathbf{w}^T M = \mathbf{0}\}$$

The column space and the row space of a matrix have the same dimension.

Definition

The **rank** of M is the number of linearly independent rows in M, which is equal to the number of linearly independent columns in M:

$$\text{rank}(M) \equiv \dim(\mathcal{R}(M)) = \dim(\mathcal{C}(M))$$

Definition

The dimension of the null space of M is called the **nullity**:

$$\text{nullity}(M) \equiv \dim(\mathcal{N}(M))$$

Rank-Nullity Theorem

$$\text{rank}(M) + \text{nullity}(M) = n = \dim(\mathbb{R}^n)$$

12.2 Linear Independence, Basis and Dimension

Definition

A set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ in a vector space V is linearly dependent if there are scalars $c_1, \dots, c_k$, at least one of which is not zero, such that $c_1\mathbf{v}_1 + \cdots + c_1\mathbf{v}_k = \mathbf{0}$.

A set of vectors that is not linearly dependent is said to be linearly independent.

Theorem

The set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ in a vector space V is linearly dependent if at least of the vectors can be expressed as a linear combination of the others.

Theorem

A set S of vectors in a vector space V is linearly dependent if it contains finitely many linearly dependent vectors. A set of vectors that is not linearly dependent is said to be linearly independent.

Definition

A subset $\mathcal{B}$ of a vector space V is a **basis** for V if

- $\mathcal{B}$ spans V, and
- $\mathcal{B}$ is linearly independent.

Definition

Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for a vector space V. Let $\mathbf{v}$ be a vector in V, and write $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Then $c_1, \dots, c_n$ are called the **coordinates** of $\mathbf{v}$ with respect to $\mathcal{B}$, and the column vector

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

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

Theorem

Let $\mathcal{B}$ be a vector space and let $\mathcal{B}$ be a basis for V. For every vector $\mathbf{v} \in V$, there is exactly one way to write $\mathbf{v}$ as a linear combination of the basis vectors in $\mathcal{B}$.

Theorem

Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for a vector space V. Let $\mathbf{u}, \mathbf{v}$ be vectors in V and let c be a scalar. Then

- $[\mathbf{u} + \mathbf{v}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{B}} + [\mathbf{v}]_{\mathcal{B}}$
- $[c\mathbf{u}]_{\mathcal{B}} = c[\mathbf{u}]_{\mathcal{B}}$
- $[c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k]_{\mathcal{B}} = c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_k[\mathbf{u}_k]_{\mathcal{B}}$

Theorem

Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for a vector space V and let $\mathbf{u}_1, \dots, \mathbf{u}_k$ be vectors in V. Then $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is linearly independent in V if and only if $\{[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_k]_{\mathcal{B}}\}$ is linearly independent in $\mathbb{R}^n$.

Theorem

Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for a vector space V.

Any set of more than n vectors in V must be linearly dependent.

Any set of fewer than n vectors in V cannot span V.

The Basis Theorem

If a vector space V has a basis with n vectors, then every basis for V has exactly n vectors.

Definition

The dimension of V, $\dim(V)$, is the number of vectors in a basis for V. A vector is called **(in)finite-dimensional** if it has a basis consisting of (in)finitely many vectors.

The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero, that of a line through the origin to be one, that of a plane through the origin to be two, that of $\mathbb{R}^3$ to be three, and so on.

Theorem

Let V be a vector space with $\dim(V) = n$. Then

- Any linearly independent set in V contains at most n vectors.
- Any spanning set for V contains at least n vectors.

- Any linearly independent set of exactly n vectors in V is a basis for V.
- Any spanning set for V consisting of exactly n vectors is a basis for V.

- Any linearly independent set in V can be extended to a basis for V.
- Any spanning set for V can be reduced to a basis for V.

Theorem

Let W be a subspace of a finite-dimensional vector space V. Then W is finite-dimensional and $\dim(W) \leq \dim(V)$.

$\dim(W) = \dim(V)$ if and only if $W = V$.

