Mathematics for Data Science II

Index :

Contents pg.no

LINEAR ALGEBRA

Types of matrices :

Definition 1.3.2. A square matrix is a matrix in which the number of rows is the same as the number of columns.

Example 1.3.2.

$$\begin{bmatrix} 0.3 & 5 & -7 \\ 2.8 & 0 & 1 \\ 0 & -2.5 & -1 \end{bmatrix}_{3\times 3}$$

This is a 3×3 matrix (3 rows and 3 columns).

- The i-th diagonal entry of a square matrix is the (i, i)-th entry.
- The diagonal of a square matrix is the set of diagonal entries.
- $0.3, 0$ and -1 are the diagonal entries of the square matrix given above.

Definition 1.3.3. A square matrix in which all entries except the diagonal are 0 is called a diagonal matrix.

Example 1.3.3.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 4.2 \end{bmatrix}_{3\times 3}$$

This is a 3×3 diagonal matrix, where diagonal entries are $1, -3$ and 4.2

Definition 1.3.4. A diagonal matrix in which all the entries in the diagonal are equal is called a scalar matrix.

Example 1.3.4.

$$\begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{bmatrix}_{3\times 3}$$

This is a 3×3 diagonal matrix, where all the diagonal entries are -3.

* Upper triangular matrix : A matrix whose all elements below diagonal are '0' is an Upper triangular matrix.

Note : In an upper triangular matrix, the Determinant equals the product of diagonal elements of the matrix.

Definition 1.3.5. The scalar matrix with all diagonal entries 1 is called the Identity matrix and is denoted by I.

Example 1.3.5.

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

This is the 3×3 identity matrix.

Trace of a matrix : The sum of diagonal elements of a matrix is called Trace of a matrix and represented by Tr(M).

1.3.6 Properties of matrix addition and multiplication

In this subsection we mention properties of matrix addition and multiplication without the proof. We want all of you to verify these on your own.

- $(A + B) + C = A + (B + C)$ (Associativity of addition)
- $(AB)C = A(BC)$ (Associativity of multiplication)
- $A + B = B + A$ (Commutativity of addition)
- In general $AB \neq BA$ (assuming both make sense)
- $\lambda(A + B) = \lambda A + \lambda B$ for some real number λ.
- $\lambda(AB) = (\lambda A)B = A(\lambda B)$ for some real number λ.
- $A(B + C) = AB + AC$
- $(A + B)C = AC + BC$

1.5 Determinant

Every square matrix A has an associated number, called its determinant and denoted by $det(A)$ or $|A|$. It is used in :

- solving a system of linear equations,
- finding the inverse of a matrix,
- calculus and more.

1.5.1 First order determinant :

If $A = [a]$, a 1×1 matrix then $det(A) = a$

Proposition 1.5.1. *The following are some important properties of determinant of matrices.*

- *Determinant of an identity matrix (of any order n) is 1. (Verify it for 2×2 and 3×3 matrices first).*
- $det(A^T) = det(A)$, *where A^T denotes the transpose of matrix A.*
- $det(AB) = det(A)det(B)$, *where both A and B are $n \times n$ matrices (Verify it for 2×2 and 3×3 matrices first).*

Proposition 1.5.2. *Determinant of inverse of a matrix*

Let is denote inverse of a matrix A by A^{-1}, then we have the $AA^{-1} = I$. Calculating determinant of both the sides we have, $det(AA^{-1}) = det(I)$, i.e., $det(A).det(A^{-1}) = 1$. Hence we get $det(A^{-1}) = \frac{1}{det(A)}$.

1.5.4 Invariance under elementary row and column operations

Important properties of the determinant include the following, which include invariance under elementary row and column operations.

1) Switching two rows or columns changes the sign.

2) Multiples of rows and columns can be added together without changing the determinant's value.

3) Scalar multiplication of a row by a constant t multiplies the determinant by t.

4) A determinant with a row or column of zeros has value 0. (we can calculate the determinant by expanding with respect to that row or column.)

Recall :

- For a 1×1 matrix $[a]$, the determinant is defined by $det([a]) = a$.

- For an $n \times n$ matrix, the determinant is defined inductively via the minors M_{1j} or cofactors C_{1j} corresponding to the first row.

- The (i, j)-th minor is the determinant of the submatrix formed by deleting the i-th row and j-th column.

- The (i, j)-th cofactor $C_{ij} = (-1)^{i+j} M_{ij}$.

Definition

$det(A) = \Sigma_{j=1}^{n} (-1)^{1+j} a_{1j} M_{1j} = \Sigma_{i=1}^{n} a_{1j} C_{1j}$

1) The determinant of a matrix with a row or column of zeros is 0.
2) The determinant of a matrix in which one row (or column) is a linear combination of other rows (resp. columns) is 0.
3) Scalar multiplication of a row by a constant t multiplies the determinant by t.
4) While computing the determinant, you can choose to compute it using expansion along a suitable row or column.

Finding solution to system of equations using Cramer's rule :

Note : cramer's rule is only applicable for square matrices which are invertible

Cramer's rule for invertible 2×2 matrices

Consider the following system of linear equations of two variables.

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$

Matrix representation : $Ax = b \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$.

Define $A_{x_1} = \begin{bmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{bmatrix}$ and $A_{x_2} = \begin{bmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{bmatrix}$.

The solution of the system of equations in 2 variables is:

$$x_1 = \frac{det(A_{x_1})}{det(A)} = \frac{b_1 a_{22} - a_{12} b_2}{a_{11} a_{22} - a_{12} a_{21}}$$

$$x_2 = \frac{det(A_{x_2})}{det(A)} = \frac{a_{11} b_2 - b_1 a_{21}}{a_{11} a_{22} - a_{12} a_{21}}$$

Define $A_{x_1} = \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix}$

$$A_{x_2} = \begin{bmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{bmatrix} \qquad A_{x_3} = \begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{bmatrix}$$

The solution of the system of equations of 3 variables is:

$$x_1 = \frac{det(A_{x_1})}{det(A)} \qquad x_2 = \frac{det(A_{x_2})}{det(A)} \qquad x_3 = \frac{det(A_{x_3})}{det(A)}$$

Inverse of a square matrix :

Let A be an $n \times n$ matrix. The inverse of A is another $n \times n$ matrix B such that $AB = BA = \mathrm{I}_{n \times n}$ and is denoted by A^{-1}.

Example

$$\begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix} \begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}$$

Recall that $det(A)det(A^{-1}) = 1$ and hence $det(A^{-1}) = \frac{1}{det(A)}$.
Conclusion : inverse of A exists $\quad \Rightarrow \quad det(A)$ has to be non-zero.

The adjugate of a square matrix

Recall that the (i,j)-th minor is the determinant of the submatrix formed by deleting the i-th row and j-th column. Notation : M_{ij}.

The (i,j)-th cofactor is defined as : $C_{ij} := (-1)^{i+j} M_{ij}$.

The cofactor matrix C is the matrix whose (i,j)-th entry is C_{ij}.

Definition

The adjugate matrix of A is defined as : $adj(A) := C^T$.

A 3×3 example of adjugate and inverse

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 8 \\ 5 & 6 & 0 \end{bmatrix}$$

$$\begin{aligned} det(A) &= 1(2 \times 0 - 8 \times 6) - 2(0 \times 0 - 8 \times 5) + 3(0 \times 6 - 2 \times 5) \\ &= -48 + 80 - 30 = 2 \end{aligned}$$

$$\begin{aligned} M_{11} &= -48, & M_{12} &= -40, & M_{13} &= -10 \\ M_{21} &= -18, & M_{22} &= -15, & M_{23} &= -4 \\ M_{31} &= 10, & M_{32} &= 8, & M_{33} &= 2 \end{aligned}$$

The cofactor matrix $\quad C = \begin{bmatrix} -48 & 40 & -10 \\ 18 & -15 & 4 \\ 10 & -8 & 2 \end{bmatrix}$

The adjugate matrix $\quad adj(A) = \begin{bmatrix} -48 & 18 & 10 \\ 40 & -15 & -8 \\ -10 & 4 & 2 \end{bmatrix}.$

Let us compute $A\frac{1}{det(A)}adj(A)$ [redacted] and $\frac{1}{det(A)}adj(A)A$.

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{bmatrix} \begin{bmatrix} -24 & 18 & 5 \\ 20 & -15 & -4 \\ -5 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -24 & 18 & 5 \\ 20 & -15 & -4 \\ -5 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{bmatrix}$$

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

If A is an $n \times n$ matrix and $det(A) \neq 0$, then A^{-1} exists and equals

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

The solution of a system of linear equations with an invertible coefficient matrix

Consider the system of linear equations $Ax = b$ where the coefficient matrix A is an invertible matrix.

Multiplying both sides by A^{-1} we obtain :

$$Ax = b$$
$$A^{-1}Ax = A^{-1}b$$
$$I_n x = A^{-1}b$$
$$x = A^{-1}b.$$

Homogeneous system of linear equations :

A system of linear equations is homogeneous if all of the constant terms are 0 i.e. $b = 0$.

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= 0 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= 0 \\
&\ldots \\
&\ldots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= 0
\end{aligned}
$$

The matrix form of a homogeneous system is $Ax = 0$.

If A is an invertible matrix then multiplying both sides by A^{-1}, we obtain $x = A^{-1}0 = 0$.

A homogeneous system of linear equations with n equations in n unknowns :

- has a unique solution 0 if its coefficient matrix is invertible, i.e. its determinant is non-zero.
- has an infinite number of solutions if its coefficient matrix is not invertible i.e. its determinant is 0.

In a homogeneous system of linear equations, if the number of equations are greater than the number of variables, the system of linear equations has trivial solution.
And if the number of variables are greater than the number of equations, it has infinite solutions including trivial solution.

Matrix representation of system of equations :

Matrix Representation

The matrix representation of this system of linear equations is $Ax = b$ where :

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

A solution is an assignment of values for x so that the equations are satisfied (i.e. hold true).

Echelon and reduced - row echelon form of matrix

(Reduced) Row echelon form

A matrix is in row echelon form if :

- The first non-zero element in each row, called the leading entry, is 1.
- Each leading entry is in a column to the right of the leading entry in the previous row.
- Rows with all zero elements, if any, are below rows having a non-zero element.
- For a non-zero row, the leading entry in the row is the only non-zero entry in its column.

Solutions of system of linear equations using RREF form :

Solutions of $Ax = b$ when A is in reduced row echelon form

Let $Ax = b$ be a system of linear equations and suppose A is in reduced row echelon form.

Suppose for some i, the i^{th} row of A is a zero row but $b_i \neq 0$. Then this system has no solution.

Reason : This means if we write the corresponding system of linear equations, the i^{th} equation reads

$$0x_1 + 0x_2 + \ldots 0x_n = b_i.$$

Since $b_i \neq 0$ this equation cannot be satisfied.

Dependent and Independent variables :

Solutions of $Ax = b$ when A is in reduced row echelon form

Let $Ax = b$ be a system of linear equations and suppose A is in reduced row echelon form.

Assume that for every zero row of A, the corresponding entry of b is also 0 (i.e. if the i^{th} row of A is zero, then so is b_i).

- If the i-th column has the leading entry of some row, we call x_i a dependent variable.
- If the i-th column does not have the leading entry of some row, we call x_i an independent variable.

The intermediate steps while reducing a matrix to **Row Reduced Echelon Form (RREF)** may not always represent solutions to the system Ax=b, but they remain **equivalent systems** that have the same solution set.

Solutions of $Ax = b$ when A is in reduced row echelon form

- Assign arbitrary values to independent variables.
- For a dependent variable, there is a unique equation in which it occurs. All other variables in that equation are independent variables and thus have values assigned. Hence, we can compute the value of the dependent variable from this equation substituting the assigned values for the other independent variables in the equation.
- The obtained values for x_i give a solution to the system.
- In fact every solution is obtained in this way.

Conclusion : If A is in reduced row echelon form, this easy procedure provides us with ALL the solutions of $Ax = b$.

Elementary row operations :

- What are elementary row operations?
- Reducing any matrix to (reduced) row echelon form using elementary row operations.
- Computing the determinant using row reduction.

Elementary Row operations

Type	Action	Example and notation
1	Interchange two rows	$\begin{bmatrix} 3 & 2 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 7 & 1 & 1 \end{bmatrix} \overset{R_1 \longleftrightarrow R_2}{\rightsquigarrow} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 3 & 2 & 1 & 1 \\ 0 & 7 & 1 & 1 \end{bmatrix}$
2	Scalar multiplication of a row by a constant t.	$\begin{bmatrix} 3 & 2 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 7 & 1 & 1 \end{bmatrix} \overset{R_1/3}{\rightsquigarrow} \begin{bmatrix} 1 & 2/3 & 1/3 & 1/3 \\ 1 & 1 & 0 & 0 \\ 0 & 7 & 1 & 1 \end{bmatrix}$
3	Adding multiples of a row to another row.	$\begin{bmatrix} 3 & 2 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 7 & 1 & 1 \end{bmatrix} \overset{R_1 - 3R_2}{\rightsquigarrow} \begin{bmatrix} 0 & -1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 7 & 1 & 1 \end{bmatrix}$

Effect of elementary row operations on determinant :

$$\begin{aligned} det(A) &= 2 \times det \begin{bmatrix} 8 & 7 \\ 6 & 9 \end{bmatrix} - 4 \times det \begin{bmatrix} 3 & 7 \\ 5 & 9 \end{bmatrix} + 1 \times det \begin{bmatrix} 3 & 8 \\ 5 & 6 \end{bmatrix} \\ &= 2(72 - 42) - 4(27 - 35) + 1(18 - 40) \\ &= 2(30) - 4(-8) + 1(-22) \\ &= 60 + 32 - 22 \\ &= 70 \end{aligned}$$

Type	Notation	Effect on determinant
1	$A \overset{R_i \leftrightarrow R_j}{\rightsquigarrow} B$	$det(A) = -det(B)$
2	$A \overset{R_i / c}{\rightsquigarrow} B$	$det(A) = c det(B)$
3	$A \overset{R_i + cR_j}{\rightsquigarrow} B$	$det(A) = det(B)$

Computing the determinant of a matrix using RREF from of that matrix :

Computing the determinant via row reduction

For a square matrix A :

Observe : Row reducing A into row echelon form produces an upper triangular matrix with diagonal entries either all 1 (if it is invertible) or some 1s and some 0s.

1. Row reduce A into row echelon form.
2. If the diagonal entries of the reduced matrix contain a 0, then its determinant is 0 and tracing the determinant back along the row reduction procedure shows that the determinant of A must be 0.
3. If the diagonal entries of the reduced matrix are all 1s its determinant is 1. Tracing back along the procedure used to row reduce using the table of how the determinant changes according to elementary row operations, we can compute the determinant of A.

$$
\begin{tikzcd}
A = \begin{bmatrix} 2 & 4 & 1 \\ 3 & 8 & 7 \\ 5 & 6 & 9 \end{bmatrix} \arrow[r, squiggly, "R_1/2"] & \begin{bmatrix} 1 & 2 & 1/2 \\ 3 & 8 & 7 \\ 5 & 6 & 9 \end{bmatrix} \arrow[r, squiggly, "R_2-3R_1, R_3-5R_1"] & \begin{bmatrix} 1 & 2 & 1/2 \\ 0 & 2 & 11/2 \\ 0 & -4 & 13/2 \end{bmatrix} \arrow[d, squiggly, "R_2/2"] \\
\begin{bmatrix} 1 & 2 & 1/2 \\ 0 & 1 & 11/4 \\ 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 2 & 1/2 \\ 0 & 1 & 11/4 \\ 0 & 0 & 35/2 \end{bmatrix} \arrow[l, squiggly, "2R_3/35"'] & \begin{bmatrix} 1 & 2 & 1/2 \\ 0 & 1 & 11/4 \\ 0 & -4 & 13/2 \end{bmatrix} \arrow[l, squiggly, "R_3+4R_2"']
\end{tikzcd}
$$

The Gaussian Elimination Method :

- The augmented matrix for a system of linear equations.
- The Gaussian elimination method to determine all solutions of a system of linear equations.
- Computing the inverse using Gaussian elimination.

The augmented matrix

Let $Ax = b$ be a system of linear equations where A is an $m \times n$ matrix and b is a $m \times 1$ column vector.

The augmented matrix of this system is defined as the matrix of size $m \times n+1$ whose first n columns are the columns of A and the last column is b.

We denote the augmented matrix by $[A|b]$ and put a vertical line between the first n columns and the last column b while writing it.

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

$$\underbrace{\qquad\qquad}_{A}$$

$$\begin{aligned} 3x_1 + 2x_2 + x_3 + x_4 &= 6 \\ x_1 + x_2 &= 2 \\ 7x_2 + x_3 + x_4 &= 8 \end{aligned}$$

where $A = \begin{bmatrix} 3 & 2 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 7 & 1 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 6 \\ 2 \\ 8 \end{bmatrix}$.

The augmented matrix is $[A|b] = \left[\begin{array}{cccc|c} 3 & 2 & 1 & 1 & 6 \\ 1 & 1 & 0 & 0 & 2 \\ 0 & 7 & 1 & 1 & 8 \end{array}\right]$.

The Gaussian elimination method

Consider the system of linear equations $Ax = b$.

1. Form the augmented matrix of the system $[A|b]$.
2. Perform the same operations on $[A|b]$ that were used to bring A into reduced row echelon form.
3. Let R be the submatrix of the obtained matrix of the first n columns and c be the submatrix of the obtained matrix consisting of the last column.

We write the obtained matrix as $[R|c]$. Notice that R is the reduced row echelon matrix obtained by row reducing A.

The solutions of $Ax = b$ are precisely the solutions of $Rx = c$.

4. Form the corresponding system of linear equations $Rx = c$.
5. Find ALL the solutions of $Rx = c$ and hence of $Ax = b$.

Since R is in reduced row echelon form, we can find ALL its solutions (as described earlier).

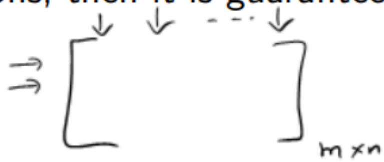
0 is always a solution of a homogeneous system of linear equations $Ax = 0$. This solution is called the *trivial solution*.

For a homogeneous system, there are only two different possibilities :

- 0 is the unique solution.

- there are infinitely many solutions other than 0.

In a homogeneous system of equations, if there are more variables than equations, then it is guaranteed to have nontrivial solutions.



Computing the inverse using the gaussian elimination method :

Computing the inverse

Computing the inverse of an invertible matrix A is equivalent to :

Finding solutions of $Ax = \begin{bmatrix}1\\0\\0\end{bmatrix}$, $Ay = \begin{bmatrix}0\\1\\0\end{bmatrix}$ and $Az = \begin{bmatrix}0\\0\\1\end{bmatrix}$.

Vector Spaces :

- Recall addition of these vectors is defined as :

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

- and scalar multiplication is defined as :

$$c(x_1, x_2, \ldots, x_n) = (cx_1, cx_2, \ldots, cx_n)$$

Properties of addition and scalar multiplication

Let v, w and v' be vectors in $\mathbb{R}^n$ and $a, b \in \mathbb{R}$.

i) $v + w = w + v$.
ii) $(v + w) + v' = v + (w + v')$.
iii) The 0 vector satisfies that $v + 0 = 0 + v = v$.
iv) The vector $-v$ satisfies that $v + (-v) = 0$.
v) $1v = v$.
vi) $(ab)v = a(bv)$.
vii) $a(v + w) = av + aw$.
viii) $(a + b)v = av + bv$

A vector space is a set with two operations (called **addition** and **scalar multiplication** with the above properties (i)-(viii).

Defintion of a vector space

A vector space V over $\mathbb{R}$ is a set along with two functions

$$+ : V \times V \to V \qquad \text{and} \qquad \cdot : \mathbb{R} \times V \to V$$

(i.e. for each pair of elements v_1 and v_2 in V, there is a unique element $v_1 + v_2$ in V, and for each $c \in \mathbb{R}$ and $v \in V$ there is a unique element $c \cdot v$ in V)

It is standard to suppress the $\cdot$ and only write cv instead of $c.v$.

The functions $+$ and $\cdot$ are required to satisfy the following rules :

Formal definition (Contd.)

i) $v_1 + v_2 = v_2 + v_1$ for all $v_1, v_2 \in V$

ii) $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$ for all $v_1, v_2, v_3 \in V$

iii) There exists an element in V denoted by 0 such that $v + 0 = v$ for all $v \in V$

iv) For each element $v \in V$ there exists an element $v' \in V$ such that $v + v' = 0$

v) For each element $v \in V$, $1v = v$

vi) For each pair of elements $a, b \in \mathbb{R}$ and each element $v \in V$, $(ab)v = a(bv)$

vii) For each element $a \in \mathbb{R}$ and each pair of elements v_1 and v_2, $a(v_1 + v_2) = av_1 + av_2$

viii) For each pair of elements $a, b \in \mathbb{R}$ and each element $v \in V$, $(a + b)v = av + bv$

Example : Matrices

Let $M_{m\times n}(\mathbb{R})$ be the set of all $m \times n$ matrices with real numbers.

Recall that we have defined addition and scalar multiplication on $M_{m\times n}(\mathbb{R})$ as follows :

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

where $A, B \in M_{m\times n}(\mathbb{R})$ and $c \in \mathbb{R}$.

Then $M_{m\times n}(\mathbb{R})$ along with addition and scalar multiplication forms a vector space.

Example : Solutions of a homogeneous system

Consider the set of solutions V of of a homogeneous system $Ax = 0$ where $A \in M_{m\times n}(\mathbb{R})$ (i.e. this is a homogeneous system of m linear equations in n variables).

Note that if $v, w \in V$ then

$$A(v + w) = Av + Aw = 0 + 0 = 0. \quad \Rightarrow v+w \in V$$

and if $c \in \mathbb{R}$ then

$$A(cv) = c(Av) = c(0) = 0. \quad \Rightarrow cv \in V$$

So addition and scalar multiplication on $\mathbb{R}^n$ restricts to the solution set. Hence it is a vector space.

This is an example of a subspace of a vector space.

Cancellation law of vector addition

If $v_1, v_2, v_3 \in V$ such that $v_1 + v_3 = v_2 + v_3$, then $v_1 = v_2$.

Corollaries:

- The vector 0 described in (iii) is unique.

  Suppose $\exists w \in V$ s.t. $v + w = v \ \forall v \in V$.
  $\cancel{v} + w = \cancel{v} + 0 \Rightarrow w = 0$.

- The vector v' described in (iv) is unique and it is standard to refer to it as $-v$.

  $v + v' = 0$ Suppose v'' also satisfies this.
  Then $v + v' = 0 = v + v''$ Cancel v $\therefore v' = v''$.

Some more important properties

In any vector space V the following statements are true.

- $0v = 0$ for each $v \in V$.

  $(0+0)v = 0v + 0v$, and $(0+0)v = 0v$ $\Big\}$ $0v = 0v + 0v$
  $\Rightarrow 0v + 0 = 0v + 0v$
  Cancel $0v \Rightarrow 0 = 0v$.

- $(-c)v = -(cv) = c(-v)$ for each $c \in \mathbb{R}$ and for each $v \in V$.

  $(c + (-c))v = cv + (-c)v$, and $(c+(-c))v = 0v = 0$ $\Rightarrow cv + (-c)v = 0$.
  $\Rightarrow (-c)v = -(cv)$.

- $c0 = 0$ for each $c \in \mathbb{R}$.

  Check this!

Linear Dependence and Linear Independence of set of vectors :

Linear combination of vectors :

Let V be a vector space and $v_1, v_2, \ldots, v_n \in V$. The linear combination of $v_1, v_2, \ldots, v_n$ with coefficients $a_1, a_2, \ldots, a_n \in \mathbb{R}$ is the vector $\sum_{i=1}^{n} a_i v_i \in V$.

A vector $v \in V$ is a linear combination of $v_1, v_2, \ldots, v_n$ if there exist some $a_1, a_2, \ldots, a_n \in \mathbb{R}$ so that $v = \sum_{i=1}^{n} a_i v_i$.

In the previous example we see that $(2, 9)$ is a linear combination of vectors $(2, 1)$ and $(-2, 3)$, as follows:

$$3(2,1) + 2(-2,3) = (6,3) + (-4,6) = (2,9)$$

Moreover, each of the vectors in the expression is a linear combination of the other two vectors.

$$\tfrac{1}{3}(2,9) - \tfrac{2}{3}(-2,3) = (2,1)$$
$$\tfrac{1}{2}(2,9) - \tfrac{3}{2}(2,1) = (-2,3)$$

Note further that we can re-write these expressions as follows :

$$3(2,1) + 2(-2,3) - (2,9) = (0,0)$$

Observe : the 0 vector is a linear combination of $(2,1), (-2,3), (2,9)$ with non-zero coefficients.

If possible let us assume we can write $(1, 2, 0)$ as a linear combination of the other two vectors as follows,

$$a(0, 2, 1) + b(2, 2, 0) = (1, 2, 0)$$

which implies, $2b = 1$, $2a + 2b = 2$ and $a = 0$. Clearly these three equations simultaneously cannot have a solution. Hence our assumption was false.

We can use the discussion above to conclude that

$$a(0, 2, 1) + b(2, 2, 0) + c(1, 2, 0) = (0, 0, 0) \text{ if and only if } a = b = c = 0.$$

i.e. the only way the 0 vector is a linear combination of $(0, 2, 1), (2, 2, 0), (1, 2, 0)$ is if the coefficients are 0.

Definition of linear dependence :

A set of vectors $v_1, v_2, \ldots, v_n$ from a vector space V is said to be linearly dependent, if there exist scalars $a_1, a_2, \ldots, a_n$, not all zero, such that

$$a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0$$

Equivalently, the 0 vector is a linear combination of $v_1, v_2, \ldots, v_n$ with non-zero coefficients.

Important Remark :

If a set is linearly dependent, then so is every superset of it.

Definition of linear independence

A set of vectors $v_1, v_2, \ldots, v_n$ from a vector space V is said to be linearly independent if $v_1, v_2, \ldots, v_n$ are not linearly dependent.

Equivalently : A set of vectors $v_1, v_2, \ldots, v_n$ from a vector space V is said to be linearly independent, if the equation

$$a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0$$

can only be satisfied when $a_i = 0$ for all $i = 1, 2, \ldots, n$.

Equivalently : A set of vectors $v_1, v_2, \ldots, v_n$ from a vector space V is said to be linearly independent if the only linear combination of $v_1, v_2, \ldots, v_n$ which equals 0 is the linear combination with all coefficients 0.

Let $v_1, v_2, \ldots, v_n$ be a set of vectors containing the 0 vector.

Suppose $v_i = 0$. Then we can choose $a_i = 1$ and $a_j = 0$ for $j \neq i$.

Then the linear combination $a_1 v_1 + a_2 v_2 + \ldots + a_n v_n$ is 0 but not all coefficients are 0.

Hence, a set of vectors $v_1, v_2, \ldots, v_n$ containing the 0 vector is always a linearly dependent set.

Conclusion : Two non-zero vectors are linearly independent precisely when they are not multiples of each other .

Conclusion : If three vectors are linearly independent then none of these vectors is a linear combination of the other two.

How to check linear independence in $\mathbb{R}^m$

How do we check if $v_1, v_2, \ldots, v_n \in \mathbb{R}^m$ are linearly independent?

In terms of coordinates, let $v_j = (v_{1j}, v_{2j}, \ldots, v_{mj})$; $j = 1, 2, \ldots, n$.