12.3 Linear Transformations

Definition

A linear transformation from a vector space V to a vector space W is a mapping $T : V \to W$ such that, for all vectors $\mathbf{u}, \mathbf{v} \in V$ and all scalars c,

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
- $T(c\mathbf{u}) = cT(\mathbf{u})$

Theorem

Let $T : V \to W$ be a linear transformation. Then

- $T(\mathbf{0}) = \mathbf{0}$
- $T(-\mathbf{v}) = -T(\mathbf{v})$ for all $\mathbf{v} \in V$
- $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$

Theorem

Let $T : V \to W$ be a linear transformation and let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a spanning set for V.

Then $T(\mathcal{B}) = \{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$ spans the range of T.

Definition

Composition of linear transformations: If $T : U \to V, S : V \to W$ are linear transformations, then the composition of S with T, pronounced S of T, is the mapping $S \circ T$ defined by

$$(S \circ T)(\mathbf{u}) = S(T(\mathbf{u})), \mathbf{u} \in U$$

Theorem

If $T : U \to V, S : V \to W$ are linear transformations, then $S \circ T : U \to W$ is a linear transformation.

Definition

Inverses of linear transformations: A linear transformation $T : V \to W$ is invertible if there is a linear transformation $T' : W \to V$ such that $T' \circ T = I_V$ and $T \circ T' = I_W$, where T' is the inverse of T.

Theorem

If a transformation T is invertible, its inverse is unique.

12.4 The Kernel and Range of a Linear Transformation

Definition

Let $T : V \to W$ be a linear transformation.

- The **kernel** of T is the set of all vectors in V that are mapped by T to $\mathbf{0}$ in W:

$$\mathrm{Ker}(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$$

- The **range** of T is the set of all vectors in W that are images of vectors in V under T:

$$\mathrm{Range}(T) = \{\mathbf{w} \in W : \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \in V\}$$

Theorem

Let $T : V \to W$ be a linear transformation. Then

- The kernel of T is a subspace.
- The range of T is a subspace.

Definition

Let $T : V \to W$ be a linear transformation.

- The **nullity** of T is the dimension of the kernel of T
- The **rank** of T is the dimension of the range of T.

The Rank Theorem

Let $T : V \to W$ be a linear transformation from a finite-dimensional vector space V into a vector space W. Then

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = \dim(V)$$

Definition

A linear transformation $T : V \to W$ is called **one-to-one** if T maps distinct vectors in V to distinct vectors in W:

$$\forall \mathbf{u}, \mathbf{v} \in V, \mathbf{u} = \mathbf{v} \Rightarrow T(\mathbf{u}) = T(\mathbf{v}), \mathbf{u} \neq \mathbf{v} \Rightarrow T(\mathbf{u}) \neq T(\mathbf{v})$$

If $\mathrm{Range}(T) = W$, then T is called **onto**:

$$\forall \mathbf{w} \in W, \exists \text{ at least one } \mathbf{v} \in V \text{ s.t. } T(\mathbf{v}) = \mathbf{w}$$

The kernel of a linear transformation T is the same as the null space of its matrix representation M_T.

Theorem

A linear transformation $T : V \to W$ is one-to-one if and only if $\mathrm{Ker}(T) = \{\mathbf{0}\}$.

Theorem

Let $\dim(V) = \dim(W) = n$. Then a linear transformation $T : V \to W$ is one-to-one if and only if it is onto.

Theorem

Let $T : V \to W$ be one-to-one. If $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a linearly independent set in V, then $T(S) = \{T(\mathbf{v}_1), \dots, T(\mathbf{v}_k)\}$ is a linearly independent set in W.

Corollary

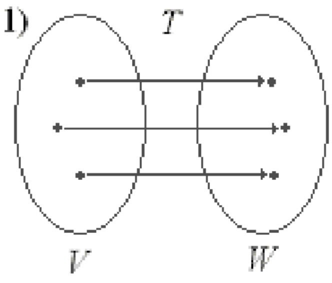
Let $\dim(V) = \dim(W) = n$. Then a one-to-one linear transf. $T : V \to W$ maps a basis for V to a basis for W.