Let us write the linear combination of these vectors with *arbitrary* coefficients $a_1, a_2, \ldots, a_n$ and equate it to 0 :

$$a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0.$$

Considering each coordinate, we have the following identities :

$$\begin{aligned}
v_{11}a_1 + v_{12}a_2 + \ldots + v_{1n}a_n &= 0 \\
v_{21}a_1 + v_{22}a_2 + \ldots + v_{2n}a_n &= 0 \\
\vdots \qquad \vdots \qquad \ddots \qquad \vdots \quad &\ \ \vdots \\
v_{m1}a_1 + v_{m2}a_2 + \ldots + v_{mn}a_n &= 0
\end{aligned}$$

Since the a_i are arbitrary (unknown), we can treat this like a homogeneous system of linear equations with coefficients v_{ij} and unknowns a_i.

For linear independence, we have to check if the only choice of a_i's satisfying the above identities is $a_i = 0$.

Equivalently, in terms of the homogeneous system of linear equations, we have to check that its only solution is the 0 solution.

Conclusion : To check $v_1, v_2, \ldots, v_n \in \mathbb{R}^m$ are linearly independent, we have to check that the homogeneous system of linear equations $Vx = 0$ has only the trivial solution, where the j^{th} column of V is v_j.

Example : 3×3

Consider the three vectors $(1,2,0)$, $(0,2,4)$ and $(3,0,0)$ in $\mathbb{R}^3$. Equate the linear combination of these three vectors with *unknown* coefficients x_1, x_2 and x_3 to 0 :

$$x_1(1,2,0) + x_2(0,2,4) + x_3(3,0,0) = (0,0,0).$$

Hence we have the system of linear equations:

$$\begin{aligned} x_1 + 0x_2 + 3x_3 &= 0 \\ 2x_1 + 2x_2 + 0x_3 &= 0 \\ 0x_1 + 4x_2 + 0x_3 &= 0 \end{aligned}$$

Since the matrix $\begin{bmatrix} 1 & 0 & 3 \\ 2 & 2 & 0 \\ 0 & 4 & 0 \end{bmatrix}$ is invertible, the system of linear equations has a unique solution $x_1 = x_2 = x_3 = 0$. **Hence the vectors** $(1,2,0)$**,** $(0,2,4)$ **and** $(3,0,0)$ **are linearly independent**.

More than 2 vectors in $\mathbb{R}^2$

Suppose we have n vectors in $\mathbb{R}^2$ where $n \geq 3$. To check linear indepence, we have to check whether the corresponding homogeneous linear system $Vx = 0$ has the unique solution $x = 0$.

Since $n \geq 3 > 2$, this is a homogeneous system with more unknowns (n) than equations (2).

We have seen in the previous week that Gaussian elimination will yield infinitely many solutions.

Hence, any set of n **vectors in** $\mathbb{R}^2$ **with** $n \geq 3$ **are linearly dependent**.

More than n vectors in $\mathbb{R}^n$

The same argument as for $\mathbb{R}^2$ in the previous slide yields :

Hence, any set of r vectors in $\mathbb{R}^n$ with $r > n$ are linearly dependent.

Relationship with determinant

To check whether a set of n vectors in $\mathbb{R}^n$ are linearly independent, we have to find the solutions of the homogeneous system $Vx = 0$ where V is an $n \times n$ matrix obtained by arranging the vectors in columns.

Since V is a square matrix, it has unique solution $x = 0$ if and only if $\cancel{A}$ V is invertible if and only if $det(\cancel{A}\ V) \neq 0$.

- If A is invertible then there exists A^{-1} such that $AA^{-1} = 1 = A^{-1}A$. Hence $det(A).det(A^{-1}) = 1$ which implies $det(A) \neq 0$.
- Now if $det(A) \neq 0$ then $A^{-1} = \frac{1}{det(A)} adj(A)$ exists.

Let us consider the vectors $(1, 4, 2)$,$(0, 4, 3)$ and $(1, 1, 0)$ in $\mathbb{R}^3$.

We obtain the matrix

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

Since $det(\cancel{A}\ V) = 1 \neq 0$, the matrix $\cancel{A}$ V is invertible and hence **the vectors $(1, 4, 2)$,$(0, 4, 3)$ and $(1, 1, 0)$ are linearly independent**.

Basis for a Vector Space :

Span of a set of vectors :

Span of a set of vectors

The span of a set S (of vectors) is defined as the set of all finite linear combinations of elements(vectors) of S, and denoted by $Span(S)$.

$$\text{i.e. } Span(S) = \{\sum_{i=1}^{n} a_i v_i \in V | a_1, a_2, \ldots, a_n \in \mathbb{R}\}$$

Example

Let $S = \{(1,0)\} \subset \mathbb{R}^2$. Then

$$Span(S) = \{a(1,0) | a \in \mathbb{R}\} = \{(a,0) | a \in \mathbb{R}.\}$$

Thus, $Span(S)$ is the X-axis in $\mathbb{R}^2$.

Spanning Set :

Spanning set for a vector space

Let V be a vector space. A set $S \subseteq V$ is a spanning set for V if $Span(S) = V$.

Example

- If $S = \{(1,0),(0,1)\}$ then $Span(S) = \mathbb{R}^2$
- If $S = \{(1,0),(0,1),(1,2)\}$ then $Span(S) = \mathbb{R}^2$
- If $S = \{(1,1),(0,1)\}$ then $Span(S) = \mathbb{R}^2$

Adding vectors to obtain a spanning set for $\mathbb{R}^3$

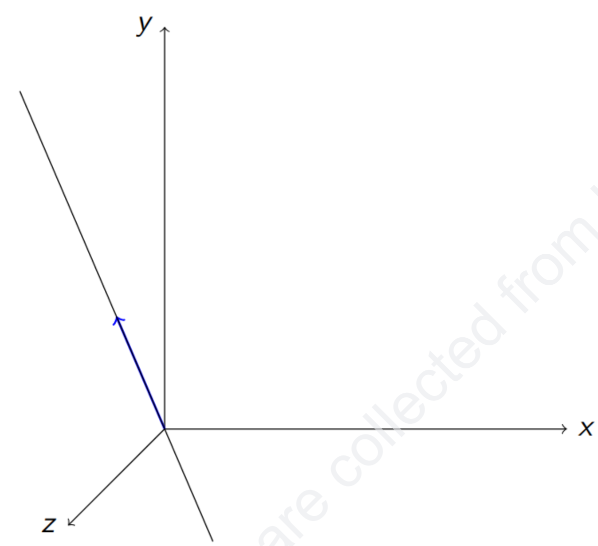
We will try to "build" a spanning set for the vector space $\mathbb{R}^3$.

Start with S_0 to be the empty set $\emptyset$. Then $Span(S_0) = Span(\emptyset) = \{(0,0,0)\}$.

Since this is not the full vector space, append a vector outside $Span(S_0)$ in $\mathbb{R}^3$ e.g. $(0,2,1)$ to S_0 and call the new set S_1.

So $S_1 = S_0 \cup \{(0,2,1)\}$.

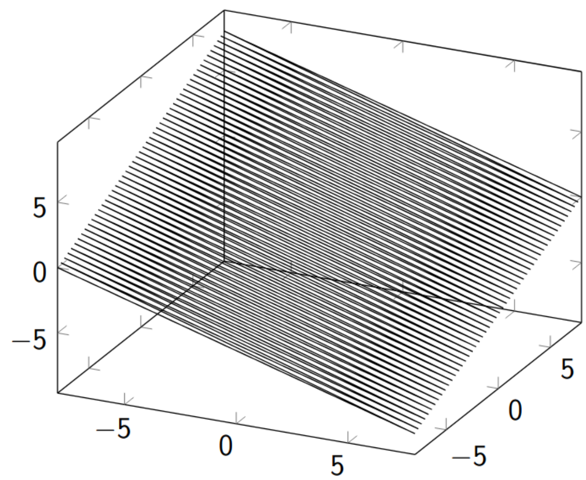
$Span(S_1)$ is the line shown in the picture below.



Choose a vector outside $Span(S_1)$ e.g. $(2,2,0)$, append it to S_1 and call the new set S_2.

So $S_2 = S_1 \cup \{(2,2,0)\}$.

$Span(S_2)$ is the plane shown in the picture.



Again choose a vector outside $Span(S_2)$, e.g. $(0, 0, 5)$, append it to S_2 and call the new set S_3.

So $S_3 = S_2 \cup \{(0, 0, 5)\}$.

Any arbitrary vector $(x, y, z) \in \mathbb{R}^3$ can be written as follows:

$$(x, y, z) = \frac{y-x}{2}(0, 2, 1) + \frac{x}{2}(2, 2, 0) + \frac{x-y+2z}{10}(0, 0, 5)$$

Hence

$$Span(S_3) = \mathbb{R}^3$$

What is a basis?

A basis B of a vector space V is a linearly independent subset of V that spans V.

Example

Let $e_i \in \mathbb{R}^n$ be the vector with i^{th} coordinate 1 and all other coordinates 0 e.g. $e_1 = (1, 0, 0, \ldots, 0)$.

The set $\varepsilon = \{e_1, e_2, \ldots, e_n\} \subseteq \mathbb{R}^n$ is a basis for $\mathbb{R}^n$. ~~consisting of~~

Equivalent conditions for B to be a basis

The following conditions are equivalent to a subset $B \subseteq V$ being a basis :

i) B is linearly independent and $Span(B) = V$.

ii) B is a maximal linearly independent set.

iii) B is a minimal spanning set.

Suppose B is a basis.
$\therefore$ B is lin. indept.
Suppose $B' = B \cup \{v\}$.
$\therefore$ $v = \sum_{i=1}^{n} a_i v_i$ where $v_1, \dots, v_n \in B$.
$\therefore$ B' is a lin. dep. set.

maximal lin. indept. means
① it is lin. indept.
② appending any vector makes it lin. dep.

minimal spanning means
① it is spanning
② it is no longer spanning if we delete any vector

How do we find a basis?

We can find a basis by any one of the methods described below :

i) Start with the $\emptyset$ and keep appending vectors which are not in the span of the set thus far obtained, until we obtain a spanning set.

ii) Take a spanning set and keep deleting vectors which are linear combinations of the other vectors, until the remaining vectors satisfy that they are not a linear combination of the other remaining ones.

Example : Method 1 : $V = \mathbb{R}^2$

Let us start with the empty set and append a non-zero vector e.g. $(1,2)$.

Now choose another vector which is not in the span of the earlier vector e.g. $(2,3)$.

$Span(\{(1,2),(2,3)\}) = \mathbb{R}^2$.

Hence this set forms a basis for $\mathbb{R}^2$.

Example : Method 2 : $V = \mathbb{R}^3$

Let us start with the set

$$S = \{(1,0,0),(1,2,0),(1,0,3),(0,2,3),(0,4,2)\}$$

Check that $Span(S) = \mathbb{R}^3$.

Now observe that, $(0,4,2) = 2(1,2,0) + \frac{2}{3}(1,0,3) - \frac{8}{3}(1,0,0)$.

So delete $(0,4,2)$.

Hence our new set of vectors is

$$S_1 = \{(1,0,0),(1,2,0),(1,0,3),(0,2,3)\}$$

Observe that $(0,2,3) = (1,2,0) + (1,0,3) - 2(1,0,0)$.

Hence delete $(0,2,3)$.

Hence our new set of vectors is

$$S_2 = \{(1,0,0),(1,1,0),(1,0,1)\}$$

None of these vectors is a linear combination of the other two vectors.

Hence S_2 forms a basis of R^3.

Rank / dimension of a vector space :

The dimension (or rank) of a vector space is the **size (or cardinality) of a basis of the vector space**.

for this course : if B is a basis of V, then the rank is the number of elements in B.

For every vector space there exists a basis,and all bases of a vector space have the same number of elements (or cardinality) ; hence, the dimension (or rank) of a vector space (say V) is uniquely defined and denoted by $dim(V)$ (or $rank(V)$) respectively.

Dimension of $\mathbb{R}^n$

Recall the i^{th} **standard basis vector** in $\mathbb{R}^n$.

$$e_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$$

i.e. the i-th co-ordinate is 1 and 0 elsewhere.

Recall that the set $\{e_1, e_2, \ldots, e_n\}$ is a basis of $\mathbb{R}^n$ called the standard basis.

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

Example :

Let use calculate the dimension of the subspace W of $\mathbb{R}^3$ spanned by $\{(1, 0, 0), (0, 1, 0), (3, 5, 0)\}$.

Observe that, $3(1, 0, 0) + 5(0, 1, 0) = (3, 5, 0)$.

Hence the set is not linearly independent.

Hence we delete the vector $(3,5,0)$ from this set.

The remaining two vectors form a linearly independent set.

Hence the set $\{(1,0,0),(0,1,0)\}$ forms a basis of the subspace W spanned by $\{(1,0,0),(0,1,0),(3,5,0)\}$.

Hence dimension of the subspace W is 2.

Example : in terms of matrices

Write the vectors which span (or generate) W as rows of a matrix :

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 5 & 0 \end{bmatrix}.$$

Apply row reduction to this matrix.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 5 & 0 \end{bmatrix} \xrightarrow{R_3-3R_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 0 \end{bmatrix} \xrightarrow{R_3-5R_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The final matrix is the row echelon form of the original matrix and its rows form a basis of the subspace W.

In particular, the number of non-zero rows is $2 = dim(W)$.

After writing the given set of vectors as rows of a matrix, convert the matrix to Row- echelon form.
Then the number of non zero rows in row echelon form is same as dimension of span of original vectors. Dim (span (vectors)).

Rank of a matrix

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

- The **column space** of A is the subspace of $\mathbb{R}^m$ spanned by the column vectors of A.
- The **row space** of A is the subspace of $\mathbb{R}^n$ spanned by the row vectors of A.
- The dimension of the column space of A is defined as the **column rank** of A.
- The dimension of row space of A is defined as the **row rank** of A.

Fact : **Column rank= Row rank** and this number is called the **rank** of A.

Let us find the rank of the matrix $A = \begin{bmatrix} 1 & 0 & 1 \\ -2 & -3 & 1 \\ 3 & 3 & 0 \end{bmatrix}$.

Reduce it to row echelon form:

$$\begin{bmatrix} 1 & 0 & 1 \\ -2 & -3 & 1 \\ 3 & 3 & 0 \end{bmatrix} \xrightarrow{R_2+2R_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 3 \\ 3 & 3 & 0 \end{bmatrix} \xrightarrow{R_3-3R_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 3 \\ 0 & 3 & -3 \end{bmatrix} \xrightarrow{-R_2/3} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 3 & -3 \end{bmatrix} \xrightarrow{R_3-3R_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

There are two non-zero rows. Hence $rank(A) = 2$.

Rank and dimension using Gaussian elimination :

Finding dimension and basis with a given spanning set

Consider a vector space W spanned by a set S.

e.g. let us consider the vector space W spanned by the set $S = \{(1,0,1),(-2,-3,1),(3,3,0)\}$.

We will use the following steps to find the dimension and a basis for W and carry out the steps for our example.

- Form a matrix with the vectors in the spanning set as the rows.

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

- Reduce to a matrix in the row echelon form.

$$\begin{bmatrix} 1 & 0 & 1 \\ -2 & -3 & 1 \\ 3 & 3 & 0 \end{bmatrix} \xrightarrow{R_2+2R_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 3 \\ 3 & 3 & 0 \end{bmatrix} \xrightarrow{R_3-3R_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 3 \\ 0 & 3 & -3 \end{bmatrix} \xrightarrow{-R_2/3} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 3 & -3 \end{bmatrix} \xrightarrow{R_3-3R_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

- The number of non-zero rows is the dimension of the vector space W.

- The vectors corresponding to the non-zero rows form the basis of the vector space W.

In the example, the final matrix is $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$.

Hence, dimension of the vector space spanned by $\{(1,0,1),(-2,-3,1),(3,3,0)\}$ is 2 and a basis is given by $(1,0,1),(0,1,-1)$.

The null space of a matrix finding nullity and a basis for the null space :

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

The subspace $W = \{x \in \mathbb{R}^n | Ax = 0\}$ of $\mathbb{R}^n$ is called the solution space of the homogeneous system of linear equation $Ax = 0$ or the null space of A.

Note that the null space is a subspace of $\mathbb{R}^n$. The dimension of the null space is called the nullity of A.

Finding the nullity and a basis for the null space

We have seen how to find the dimension and a basis for the row space of A using row reduction.

We will use row reduction to also find the nullity and a basis for the null space of A.

Recall first how to find the solution space for a system $Ax = b$ i.e. Gaussian elimination.

- Form the augmented matrix $[A|b]$
- Apply the same row reduction operations on the augmented matrix that are used to row reduce A to obtain the augmented matrix $[R|c]$ where R is the matrix in reduced row echelon form obtained from A.
- If the i-th column has the leading entry of some row, we call x_i a dependent variable.
- If the i-th column does not have the leading entry of some row, we call x_i an independent variable.

$$\text{nullity}(A) = \text{number of independent variables .}$$

- Assign arbitrary value t_i to the i^{th} independent variable.
- Compute the value of each dependent variables in terms of t_is from the unique row it occurs in.
- Every solution is obtained by letting t_is vary in $\mathbb{R}$.

The vectors obtained by substituting $t_i = 1$ and $t_j = 0 \forall j \neq i$ as i varies constitutes a basis of the null space of A (i.e. the solution space of $Ax = 0$).

Example : 3×3 matrix

Consider the (matrix representation of the) homogeneous system of linear equations of the form $Ax = 0$, where $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}$

The augmented matrix is $\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 2 & 2 & 2 & 0 \\ 3 & 3 & 3 & 0 \end{array}\right]$.

Row reduction yields :

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 2 & 2 & 2 & 0 \\ 3 & 3 & 3 & 0 \end{array}\right] \underset{R_2-2R_1}{\overset{R_3-3R_1}{\rightsquigarrow}} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Independent variables : x_2, x_3, dependent variable : x_1.

Hence, $nullity(A) = 2$.

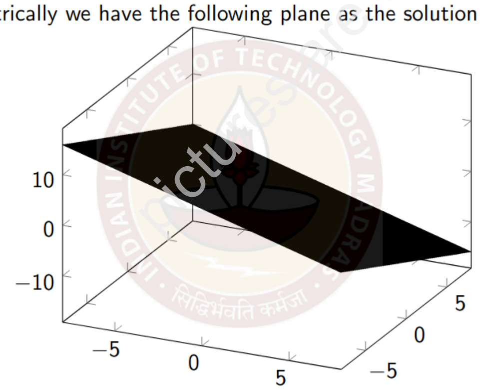
Put $x_2 = t_1$ and $x_3 = t_2$. Then the equation yields
$x_1 = -x_2 - x_3 = -t_1 - t_2$.

Hence, the null space of A (i.e. the solution space of $Ax = 0$) is
$\{(-t_1 - t_2, t_1, t_2) | t_1, t_2 \in \mathbb{R}\}$.

$t_1 = 1, t_2 = 0$ yields the basis vector $(-1, 1, 0)$.
$t_1 = 0, t_2 = 1$ yields the basis vector $(-1, 0, 1)$.

Hence, a basis for the null space is $(-1, 1, 0), (-1, 0, 1)$.

Geometrically we have the following plane as the solution space :



The rank-nullity theorem

Theorem

For an $m \times n$ matrix A, $rank(A) + nullity(A) = n$.

How to check if a set of n vectors is a basis for $\mathbb{R}^n$

Short answer : Use determinants.

Suppose we are given n vectors of $\mathbb{R}^n$.

We write them as columns of a matrix, thus obtaining an $n \times n$ (square) matrix.

If the determinant of the matrix is 0, then the given set of vectors does not form a basis, otherwise it forms a basis.

Examples :
The standard basis $(1,0),(0,1)$ yields the identity matrix I with determinant 1.
The vectors $(1,-2),(5,-10)$ yields the matrix $\begin{bmatrix} 1 & 5 \\ -2 & -10 \end{bmatrix}$ with determinant 0. This is not a basis for $\mathbb{R}^2$.

<u>linear mapping :</u>

What is a linear mapping

A linear mapping f from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be defined as follows :

$$f(x_1, x_2, \ldots, x_n) = \left(\sum_{j=1}^{n} a_{1j}x_j, \sum_{j=1}^{n} a_{2j}x_j, \ldots, \sum_{j=1}^{n} a_{mj}x_j \right).$$

where the coefficients a_{ij}s are real numbers (scalars). A linear mapping can be thought of as a collection of linear combinations.

We can write the expressions on the RHS in matrix form as Ax

where $A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$ and $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

Linearity of linear mappings

It follows that a linear mapping satisfies linearity, i.e. for any $c \in \mathbb{R}$ (scalar)

$$f(x_1 + cy_1, x_2 + cy_2, \dots, x_n + cy_n) = f(x_1, x_2, \dots, x_n) + cf(y_1, y_2, \dots, y_n).$$

$$f(x_1+cy_1, x_2+cy_2, \dots, x_n+cy_n) = A\begin{bmatrix} x_1+cy_1 \\ x_2+cy_2 \\ \vdots \\ x_n+cy_n \end{bmatrix}$$

$$= A\left\{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + c\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}\right\} = A\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + cA\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$= f(x_1, \dots, x_n) + cf(y_1, \dots, y_n).$$

Linear transformation :

Formal definition

A function $f : V \to W$ between two vector spaces V and W is said to be a linear transformation if for any two vectors v_1 and v_2 in the vector space V and for any $c \in \mathbb{R}$ (scalar) the following conditions hold :

- $f(v_1 + v_2) = f(v_1) + f(v_2)$ ✓
- $f(cv_1) = cf(v_1)$ ✓

Equivalent to linearity: $f(v_1 + cv_2) = f(v_1) + cf(v_2)$.

$f(v_1 + cv_2) = f(v_1) + \underline{\underline{f(cv_2)}} = f(v_1) + cf(v_2)$

Linear mappings are linear transformation.

Examples :

1. $f : \mathbb{R}^2 \to \mathbb{R}^2$ $\quad f(x, y) = (2x, y)$
2. $f : \mathbb{R}^2 \to \mathbb{R}^2$ $\quad f(x, y) = (2x, 0)$
3. $f : \mathbb{R}^3 \to \mathbb{R}^3$ $\quad f(x, y, z) = (\frac{x}{2}, 3y, 5z)$
4. $f : \mathbb{R}^3 \to \mathbb{R}^4$ $\quad f(x, y, z) = (4y - z, 3y + \frac{11}{19}z, 5x - 2z, 23y)$
5. $f : \mathbb{R} \to \mathbb{R}^3$ $\quad f(t) = (t, 3t, \frac{23}{89}t)$
6. $f : \mathbb{R}^2 \to \mathbb{R}$ $\quad f(x, y) = x.$

1-1 and onto functions

Recall that a function $f : V \to W$ is 1-1 (or injective) if $f(v_1) = f(v_2)$ implies $v_1 = v_2$.

Recall that a function $f : V \to W$ is onto (or surjective) if for every $w \in W$ there exists $v \in V$ such that $f(v) = w$.

For a linear transformation, being 1-1 is equivalent to $f(v) = 0$ implies $v = 0$.

$f : V \to W$ is a lin. trans.

Assume f is 1-1. Then $f(v_1) = f(v_2) \Rightarrow v_1 = v_2$.

$f(0) = 0$. If $f(v) = 0 \Rightarrow f(v) = f(0)$
$\Rightarrow v = 0$.

$f(0+0) = f(0) + f(0)$
$\Rightarrow f(0) = f(0) + f(0)$
$\Rightarrow f(0) = 0$.

Conversely, assume $f(v) = 0 \Rightarrow v = 0$.

$f(v_1) = f(v_2) \Rightarrow f(v_1 - v_2) = 0.$
$\Rightarrow v_1 - v_2 = 0$
$\Rightarrow v_1 = v_2$.

$f(v_1) - f(v_2) = 0.$
$f(v_1) + f(-v_2) = 0$
$f(v_1 - v_2) = 0.$

$f(v) + f(-v) = f(v + (-v)) = f(0) = 0$
$\Rightarrow f(-v) = -f(v)$

What is an isomorphism

Recall that a function $f : V \to W$ is bijective (or a bijection) if it is 1-1 and onto.

Note that being a bijection is equivalent to : for any $w \in W$ there exists a unique $v \in V$ such that $f(v) = w$.

A linear transformation $f : V \to W$ between two vector spaces V and W is said to be an isomorphism if it is a bijection.

Example 1 seen earlier is an isomorphism :

$$f : \mathbb{R}^2 \to \mathbb{R}^2 \qquad f(x, y) = (2x, y).$$

$$\left.\begin{aligned} f(x,y) = (0,0) \implies (2x,y) = (0,0) &\implies 2x = 0, y = 0 \\ &\implies x = 0, y = 0 \\ &\implies (x,y) = (0,0) \end{aligned}\right\} f \text{ is } 1\text{-}1$$

$$\left.\begin{aligned} &\text{For } (u,v) \in \mathbb{R}^2 \text{ consider } x = u/2,\ y = v. \\ &\therefore f(x,y) = (2x,y) = (2 \cdot u/2, v) = (u,v) \end{aligned}\right\} \text{onto}$$

Non Examples :

Example 2 seen earlier is not an isomorphism :

$$f : \mathbb{R}^2 \to \mathbb{R}^2 \qquad f(x, y) = (2x, 0).$$

There is no pre-image for the vector (u, v), where v is non-zero. e.g. $(0, 1)$ has no pre-image. So f is not surjective. Also $f(x, y) = (0, 0)$ implies $(2x, 0) = (0, 0)$, hence $x = 0$. But there is no restriction on y, e.g. $f(0, 1) = (0, 0)$. Hence f is not 1-1 either.

Similarly, the fifth example $f : \mathbb{R} \to \mathbb{R}^3 \quad ; \quad f(t) = (t, 3t, \frac{23}{89}t)$ is 1-1 but not onto.

Also the sixth example $f : \mathbb{R}^2 \to \mathbb{R} \quad ; \quad f(x, y) = x$ is onto but not 1-1.

Bases determine linear transformations

Let V be a vector space with basis $\{v_1, v_2, \ldots, v_n\}$.
Let $f : V \to W$ be a linear transformation. Then the ordered vectors $f(v_1), f(v_2), \ldots f(v_n)$ uniquely determine f.

Let $v \in V$. $v = \sum_{i=1}^{n} c_i v_i$.

$$\boxed{f(v) = f\left(\sum_{i=1}^{n} c_i v_i\right) = \sum_{i=1}^{n} c_i f(v_i)}$$

is determined by $c_1, \ldots, c_n$ & $f(v_1), f(v_2), \ldots, f(v_n)$.

Suppose $w_1, \ldots, w_n$ is a specified set of vectors in W

There is a unique lin. trans. f s.t. $f(v_i) = w_i$.

Define $f(v) = \sum_{i=1}^{n} c_i w_i$ where $v = \sum_{i=1}^{n} c_i v_i$.

$f(v_k) = w_k$.

Example

Consider the standard basis $\{(1,0),(0,1)\}$ of $\mathbb{R}^2$. What linear transformation $f : \mathbb{R}^2 \to \mathbb{R}^2$ do we obtain by extending :

$$f((1,0)) = (2,0) = 2(1,0) \quad w_1$$
$$f((0,1)) = (0,1) \quad w_2$$

$(x, y) = x(1,0) + y(0,1)$.

$f(x,y) = x(2,0) + y(0,1) = (2x,0) + (0,y)$.

$= (2x, y)$.

Example : changing the basis

Note that if we choose a different basis for $\mathbb{R}^2$, then we may get a different linear transformation. In the previous example, consider the basis $\{(1,0),(1,1)\}$ instead of the standard basis for $\mathbb{R}^2$. Let us calculate the linear transformation f that we obtain by extending :

$$f((1,0)) = (2,0) = 2(1,0) \quad w_1$$
$$f((1,1)) = (0,1) \quad w_2$$

Note that every element (x,y) is uniquely represented in terms of this basis as $(x,y) = (x-y)(1,0) + y(1,1)$.