By the invertible matrix theorem, a matrix A is invertible if and only if its null space contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\} \iff A$ is invertible. By the zero kernel condition, a linear transformation T is invertible if and only if its kernel contains only the zero vector, $\mathrm{Ker}(T) = \{\mathbf{0}\}$:

- Invertible linear transformations $T : \mathbb{R}^n \to \mathbb{R}^n$ map different input vectors $\mathbf{x}$ to different output vectors $\mathbf{y} = T(\mathbf{x})$; it's possible to build an inverse $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ that restores every $\mathbf{y}$ back to the $\mathbf{x}$ it came from.
- In contrast, a non-invertible linear transformation S sends all vectors $\mathbf{x} \in \mathrm{Ker}(S)$ to the zero vector $S(\mathbf{x}) = \mathbf{0}$. There is no way to undo the action of S since we can't determine the original $\mathbf{x}$ that was sent to $\mathbf{0}$.

Theorem

A linear transformation $T : V \to W$ is invertible if and only if it is one-to-one and onto.



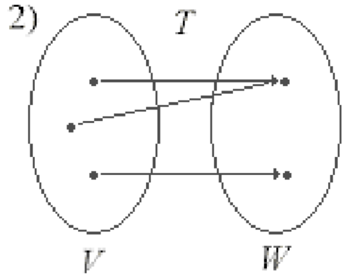
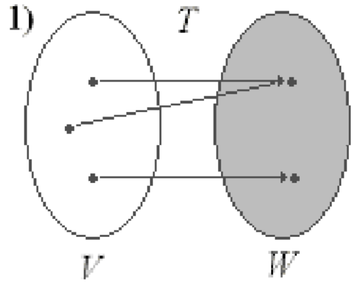


Figure 1

(a) T is one-to-one

(b) T is not one-to-one



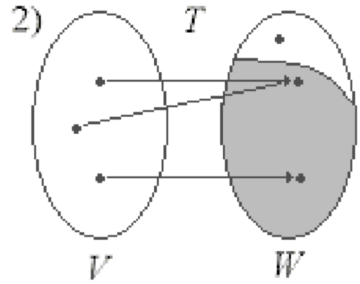


Figure 2

(c) T is onto

(d) T is not onto

Definition

A linear transformation that is one-to-one and onto is called an **isomorphism** (iso = equal, morph = shape). If V and W are two vector spaces such that there is an isomorphism from V to W, then V is isomorphic to W:

$$V \simeq W$$

Theorem

Let V, W be two finite-dimensional vector spaces (over the same field of scalars). Then V is isomorphic to W if and only if

$$\dim(V) = \dim(W)$$

The Fundamental Theorem of Invertible Matrices

Let A be an $n \times n$ matrix. The following statements are equivalent:

1. A is invertible.
2. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.
3. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
4. The reduced row echelon form of A is I_n.
5. A is a product of elementary matrices.
6. $\operatorname{rank}(A) = n$.
7. $\operatorname{nullity}(A) = 0$.
8. The column vectors of A are linearly independent.
9. The column vectors of A span $\mathbb{R}^n$.
10. The column vectors of A form a basis for $\mathbb{R}^n$.
11. The row vectors of A are linearly independent.
12. The row vectors of A span $\mathbb{R}^n$.
13. The row vectors of A form a basis for $\mathbb{R}^n$.
14. $\det(A) \neq 0$.
15. 0 is not an eigenvalue of A.
16. T is invertible.
17. T is one-to-one.
18. T is onto.
19. $\operatorname{Ker}(T) = \{\mathbf{0}\}$.
20. $\operatorname{Range}(T) = W$.

12.5 Inner Product Spaces

Definition

The **inner product** takes pairs of vectors as inputs and produces scalars as outputs ($\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$):

$$\mathbf{u}^T\mathbf{v} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = u_1v_1 + u_2v_2 + u_3v_3$$

Definition

The **outer product** takes some vectors $\mathbf{u}, \mathbf{v}$ as inputs and produces matrices as outputs ($\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$) such that every column is a multiple of the vector $\mathbf{u}$, and every row is a multiple of the vector $\mathbf{v}^T$:

$$\mathbf{u}\mathbf{v}^T = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \end{bmatrix}$$

Definition

An inner product on a vector space V is an operation that assigns to every pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ such that the following properties hold for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars c:

- $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
- $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
- $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
- $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$
- $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

A vector space with such an inner product is called an **inner product space**.