Basis : $(1,0), (1,1)$
$w_1 = (2,0)$
$w_2 = (1,1)$

$f(x,y) = (x-y)(2,0) + y(1,1)$
$= (2x-2y+y, y)$
$= (2x-y, y)$.

$$f(x,y) = (x-y)\, f(1,0) + y\, f(1,1)$$
$$= \boxed{(x-y)}\,(2,0) + \boxed{y}\,(0,1) = (2x-2y, y).$$
$$= (2(x-y), 0) + (0,y)$$

An important property of finite dimensional vector spaces

Let V be a vector space with dimension n. Choose a basis $\{v_1, v_2, \dots, v_n\}$.

Define $f : V \to \mathbb{R}^n$ by extending the function sending the basis vector v_i to the standard basis vector $e_i \in \mathbb{R}^n$ for each i.

Then f is an isomorphism.

$v = \sum c_i v_i$

$f(v) = \sum c_i e_i \implies f(v_i) = e_i$.

Onto : $(x_1, \dots, x_n) \in \mathbb{R}^n$. Let $v = \sum x_i v_i$.

Then $f(v) = \sum x_i e_i = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = (x_1, \dots, x_n)$

$v = \sum c_i v_i$

1-1 : $f(v) = 0 \implies \sum_{i=1}^{n} c_i e_i = (0, \dots, 0)$

$\implies (c_1, \dots, c_n) = (0, \dots, 0)$

$\implies c_i = 0 \implies v = 0v_1 + 0v_2 + \dots + 0v_n = 0$.

Example

Recall that we have computed that a basis for the subspace $W = \{(x, y, z) | x + y + z = 0\}$ is $(-1, 1, 0), (-1, 0, 1)$.

Then the homomorphism f obtained by extending $f(-1, 1, 0) = (1, 0)$ and $f(-1, 0, 1) = (0, 1)$ is an isomorphism.

Note that $(x, y, z) \in W$ can be uniquely expressed as $(x, y, z) = y(-1, 1, 0) + z(-1, 0, 1)$.

Hence, $f : W \to \mathbb{R}^2$ is $f(x, y, z) = y(1, 0) + z(0, 1) = (y, z)$.

onto : $(y, z) \in \mathbb{R}^2$. Choose $x = -y - z$ & consider $(x, y, z) \in W$.

$\therefore f(x, y, z) = (y, z)$.

1-1 : $f(x, y, z) = 0$ for $(x, y, z) \in W$ $\implies (y, z) = (0, 0) \implies y = 0, z = 0$

$\implies x = -y - z = 0$.

$\implies (x, y, z) = 0$.

Linear transformation in matrix form :

The matrix corresponding to a linear transformation with respect to ordered bases

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

Let $\beta = v_1, v_2, \ldots, v_n$ be an ordered basis of V and $\gamma = w_1, w_2, \ldots, w_m$ be an ordered basis of W.

Each $f(v_i)$ can be uniquely written as a linear combination of w_js, where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

$$
\begin{aligned}
f(v_1) &= a_{11}w_1 + a_{21}w_2 + \ldots + a_{m1}w_m \\
f(v_2) &= a_{12}w_1 + a_{22}w_2 + \ldots + a_{m2}w_m \\
&\vdots \\
f(v_n) &= a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{mn}w_m
\end{aligned}
$$

The matrix corresponding to a linear transformation with respect to ordered bases

The matrix corresponding to the linear transformation f with respect to the ordered bases β and γ is given by $\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & \vdots & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$.

e.g. Let $V = W = \mathbb{R}^2, \beta = \gamma = (1,0), (1,1)$ and $f(x,y) = (2x, y)$.

$$
\begin{aligned}
f(1,0) &= (2,0) = 2(1,0) + 0(1,1) \\
f(1,1) &= (2,1) = 1(1,0) + 1(1,1)
\end{aligned}
$$

Hence the matrix corresponding to f w.r.t. the ordered bases $\{(1,0),(1,1)\}$ is $\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$

Recovering the linear transformation

Let $\beta = v_1, v_2, \ldots, v_n$ and $\gamma = w_1, w_2, \ldots, w_m$ be ordered bases of V and W respectively. Suppose A is an $m \times n$ matrix. What is the corresponding linear transformation?

Let $v \in V$. Express $v = \sum_{j=1}^{n} c_j v_j$. Define

$$f(v) = \sum_{j=1}^{n} c_j \sum_{i=1}^{m} A_{ij} w_i.$$

Check that f is a linear transformation!

Letting $c_k = 1$ and $c_j = 0$ for all $j \neq k$, we get that
$$f(v_k) = A_{1k} w_1 + A_{2k} w_2 + \ldots + A_{mk} w_m.$$

Hence the matrix corresponding to f is indeed A.

Fixed ordered bases : Linear transformations $\leftrightarrow$ matrices

Let β and γ be ordered bases for vector spaces V and W respectively where $n = dim(V)$ and $m = dim(W)$.

There is a bijection :
{ linear transformations from V to W } $\leftrightarrow$ { $m \times n$ matrices } .

$f \longmapsto A$

$f(v_j) = \sum_{i=1}^{m} a_{ij} w_i$

$f(v_j) = \sum_{i=1}^{m} A_{ij} w_i$

Extend f to a lin. trans.

$f \longleftarrow\!\!| \; A$

Example :

Let $W = \{(x, y, z) | x + y + z = 0\}$, $V = \mathbb{R}^2$.

Let $\beta = (-1, 1, 0), (-1, 0, 1)$ and let γ be the standard basis of $\mathbb{R}^2$.

Recall that the isomorphism $f(x, y, z) = (y, z)$ from W to $\mathbb{R}^2$ was obtained by extending $f(-1, 1, 0) = (1, 0)$ and $f(-1, 0, 1) = (0, 1)$.

$$f(-1,1,0) = 1(1,0) + 0(0,1)$$
$$f(-1,0,1) = 0(1,0) + 1(0,1)$$

Hence, the matrix corresponding to the linear transformation f with respect to the ordered bases β and γ is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Suppose in the previous example, we change the basis γ from the standard basis to the ordered basis $(1, 0), (1, 1)$.

$$f(-1, 1, 0) = 1(1, 0) + 0(1, 1)$$
$$f(-1, 0, 1) = -1(1, 0) + 1(1, 1).$$

Hence, the matrix corresponding to the linear transformation f with respect to the new ordered bases $\beta = (-1, 1, 0), (-1, 0, 1)$ and $\gamma = (1, 0), (1, 1)$ is $\begin{bmatrix} 1 & \cancel{0}\,{-1} \\ \cancel{1}\,0 & 1 \end{bmatrix}$.

Thus, changing the ordered bases gives us different matrices corresponding to the same linear transformation.

Image and kernel of linear transformations

Definitions of kernel and image :

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

Define the kernel of f (denoted by $ker(f)$) as :

$$ker(f) = \{v \in V \mid f(v) = 0\}.$$

Define the image of f (denoted by $Im(f)$) as :

$$Im(f) = \{w \in W \mid \exists v \in V \text{ for which } f(v) = w\}.$$

$Im(f)$ is another name for the "range of the function f" which we have studied in Maths-1.

Examples :

Consider $f : \mathbb{R}^2 \to \mathbb{R}^2 \qquad f(x, y) = (2x, y)$.

Then $ker(f) = \{(0, 0)\}$ and $Im(f) = \mathbb{R}^2$.

Another example : consider $f : \mathbb{R}^2 \to \mathbb{R}^2 \qquad f(x, y) = (2x, 0)$.

Then $ker(f) = \{(0, y) | y \in \mathbb{R}\}$ i.e. the Y-axis.

Also $Im(f) = \{(x, 0) | x \in \mathbb{R}\}$ i.e. the X-axis.

A linear transformation f is 1-1 if and only if $ker(f) = 0$.

Writing this out for linear transformations, we see that : a linear transformation $f : V \to W$ is onto if and only if $Im(f) = W$

Kernels and null spaces

Let $f : V \to W$ be a linear transformation. Let $\beta = v_1, v_2, \dots, v_n$ and $\gamma = w_1, w_2, \dots, w_m$ be ordered bases of V and W respectively.

Let A be the matrix corresponding to f with respect to β and γ.

Recall that for $v = \sum_{j=1}^{n} c_j v_j \in V$, $\quad f(v) = \sum_{j=1}^{n} c_j \sum_{i=1}^{m} A_{ij} w_i$.

Hence, $\quad f(v) = 0 \iff \sum_{j=1}^{n} A_{ij} c_j = 0$ for all i.

Thus, $v = \sum_{j=1}^{n} c_j v_j \in ker(f)$

$$\iff c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \text{ is in the null space of } A.$$

Images and column spaces

Let $f : V \to W$ be a linear transformation. Let $\beta = v_1, v_2, \dots, v_n$ and $\gamma = w_1, w_2, \dots, w_m$ be ordered bases of V and W respectively.

Let A be the matrix corresponding to f with respect to β and γ.

Recall that for $v = \sum_{j=1}^{n} c_j v_j \in V$, $\quad f(v) = \sum_{j=1}^{n} c_j \sum_{i=1}^{m} A_{ij} w_i$.

Let $w = \sum_{i=1}^{m} d_i w_i \in W$. Then $w \in Im(f)$ precisely when there exist scalars $c_j; j = 1, 2, \dots, n$ such that $\sum_{j=1}^{n} A_{ij} c_j = d_i$ for all i.

Equivalently $w = \sum_{i=1}^{m} d_i w_i \in Im(f)$ if there exists a column

vector $c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ such that the column vector $d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix} = Ac.$

Hence, $w = \sum_{i=1}^{m} d_i w_i \in Im(f)$

$$\iff d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix} \text{ is in the column space of } A.$$

Bases for the kernel and image of a linear transformation

Let $f : V \to W$ be a linear transformation. Let $\beta = v_1, v_2, \ldots, v_n$ and $\gamma = w_1, w_2, \ldots, w_m$ be ordered bases of V and W respectively.

Let A be the matrix corresponding to f with respect to β and γ.

The relation between kernels and null spaces derived earlier actually yields an isomorphism between them.

In particular, the vectors $\begin{bmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{1n} \end{bmatrix}, \begin{bmatrix} c_{21} \\ c_{22} \\ \vdots \\ c_{2n} \end{bmatrix}, \ldots, \begin{bmatrix} c_{k1} \\ c_{k2} \\ \vdots \\ c_{kn} \end{bmatrix}$ form a basis for the null space of A precisely when $v_1', v_2', \ldots, v_k' \in ker(f)$, where $v_i' = \sum_{j=1}^{n} c_{ij} v_j$, form a basis for $ker(f)$.

Similarly, the relation between images and column spaces derived earlier yields an isomorphism between them.

In particular, the vectors $\begin{bmatrix} d_{11} \\ d_{12} \\ \vdots \\ d_{1m} \end{bmatrix}, \begin{bmatrix} d_{21} \\ d_{22} \\ \vdots \\ d_{2m} \end{bmatrix}, \dots, \begin{bmatrix} d_{r1} \\ d_{r2} \\ \vdots \\ d_{rn} \end{bmatrix}$ form a basis for the column space of A precisely when $w_1', w_2', \dots, w_r' \in im(f)$, where $w_i' = \sum_{j=1}^{m} d_{ij} w_j$, form a basis for $im(f)$.

Note further that under this isomorphism, the columns of A, which form a spanning set of the column space of A, correspond to the images $f(v_i)$, which form a spanning set for $im(f)$.

We can thus use row reduction to obtain these bases.

Examples of finding bases for the kernel and image of a linear transformation :

Recall :

- Kernel of a linear transformation
- For a linear transformation $f : V \to W$, let A be the matrix corresponding to f after choosing ordered bases $\beta = v_1, v_2, \dots, v_n$ and $\gamma = w_1, w_2, \dots, w_m$ for V and W respectively.

$$\begin{array}{lccc} \text{The isomorphism} & \mathbb{R}^n & \xrightarrow{\sim} & V \\ \text{given by} \quad c = & \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} & \mapsto & v = \sum_{j=1}^{n} c_j v_j \\ \text{restricts to an isomorphism} & \mathcal{N}(A) & \xrightarrow{\sim} & ker(f) \end{array}$$

- In particular, a basis of $\mathcal{N}(A)$ will yield a basis of $ker(f)$.

- Image of a linear transformation
- For a linear transformation $f : V \to W$, let A be the matrix corresponding to f after choosing ordered bases $\beta = v_1, v_2, \ldots, v_n$ and $\gamma = w_1, w_2, \ldots, w_m$ for V and W respectively.

$$\begin{array}{lccc}
\text{The isomorphism} & \mathbb{R}^m & \xrightarrow{\sim} & W \\
\text{given by} & d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix} & \mapsto & w = \sum_{j=1}^{n} d_j w_j \\
\text{restricts to an isomorphism} & \mathcal{C}(A) & \xrightarrow{\sim} & im(f)
\end{array}$$

- In particular, a basis of $\mathcal{C}(A)$ will yield a basis of $im(f)$.

Example

Consider $T : \mathbb{R}^4 \to \mathbb{R}^3$ defined by $T(x_1, x_2, x_3, x_4) = (2x_1 + 4x_2 + 6x_3 + 8x_4, x_1 + 3x_2 + 5x_4, x_1 + x_2 + 6x_3 + 3x_4)$

Choose β and γ to be the standard (ordered) bases for $\mathbb{R}^4$ and $\mathbb{R}^3$ respectively. The corresponding matrix is $\begin{bmatrix} 2 & 4 & 6 & 8 \\ 1 & 3 & 0 & 5 \\ 1 & 1 & 6 & 3 \end{bmatrix}$.

Row reduction yields :

$$\begin{bmatrix} 2 & 4 & 6 & 8 \\ 1 & 3 & 0 & 5 \\ 1 & 1 & 6 & 3 \end{bmatrix} \xrightarrow{R_1/2} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 0 & 5 \\ 1 & 1 & 6 & 3 \end{bmatrix} \xrightarrow[R_3 - R_1]{R_2 - R_1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -3 & 1 \\ 0 & -1 & 3 & -1 \end{bmatrix}$$

$$\xrightarrow{R_3 + R_2} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_1 - 2R_2} \begin{bmatrix} 1 & 0 & 9 & 2 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus the reduced row echelon form is $\begin{bmatrix} 1 & 0 & 9 & 2 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

The non-pivot columns (resp. independent variables) are the third and fourth (resp. X_3 and X_4).

Putting $X_3 = s$ and $X_4 = t$ and using the system of equations gives $X_1 = -9s - 2t$ and $X_2 = 3s - t$.

Substituting $s = 1, t = 0$ and $s = 0, t = 1$ gives the basis vectors $(-9, 3, 1, 0), (-2, -1, 0, 1)$ for the null space.

Since we have chosen β to be the standard ordered basis, the basis for the $ker(T)$ is also the same, i.e.
$-9e_1 + 3e_2 + 1e_3 + 0e_4 = (-9, 3, 1, 0)$ and
$(-2e_1 - 1e_2 + 0e_3 + 1e_4 = (-2, -1, 0, 1)$.

Moreover, the pivot columns are the first and second columns.

Hence the column space is spanned by the first 2 columns of the matrix A, i.e. $(2, 1, 1), (4, 3, 1)$.

Since we have chosen γ to be the standard ordered basis, the basis for $im(T)$ is also the same, i.e. $2e_1 + 1e_2 + 1e_3 = (2, 1, 1)$ and $4e_1 + 3e_2 + 1e_3 = (4, 3, 1)$.

Another example :

Let $V = \mathbb{R}^2, W = \{(x, y, z) | x + y + z = 0\}$. Let the respective ordered bases be $\beta = (1, 1), (1, -1)$ and $\gamma = (-1, 1, 0), (-1, 0, 1)$. $T(x, y) = (0, x + 2y, -x - 2y)$ defines a linear transformation from V to W.

The corresponding matrix is $\begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix}$.

Putting this into row reduced echelon form yields :

$$\begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix} \xrightarrow{R_1/3} \begin{bmatrix} 1 & -1/3 \\ -3 & 1 \end{bmatrix} \xrightarrow{R_2+3R_1} \begin{bmatrix} 1 & -1/3 \\ 0 & 0 \end{bmatrix}$$

The pivot column (dependent variable) is the first column while the non-pivot column (independent variable) is the second column.

Hence, the null space is given by putting $X_2 = t$ and using the corresponding system to obtain $X_1 = t/3$.

A basis for the null space is thus the singleton set consisting of the vector obtained by substituting $t = 1$ i.e. $\{(1/3, 1)\}$.

Therefore a basis for $ker(T)$ is the singleton set consisting of the vector $1/3(1, 1) + 1(1, -1) = (4/3, -2/3)$.

Similarly, a basis for the column space of A is given by the singleton set consisting of the first column (pivot column) i.e. $(3, -3)$.

Therefore a basis for $ker(T)$ is the singleton set consisting of $3(-1, 1, 0) + (-3)(-1, 0, 1) = (0, 3, -3)$.

The rank-nullity theorem for linear transformations

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

The rank of T (denoted $rank(T)$) is the dimension of $Im(T)$.

The nullity of T (denoted $nullity(T)$) is the dimension of $ker(T)$.

Reinterpreting the rank-nullity theorem for matrices, we obtain :

$$rank(T) + nullity(T) = dim(V)$$

Equivalence and similarity of matrices :

Equivalence of matrices

Let A and B be two matrices of order $m \times n$. We say A is equivalent to B if $B = QAP$ for some invertible $n \times n$ matrix P and for some invertible $m \times m$ matrix Q.

Other characterisations:

1) A can be transformed into B by a combination of elementary row and column operations.
2) $rank(A) = rank(B)$ $\quad QAP = \begin{bmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{bmatrix} = Q'BP'$

Equivalence of matrices is an equivalence relation i.e.

- A is equivalent to itself $\quad A = I_{m\times m} A I_{n\times n}$ $\quad B = QAP$
- A is equivalent to B implies B is equivalent to A. $\quad \Rightarrow A = Q^{-1}BP^{-1}$
- A is equivalent to B and B to C implies A is equivalent to C.

$B = QAP$, $C = Q'BP'$ $\Rightarrow C = Q'QAPP'$

Example

Consider the linear transformation $f : \mathbb{R}^3 \to \mathbb{R}^2$, defined as :

$$f(x, y, z) = (x + y, y + z).$$

Consider two ordered bases for $\mathbb{R}^3$:
$\beta_1 = (1, 0, 0), (0, 1, 0), (0, 0, 1)$ and $\beta_2 = (1, 1, 0), (0, 1, 1), (0, 0, 1)$.

Similarly, consider two ordered bases for $\mathbb{R}^2$:

$$\gamma_1 = (1, 0), (0, 1) \text{ and } \gamma_2 = (1, 0), (1, 1).$$

$f(1, 0, 0) = (1, 0)$,
$f(0, 1, 0) = (1, 1) = 1(1, 0) + 1(0, 1)$,
$f(0, 0, 1) = (0, 1)$.

Hence the matrix corresponding to f with respect to the bases β_1 and γ_1 is $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

Example (contd.)

$(1,1,0) = 1e_1 + 1e_2 + 0e_3$

$f(1, 1, 0) = (2, 1) = 1(1, 0) + 1(1, 1)$,
$f(0, 1, 1) = (1, 2) = -1(1, 0) + 2(1, 1)$,
$f(0, 0, 1) = (0, 1) = -1(1, 0) + 1(1, 1)$.

$(1,0) = 1(1,0) + 0(1,1)$
$(0,1) = -1(1,0) + 1(1,1)$

Hence the matrix corresponding to f with respect to the bases β_2 and γ_2 is $B = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 2 & 1 \end{bmatrix}$.

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

Choose $P = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$ and $Q = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$. Then

$$QAP = \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 2 & 1 \end{bmatrix} = B$$

Hence A and B are equivalent to each other.

Consider a linear transformation $T : V \to W$, two ordered bases β_1 and β_2 for V, and two ordered bases γ_1 and γ_2 for W.

Let A be the matrix corresponding to T with respect to the bases β_1 and γ_1 and B be the matrix corresponding to T with respect to the bases β_2 and γ_2.

Then A is equivalent to B !

$P \to$ express the ordered basis β_2 in terms of the ordered basis β_1.

$Q \to$ express the ordered basis γ_1 in terms of γ_2.

Then $B = QAP$.

Similar Matrices :

An $n \times n$ matrix A is similar to an $n \times n$ matrix B if there exists an $n \times n$ invertible matrix P such that $B = P^{-1}AP$.

Note that similarity is an equivalence relation, i.e. :

- A is similar to itself $P = I$: $A = I^{-1} A I$.
- A is similar to B implies B is similar to A. ✓
- A is similar to B and B to C implies A is similar to C. ✓

$B = P^{-1}AP \Rightarrow PBP^{-1} = A \Rightarrow A = (P^{-1})^{-1} B P^{-1}$

$B = P^{-1}AP$, $C = Q^{-1}BQ$

$\Rightarrow C = Q^{-1}(P^{-1}AP)Q = Q^{-1}P^{-1}APQ$

$= (PQ)^{-1} A (PQ)$

Properties of similar matrices :

Suppose A and B are similar matrices. Then the following properties hold :

- A and B are equivalent.
- A and B have the same rank.
- $det(B) = det(P^{-1}AP) = det(P^{-1})det(A)det(P)$
$$= \frac{1}{det(P)} det(A)det(P) = det(A).$$
- Several other invariants of A and B are the same such as the characteristic polynomial, minimal polynomial and eigen values (with multiplicity).

An example of similar matrices

Consider the linear transformation $f : \mathbb{R}^3 \to \mathbb{R}^3$ where $f(x, y, z) = (-x + y + z, x - y + z, x + y - z)$.

Let $\beta = \gamma$ both be the standard ordered basis of $\mathbb{R}^3$.

Then we get :

$$\begin{aligned} f(1,0,0) = (-1,1,1) &= -1(1,0,0) + 1(0,1,0) + 1(0,0,1) \\ f(0,1,0) = (1,-1,1) &= 1(1,0,0) - 1(0,1,0) + 1(0,0,1) \\ f(0,0,1) = (1,1,-1) &= 1(1,0,0) + 1(0,1,0) - 1(0,0,1) \end{aligned}$$

Hence the matrix of f corresponding to the standard ordered basis is $\begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$.

Consider another ordered basis $\beta' = (1,1,1), (-1,1,0), (-1,0,1)$.

Then we have the following:

$$\begin{aligned}
f(1,1,1) &= (1,1,1) &&= 1(1,1,1) + 0(-1,1,0) + 0(-1,0,1) \\
f(-1,1,0) &= (2,-2,0) &&= 0(1,1,1) - 2(-1,1,0) + 0(-1,0,1) \\
f(-1,0,1) &= (2,0,-2) &&= 0(1,1,1) + 0(-1,1,0) - 2(-1,0,1)
\end{aligned}$$

Hence the matrix of f corresponding to the ordered basis β' is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$.

Let $P = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. Then $P^{-1} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix}$.

Then $P^{-1}AP = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

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

Hence A and B are similar matrices.

Linear transformations and similarity of matrices

Consider a linear transformation $T : V \to V$ and two ordered bases β and γ for V.

Let A be the matrix corresponding to T with respect to the basis β and B the matrix corresponding to T with respect to the basis γ.

Then A is similar to B ! $B = P^{-1}AP$

$P \to$ Express γ in terms of β.

$P^{-1} \to$ Express β " " " γ.

Why do we care about similarity? Because under some basis, we hope that the corresponding matrix is a diagonal matrix which gives an easy geometric understanding of the linear transformation.

Affine subspaces and affine mappings :

Affine Subspaces :

Let V be a vector space. An affine subspace of V is a subset L such that there exists $v \in V$ and a vector subspace $U \subseteq V$ such that

$$L = v + U := \{v + u | u \in U\}.$$

We say an affine subspace L is n-dimensional if the corresponding subspace U is n-dimensional.

The subspace U corresponding to an affine subspace is unique.

However the vector v is not unique and in fact can be any vector in L.

$v' \in L$
$\Rightarrow v' = v + u_1$ where $u_1 \in U$
$\Rightarrow v' - v = u_1 \in U$

$L = v + U$
$= v' + U'$

$v - v' \in U'$
$\Rightarrow v' - v \in U'$

$v \in L$, $v' \in L$
Let $u \in U$
Then $v + u \in L = v' + U'$
$\Rightarrow v + u = v' + u'$ for some $u' \in U'$
$\Rightarrow u = (v' - v) + u' \in U'$

The subspace U corresponding to an affine subspace is unique.

However the vector v is not unique and in fact can be any vector in L.

$L = v + U$
$= v' + U'$
$v - v' \in U$ & U'
$\Rightarrow U = U'$.

Affine subspaces are thus translates of a vector subspace of V.

Affine subspaces in $\mathbb{R}^2$

$L = \{(x,y)\} = (x,y) + \{(0,0)\}$

- Points

$L \equiv y = mx + c$ is a translate of $y = mx$

$(0,c) \in L$

- Lines

$L = \underbrace{(0,c)}_{v} + \underbrace{\{(x,mx) \mid x \in \mathbb{R}\}}_{U}$

- the entire plane $\mathbb{R}^2$

$\underbrace{\mathbb{R}^2}_{L} = \underbrace{(0,0)}_{v} + \underbrace{\mathbb{R}^2}_{U}$

A subset which is not an affine subspace : the parabola $y = x^2 + 1$ or the curve $y^2 = x^3$.

Affine subspaces in $\mathbb{R}^3$

- Points $\quad v + \lambda(0,0)$
- Lines $\quad v + \lambda v_1$
- Planes $\quad v + \lambda_1 v_1 + \lambda_2 v_2$
- the entire space $\mathbb{R}^3$ $\quad \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3$.

Example: Two-dimensional affine subspaces in $\mathbb{R}^3$ can be expressed as

$$l = v + \boxed{\lambda_1 v_1 + \lambda_2 v_2}$$

$U = \{\lambda_1 v_1 + \lambda_2 v_2 \mid \lambda_1, \lambda_2 \in \mathbb{R}\}$

where $\lambda_1, \lambda_2 \in \mathbb{R}$ and v, v_1, v_2 are vectors in $\mathbb{R}^3$. $\quad = \mathrm{Span}(v_1, v_2)$

The solution set of system of linear equations as Affine subspace of Rn :

The solution set to a system of linear equations

Let $Ax = b$ be a linear system of equations.

- $b = 0$: In this case, it is a homogeneous system and as seen before, the solution set is a subspace of $\mathbb{R}^n$, namely the null space $\mathfrak{N}(A)$ of A.
- $b \notin$ column space of A : In this case, $Ax = b$ does not have a solution, so the solution set is the empty set.
- $b \in$ column space of A : In this case, the solution set L is an affine subspace of $\mathbb{R}^n$. Specifically, it can be described as $L = v + \mathfrak{N}(A)$ where v is any solution of the equation $Ax = b$.

Affine mappings of affine subspaces

Let L and L' be affine subspaces of V and W respectively. Let $f : L \to L'$ be a function. Consider any vector $v \in L$ and the unique subspace $U \subseteq V$ such that $L = v + U$. Note that $f(v) \in L'$ and hence $L' = f(v) + U'$ where U' is the unique subspace of W corresponding to L'. Then f is an affine mapping from L to L' if the function $g : U \to U'$ defined by $g(u) = f(u + v) - f(v)$ is a linear transformation.

For a linear transformation $T : U \to U'$ and fixed vectors $v \in L$ and $v' \in L'$, an affine mapping f can be obtained by defining $f(v + u) = v' + T(u)$, and in fact every affine mapping is obtained in this way.