Example: The dot product on $\mathbf{R}^n$ is an inner product.

Example: Continuous functions on an interval $[a, b]$ on the vector space $C[a, b]$ are inner products:

Definition

Let f, g be continuous functions on an interval $[a, b]$. Then the inner product on the vector space $C[a, b]$ is defined as:

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

Example: Let $f(x) = 1, g(x) = x^3$ on $C[0, 2]$. Their inner product is

$$\langle f, g \rangle = \int_0^2 x^3\, dx = \frac{1}{4}x^4\Big|_0^2 = 4$$

Theorem

Let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be vectors in an inner product space V and let c be a scalar. Then

- $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
- $\langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
- $\langle \mathbf{u}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle = 0$

Definition

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

- the **norm** or magnitude of $\mathbf{v}$ is: $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$
- their **distance** is: $d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$
- $\langle \mathbf{u}, \mathbf{v} \rangle = 0 \iff \mathbf{u} \perp \mathbf{v}$

Note: $||\mathbf{v}||$ is always defined, since $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ by the definition of the inner product.

Example: Let $f(x) = 1, g(x) = x^3$ on $C[0, 2]$. Then their norms and distance are:

$$||f|| = \sqrt{\langle f, f \rangle} = \sqrt{\int_0^2 f(x)f(x)\, dx} = \sqrt{x\big|_0^2} = \sqrt{2 - 0} = \sqrt{2} \qquad ||g|| = \sqrt{\langle g, g \rangle} = \sqrt{\int_0^2 g(x)g(x)\, dx} = \sqrt{x^7/7\big|_0^2} = \sqrt{2^7/7}$$

$$||f - g|| = \sqrt{\langle f - g, f - g \rangle} = \sqrt{\int_0^2 (f(x) - g(x))^2\, dx}$$

Example: Let $f(x) = 1, g(x) = \sin x$. In the vector space $C[0, \pi]$ with inner product $\langle f, g \rangle = \int_0^\pi f(x)g(x)\, dx$, are f, g orthogonal?

Example: Let $f(x) = 1, g(x) = \sin x$. In the vector space $C[-\pi, \pi]$ with inner product $\langle f, g \rangle = \int_{-\pi}^\pi f(x)g(x)\, dx$, are f, g orthogonal?

The Pythagorean Theorem

Let $\mathbf{u}, \mathbf{v}$ be vectors in an inner product space V. Then $\mathbf{u}, \mathbf{v}$ are orthogonal if and only if $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$.

$$\begin{aligned}
||\mathbf{u} + \mathbf{v}||^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle && (8) \\
&= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle && (9) \\
&= ||\mathbf{u}||^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + ||\mathbf{v}||^2 && (10) \\
&= ||\mathbf{u}||^2 + ||\mathbf{v}||^2 \qquad \text{if and only if } \langle \mathbf{u}, \mathbf{v} \rangle = 0 && (11)
\end{aligned}$$

Definition

Given a vector space V with an inner product, a list $\{v_1; v_2; \ldots; v_n\}$ in V is **orthogonal** if $\langle v_i, v_j \rangle = 0$ for all $i \neq j$.

The list is orthonormal if it is orthogonal, and $||v_i|| = 1$ for all i.

Theorem

Any orthogonal list of non-zero vectors is linearly independent.

References

[1] Matthew Towers, UCL. "MATH0007: Algebra for Joint Honours Students". In: (2020). URL: https://www.ucl.ac.uk/~ucahmto/0007/_book/.

[2] Poole, David. "Linear Algebra: A Modern Introduction". In: (2012).

[3] TU Delft. "Mastering Linear Algebra". In: (2024). URL: https://online-learning.tudelft.nl/programs/mastering-linear-algebra/.