$g(u) = f(u+v) - f(v)$ $\quad g(u') = f(u'+v) - f(v)$

$g(u+u') \stackrel{?}{=} g(u) + g(u')$ $\quad g(u+u') = f(u+u'+v) - f(v)$

$f(u+u'+v) - f(v) = f(u+v) - f(v) + f(u'+v) - f(v) \Rightarrow f(u+u'+v) = f(u+v) + f(u'+v) - f(v)$

$\Rightarrow f(u+u'+v) + \boxed{f(v)} = f(u+v) + f(u'+v)$

An example and an important special case

Let $T'(x, y, z) = (2x + 3y + 2, 4x - 5y + 3)$. Then this is an affine mapping from $\mathbb{R}^3$ to $\mathbb{R}^2$.

$T'(x,y,z) = \underbrace{(2,3)}_{w} + \underbrace{(2x+3y,\ 4x-5y)}_{T(x,y,z)}$ is a lin. trans.

Let $T : V \to W$ be a linear transformation and $w \in W$, then the mapping

$$T' : V \to W$$
$$T'(v) = w + T(v)$$

is an affine mapping from V to W.

Lengths and angles :

The dot product of two vectors in R 2:

Consider the two vectors $(3, 4)$ and $(2, 7)$ in $\mathbb{R}^2$. The dot product of these two vectors gives us a scalar as follows:

$$(3, 4) \cdot (2, 7) = 3 \times 2 + 4 \times 7 = 6 + 28 = 34$$

For two general vectors (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$, the dot product of these two vectors is the scalar computed as follows :

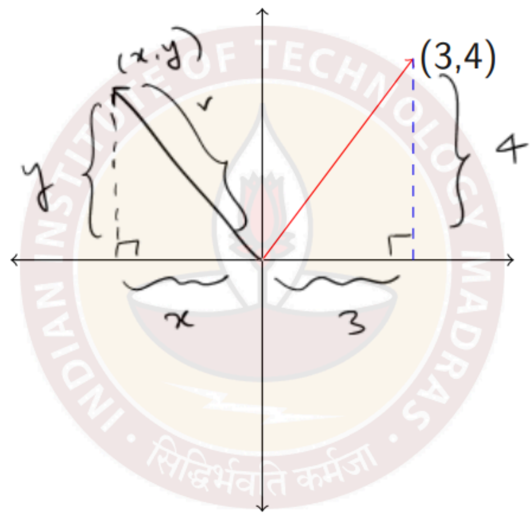
$$(x_1, y_1) \cdot (x_2, y_2) = x_1x_2 + y_1y_2.$$

Length of a vector :

Let us find the length of the vector $(3, 4)$ in $\mathbb{R}^2$.

$v\ (x,y)$

length is $\sqrt{x^2+y^2}$.



Using Pythagoras' theorem, the length of the vector $(3, 4)$ is $\sqrt{3^2 + 4^2} = 5$ units.

Observe that $(3, 4) \cdot (3, 4) = 3^2 + 4^2$, and hence the length of $(3, 4)$ is the square root of the dot product of the vector with itself.

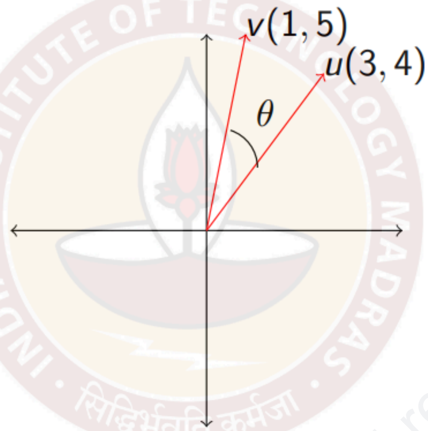
$$\text{Length of the vector } (3, 4) = \sqrt{(3, 4) \cdot (3, 4)} = \sqrt{3^2 + 4^2} = \cancel{\sqrt{5}}\ \sqrt{25} = 5.$$

More generally, the length of the vector $(x, y) \in \mathbb{R}^2$ is $\sqrt{x^2 + y^2} = \sqrt{(x, y) \cdot (x, y)}$.

$$\sqrt{x^2+y^2} = \sqrt{(x,y)\cdot(x,y)}$$

The angle between two vectors in $\mathbb{R}^2$

- The angle between the vectors u and v and measures how far the direction is of v from u (or vice versa). e.g. θ is the angle between $u = (3, 4)$ and $v = (1, 5)$.

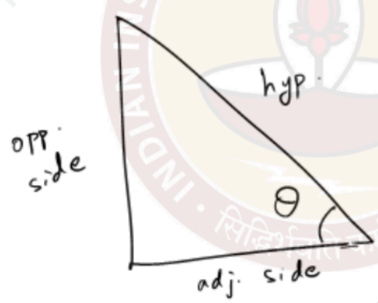


- It is measured in degrees (between 0 and 360) or radians (between 0 and 2π).
- The angle is often described or measured by computing its trignometric functions (e.g. *sin*, *cos* , *tan*).

The dot product and the angle between two vectors in $\mathbb{R}^2$

Let u and v be two vectors in $\mathbb{R}^2$. Then we can compute the angle θ between the vectors u and v using the dot products as :

$$cos(\theta) = \frac{u \cdot v}{\sqrt{(v \cdot v) \times (u \cdot u)}} \quad i.e. \quad \theta = cos^{-1}\left(\frac{u \cdot v}{\sqrt{(v \cdot v) \times (u \cdot u)}}\right).$$



$$\cos(\theta) = \frac{\text{length of adj. side}}{\text{length of hypotenuse}}$$

$$\sin(\theta) = \frac{\text{length of the opp. side}}{\text{length of hypotenuse}}$$

$$\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)} = \frac{\text{length(opp.)}}{\text{length (adj.)}}.$$

For two general vectors (x_1, y_1, z_1) and (x_2, y_2, z_2) in $\mathbb{R}^3$, the dot product of these two vectors is the scalar computed as follows :

$$(x_1, y_1, z_1) \cdot (x_2, y_2, z_2) = x_1x_2 + y_1y_2 + z_1z_2.$$

More generally, the length of the vector $(x, y, z) \in \mathbb{R}^3$ is $\sqrt{x^2 + y^2 + z^2} = \sqrt{(x, y, z) \cdot (x, y, z)}$.

The angle between two vectors in $\mathbb{R}^3$ and the dot product

The angle between the vectors u and v in $\mathbb{R}^3$ is the angle between them computed by passing a plane through them.

It measures how far the direction is of v from u (or vice versa) on that plane.

Let u and v be two vectors in $\mathbb{R}^3$. Then we can compute the angle θ between the vectors u and v using the dot product as :

$$cos(\theta) = \frac{u \cdot v}{\sqrt{(v \cdot v) \times (u \cdot u)}} \quad i.e. \quad \theta = cos^{-1}\left(\frac{u \cdot v}{\sqrt{(v \cdot v) \times (u \cdot u)}}\right).$$

Dot products in R n : length and angle :

Let $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ be vectors in $\mathbb{R}^n$.

The dot product of the two vectors u and v is defined as
$u \cdot v = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$.

The length of the vector u is denoted by $||u||$ and defined by
$||u|| = \sqrt{u \cdot u}$.

The angle θ between the two vectors u and v is measured on the 2-dimensional plane spanned by u and v and can be computed as :

$$cos(\theta) = \frac{u \cdot v}{||u|| \times ||v||} \quad i.e. \quad \theta = cos^{-1}\left(\frac{u \cdot v}{\sqrt{(v \cdot v) \times (u \cdot u)}}\right).$$

Inner products and norms on a vector space :

Inner product on a vector space:

An inner product on a vector space V is a function
$\langle\ .\ ,\ .\ \rangle : V \times V \to \mathbb{R}$ satisfying the following :

- $\langle v, v\rangle > 0$ for all $v \in V \setminus \{0\}$; $\langle v, v\rangle = 0$ if and only if $v = 0$.
- $\langle v_1 + v_2, v_3\rangle = \langle v_1, v_3\rangle + \langle v_2, v_3\rangle$
- $\langle v_1, v_2\rangle = \langle v_2, v_1\rangle$
- $\langle cv_1, v_2\rangle = c\langle v_1, v_2\rangle = \langle v_1, cv_2\rangle$. $, c \in \mathbb{R}$.

$\langle v_1, v_2 + v_3\rangle = \langle v_1, v_2\rangle + \langle v_1, v_3\rangle$.

A vector space V together with an inner product $\langle ., .\rangle$ is called an inner product space.

The dot product is an example of an inner product

Recall that the dot product of $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ be in $\mathbb{R}^n$ is

$$u \cdot v = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

. This yields a function

$$\langle\ .\ ,\ .\ \rangle : V \times V \to \mathbb{R} \quad ; \quad \langle u, v\rangle = u \cdot v.$$

Norm on a vector space :

A norm on a vector space V is a function

$$||.|| : V \to \mathbb{R}$$

$$x \mapsto ||x||$$

satisfying the following conditions:

- $||x + y|| \leq ||x|| + ||y||$, for all $x, y \in V$

  $v + w$ $\quad$ v $\quad$ w

- $||cx|| = |c|||x||$ for all $c \in \mathbb{R}$ and for all $x \in V$

  cv $\quad$ $|c|\ ||v||$

- $||x|| \geq 0$ for all $x \in V$; $||x|| = 0$ if and only if $x = 0$

  $||v||$ $\quad$ v $\quad$ $||v|| = 0$ iff $v = 0$.

Length as an example of a norm

Recall that the length of a vector $u = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is

$$||u|| = \sqrt{(x_1^2 + x_2^2 + \ldots + x_n^2)} \qquad .$$

The length function $\mathbb{R}^n \to \mathbb{R}$ is a norm on $\mathbb{R}^n$.

The inner product induces a norm

Let V be an inner product space with inner product $\langle\ .\ ,\ .\ \rangle$.

Then the function $||.|| : V \to \mathbb{R}$ defined by $||v|| = \sqrt{\langle v, v\rangle}$ is a norm on V.

$$||v|| = 0 \iff \sqrt{\langle v,v\rangle} = 0 \iff \langle v,v\rangle = 0 \iff v = 0.$$

If $v \neq 0$, $\langle v,v\rangle > 0 \Rightarrow \sqrt{\langle v,v\rangle} > 0 \Rightarrow ||v|| > 0$.

$$||cv|| = \sqrt{\langle cv, cv\rangle} = \sqrt{c \times c \langle v,v\rangle} = \sqrt{c^2}\sqrt{\langle v,v\rangle} = |c|\,||v||.$$

$$||v+w|| = \sqrt{\langle v+w, v+w\rangle} = \sqrt{\langle v,v\rangle + \langle v,w\rangle + \langle w,v\rangle + \langle w,w\rangle} = \sqrt{\langle v,v\rangle + 2\langle v,w\rangle + \langle w,w\rangle}$$

$$||v+w||^2 = \langle v,v\rangle + 2\langle v,w\rangle + \langle w,w\rangle = ||v||^2 + ||w||^2 + 2\langle v,w\rangle \leq ||v||^2 + ||w||^2 + 2\,||v||\,||w|| = (||v|| + ||w||)^2.$$

Orthogonality and linear independence :

Recall that if θ is the angle between two vectors u and v (of $\mathbb{R}^n$) on the subspace spanned by them, then

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

Recall also that the dot product and the length are special cases of an inner product $\langle\ .\ ,\ .\ \rangle$ is and a norm on $\mathbb{R}^n$.

The geometric intuition of orthogonal vectors

If the angle θ between two vectors u and v in $\mathbb{R}^n$ is a right angle (i.e. $90°$), then

$$cos(\theta) = 0 = \frac{u \cdot v}{||u|| ||v||}.$$

Then $u \cdot v = 0$.

e.g. $(1,2,3)$ and $(2,2,-2))$ are orthogonal.

$$(1,2,3)\cdot(2,2,-2) = 1\times 2 + 2\times 2 + 3\times(-2)$$
$$= 2+4-6 = 0.$$

Orthogonal vectors

Two vectors u and v of an inner product space V are said to be orthogonal if $\langle u, v\rangle = 0$.

e.g. consider $\mathbb{R}^2$ with the inner product

$$\langle u, v\rangle = x_1y_1 - (x_1y_2 + x_2y_1) + 2x_2y_2$$

where $u = (x_1, x_2)$ and $v = (y_1, y_2)$.

Then the vectors $(1,1)$ and $(1,0)$ are orthogonal (w.r.t. this inner product).

$$\langle (1,1), (1,0)\rangle = 1\times 1 - (1\times 0 + 1\times 1) + 2\times 1\times 0$$
$$= 1-0-1+0$$
$$= 0.$$

An orthogonal set of vectors

An orthogonal set of vectors of an inner product space V is a set of vectors whose elements are mutually orthogonal.

Explicitly, if $S = \{v_1, v_2, \dots, v_k\} \subseteq V$, then S is an orthogonal set of vectors if

$$\langle v_i, v_j \rangle = 0 \text{ for } i, j \in \{1, 2, \dots, k\} \text{ and } i \neq j.$$

e.g. consider $\mathbb{R}^3$ with the usual inner product i.e. the dot product. Then the set $S = \{(4, 3, -2), (-3, 2, -3), (-5, 18, 17)\}$ is an orthogonal set of vectors.

$(4,3,-2) \cdot (-3,2,-3) = 4 \times (-3) + 3 \times 2 + (-2) \times (-3)$
$= -12 + 6 + 6 = 0.$

$(4,3,-2) \cdot (-5,18,17) = -20 + 54 - 34 = 0$

$(-3,2,-3) \cdot (-5,18,17) = 15 + 36 - 51 = 0.$

Orthogonality and linear independence

Let $\{v_1, v_2, \dots, v_k\}$ be an orthogonal set of non-zero vectors in the inner product space V.

Then $\{v_1, v_2 \dots, v_k\}$ is a linearly independent set of vectors.

Suppose $\sum_{i=1}^{k} c_i v_i = 0$

Then $\left\langle \sum_{i=1}^{k} c_i v_i, v_1 \right\rangle = \langle 0, v_1 \rangle = 0.$

$\therefore \sum_{i=1}^{k} \langle c_i v_i, v_1 \rangle = 0$

$\therefore \sum_{i=1}^{k} c_i \langle v_i, v_1 \rangle = 0$

$\therefore c_1 \underbrace{\langle v_1, v_1 \rangle}_{\neq 0} = 0 \Rightarrow c_1 = 0.$

What is an orthogonal basis

Let V be an inner product space. A basis consisting of mutually orthogonal vectors is called an orthogonal basis.

Since an orthogonal set of vectors is already linearly independent, an orthogonal set is a basis precisely when it is a maximal orthogonal set (i.e. there is no orthogonal set strictly containing this one).

If $dim(V) = n$, then

$$\boxed{\text{orthogonal basis} \equiv \text{orthogonal set of } n \text{ vectors.}}$$

Examples of orthogonal bases :

1. the standard basis.
2. $\{(4, 3, -2), (-3, 2, -3), (-5, 18, 17)\} \subseteq \mathbb{R}^3$.
3. consider $\mathbb{R}^2$ with the inner product $\langle (x_1, x_2), (y_1, y_2) \rangle = x_1y_1 - (x_1y_2 + x_2y_1) + 2x_2y_2$. Then $\{(1, 1), (1, 0)\}$ is an orthogonal basis.

- An orthogonal set of vectors $\{v_1, v_2, \ldots, v_k\}$ of an inner product space V is a set of vectors whose elements are mutually orthogonal. i.e.

$$\langle v_i, v_j \rangle = 0 \text{ for } i, j \in \{1, 2, \ldots, k\} \text{ and } i \neq j.$$

- An orthogonal set of vectors is linearly independent.
- A maximal orthogonal set is a basis and is called an orthogonal basis.

What is an orthonormal set?

An orthonormal set of vectors of an inner product space V is an orthogonal set of vectors such that the norm of each vector of the set is 1.

Explicitly, if $S = \{v_1, v_2, \ldots, v_k\} \subseteq V$, then S is an orthonormal set of vectors if

$$\begin{aligned} &\langle v_i, v_j\rangle = 0 \qquad \text{for } i,j \in \{1,2,\ldots,k\} \text{ and } i \neq j \\ \text{and} \qquad &||v_i|| = 1 \qquad \forall i \in \{1,2,\ldots,k\} \end{aligned}$$

e.g. consider $\mathbb{R}^4$ with the usual inner product i.e. the dot product. Then the set

$\frac{1}{3}+\frac{1}{3}+\frac{1}{3}=1$ $\quad \frac{4}{42}+\frac{1}{42}+\frac{1}{42}+\frac{36}{42}=1$ $\quad \frac{4}{9}+\frac{4}{9}+\frac{1}{9}=1.$

$$\left\{\left(\frac{-1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0\right), \left(\frac{2}{\sqrt{42}}, \frac{1}{\sqrt{42}}, \frac{1}{\sqrt{42}}, \frac{6}{\sqrt{42}}\right), \left(\frac{2}{3}, 0, \frac{2}{3}, \frac{-1}{3}\right)\right\}$$

is an orthonormal set of vectors.

What is an orthonormal basis?

An orthonormal basis is an orthonormal set of vectors which forms a basis.

Equivalently : An orthonormal basis is an orthogonal basis where the norm of each vector is 1.

Equivalently : An orthonormal basis is a maximal orthonormal set.

Example : The standard basis w.r.t. the usual inner product forms an orthonormal basis.

$$\langle e_i, e_j \rangle = (0,0,\ldots,0,\overset{i^{th}}{1},0,\ldots,0)\cdot(0,0,\ldots,\overset{j^{th}}{1},0,\ldots,0) \quad i\neq j$$
$$= 0\times 0 + \cdots + 1\times 0 + 0\cdots + 0 + 0\times 1 + 0 + \cdots + 0 = 0.$$

$$||e_i|| = \sqrt{\langle e_i, e_i\rangle} = \sqrt{0\times 0 + \cdots + 1\times 1 + 0 + \cdots + 0} = \sqrt{1} = 1.$$

Another example

Consider $\mathbb{R}^3$ with the usual inner product and the set $\beta = \left\{\frac{1}{3}(1,2,2), \frac{1}{3}(-2,-1,2), \frac{1}{3}(2,-2,1)\right\}$. Then β forms an orthonormal basis of $\mathbb{R}^3$.

v_1, v_2, v_3

$||v_1||^2 = \langle v_1, v_1\rangle = \left\langle \frac{1}{3}(1,2,2), \frac{1}{3}(1,2,2)\right\rangle = \frac{1}{9}(1\times1+2\times2+2\times2)$
$= 9/9 = 1.$

$||v_2||^2 = \langle v_2, v_2\rangle = \left\langle \frac{1}{3}(-2,-1,2), \frac{1}{3}(-2,-1,2)\right\rangle$
$= \frac{1}{9}\{(-2)\times(-2) + (-1)\times(-1) + 2\times2\} = 9/9 = 1.$

$||v_3||^2 = 1.$

$\langle v_1, v_2\rangle = \left\langle \frac{1}{3}(1,2,2), \frac{1}{3}(-2,-1,2)\right\rangle = \frac{1}{9}\langle(1,2,2),(-2,-1,2)\rangle$
$= \frac{1}{9}(1\times(-2) + 2\times(-1) + 2\times2) = 0.$

$\langle v_1, v_3\rangle = \langle v_2, v_3\rangle = 0$

$|\beta| = 3$ & β is lin. indept. & $\dim(\mathbb{R}^3) = 3$

$\Rightarrow \beta$ is an o.n. basis.

Obtaining orthonormal sets from orthogonal sets

Let V be an inner product space. If $\Gamma = \{v_1, v_2, \ldots, v_k\}$ is an orthogonal set of vectors, then we can obtain an orthonormal set of vectors β from Γ by

$$\beta = \left\{\frac{v_1}{||v_1||}, \frac{v_2}{||v_2||}, \ldots, \frac{v_k}{||v_k||}\right\}.$$

Example : Consider $\mathbb{R}^2$ with the usual inner product and the orthogonal basis $\Gamma = \{(1,3),(-3,1)\}$

Then $\beta = \left\{\frac{1}{\sqrt{10}}(1,3), \frac{1}{\sqrt{10}}(-3,1)\right\}$ is an orthonormal basis of $\mathbb{R}^2$.

$\checkmark\ \langle v_i, v_j\rangle = 0$

$\Rightarrow \left\langle \frac{v_i}{||v_i||}, \frac{v_j}{||v_j||}\right\rangle = \frac{1}{||v_i||\,||v_j||}\langle v_i, v_j\rangle$
$= 0.$

$\left|\left|\frac{v_i}{||v_i||}\right|\right| = \frac{1}{||v_i||}\,||v_i|| = 1.$

Why are orthonormal bases important?

Suppose $\Gamma = \{v_1, v_2, \ldots, v_n\}$ is an orthonormal basis of an inner product space V and let $v \in V$.

Then v can be written as

$$v = c_1 v_1 + c_2 v_2 + \ldots + c_n v_n.$$

How do we find $c_1, c_2, \ldots, c_n$? For any basis, this means writing a system of linear equations and solving it.

But since Γ is orthonormal, we can use the inner product and compute $c_i = \langle v, v_i \rangle$.

$$\langle v, v_i \rangle = \langle c_1 v_1 + c_2 v_2 + \cdots + c_i v_i + \cdots + c_n v_n , v_i \rangle$$
$$= c_1 \cancel{\langle v_1, v_i \rangle} + c_2 \cancel{\langle v_2, v_i \rangle} + \cdots + c_i \boxed{\langle v_i, v_i \rangle} + \cdots$$
$$+ c_n \cancel{\langle v_n, v_i \rangle}$$
$$= c_i \langle v_i, v_i \rangle = c_i \|v_i\|^2 = c_i.$$

Example

$\left\{\frac{1}{\sqrt{10}}(1,3), \frac{1}{\sqrt{10}}(-3,1)\right\}$ is an orthonormal basis of $\mathbb{R}^2$. Write $(2,5)$ as a linear combination in terms of these basis vectors.

$$(2,5) = c_1 \frac{1}{\sqrt{10}}(1,3) + c_2 \frac{1}{\sqrt{10}}(-3,1)$$

$$c_1 = \left\langle (2,5), \frac{1}{\sqrt{10}}(1,3) \right\rangle = \frac{1}{\sqrt{10}}(2 \times 1 + 5 \times 3)$$
$$= \frac{1}{\sqrt{10}} 17.$$

$$c_2 = \left\langle (2,5), \frac{1}{\sqrt{10}}(-3,1) \right\rangle = \frac{-1}{\sqrt{10}}.$$
$$= \frac{1}{\sqrt{10}}(2 \times (-3) + 5 \times 1)$$

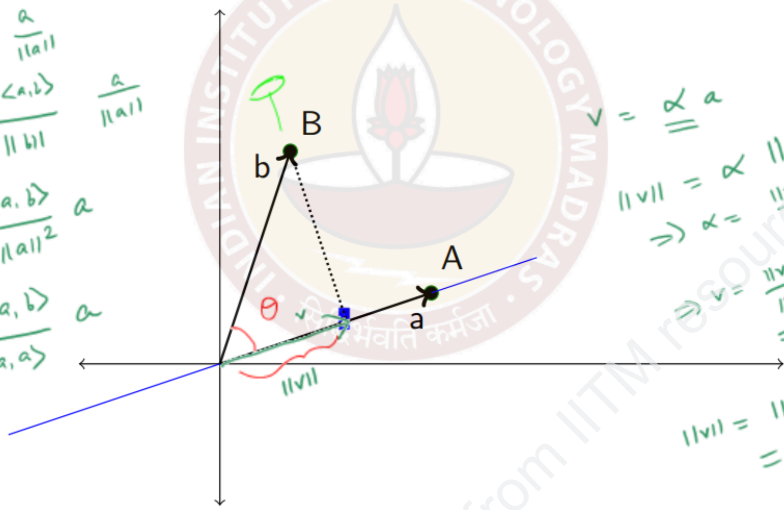
$$(2,5) = \frac{17}{\sqrt{10}} v_1 + \frac{-1}{\sqrt{10}} v_2 = \frac{17}{\sqrt{10}} \times \frac{1}{\sqrt{10}}(1,3) - \frac{1}{\sqrt{10}} \times \frac{1}{\sqrt{10}}(-3,1)$$
$$= \frac{17}{10}(1,3) - \frac{1}{10}(-3,1).$$

Projections using inner products :

Shortest distances in $\mathbb{R}^2$

A and B are points in the plane $\mathbb{R}^2$ and we want to find the nearest point from B on the line passing through A and the origin. Drop a perpendicular from B on to the line. Let a and b be the vectors corresponding to the points A and B respectively.

$$v = \|v\| \frac{a}{\|a\|} = \frac{\|b\| \langle a, b\rangle}{\|a\| \, \|b\|} \frac{a}{\|a\|} = \frac{\langle a, b\rangle}{\|a\|^2} a = \frac{\langle a, b\rangle}{\langle a, a\rangle} a$$



$$v = \underline{\underline{\alpha}}\, a$$

$$\|v\| = \alpha \|a\| \Rightarrow \alpha = \frac{\|v\|}{\|a\|}$$

$$\Rightarrow v = \frac{\|v\|}{\|a\|} a = \|v\| \frac{a}{\|a\|}$$

$$\|v\| = \|b\| \cos\theta = \|b\| \frac{\langle a, b\rangle}{\|a\| \, \|b\|}$$

The projection of a vector to a subspace

Let V be an inner product space , $v \in V$ and $W \subseteq V$ be a subspace. Then the projection of v onto W is the vector in W, denoted by $proj_W(v)$, computed as follows :

Find an orthonormal basis $\{v_1, v_2, \dots v_n\}$ for $\cancel{V}$ W.

Define $$proj_W(v) = \sum_{i=1}^{n} \langle v, v_i \rangle v_i .$$

Fact : The definition is independent of the chosen orthonormal basis (i.e. the expression on the RHS does not change even if you choose a different orthonormal basis).

Find $w \in W$ s.t. $\|v - w\|$ is smallest.
Ans. $\therefore w = proj_W(v)$.

The projection of v onto W is the vector in W closest to v. Note that "closest" is in terms of the distance based on the norm induced by the inner product.

Previous examples

$V = \mathbb{R}^2$, $W = \langle(3,1)\rangle$, $v = (1,3)$. Then $proj_W(v) = (1.8, 0.6)$. ✓

$\frac{1}{\sqrt{10}}(3,1)$ is an o.n. basis for W.

$proj_W(v) = \langle v, \frac{1}{\sqrt{10}}(3,1)\rangle \frac{1}{\sqrt{10}}(3,1) = \langle (1,3), (3,1)\rangle \frac{1}{10}(3,1)$
$= \frac{(1\times3+3\times1)}{10}(3,1) = \frac{6}{10}(3,1)$
$= (1.8, 0.6)$.

$V = \mathbb{R}^3$, $W = \langle(1,0,0),(0,1,0)\rangle$, $v = (2,3,5)$.
Then $proj_W(v) = (2,3,0)$.

o.n. basis $\langle (1,0,0), (0,1,0)\rangle$.

$proj_W(v) = \langle (2,3,5), (1,0,0)\rangle (1,0,0) + \langle (2,3,5), (0,1,0)\rangle (0,1,0)$
$= 2(1,0,0) + 3(0,1,0) = (2,3,0)$.

Projection on a vector and orthogonal bases

Let V be an inner product space and $v, w \in V$. Define

$$proj_w(v) = proj_{\langle w\rangle}(v).$$

Note that an orthonormal basis for $\langle w\rangle$ is $\frac{w}{||w||}$ and hence

$$proj_w(v) = \left\langle v, \frac{w}{||w||}\right\rangle \frac{w}{||w||} = \frac{\langle v, w\rangle}{||w||^2} w = \frac{\langle v, w\rangle}{\langle w, w\rangle} w.$$

Similarly, if $\{v_1, v_2, \ldots v_n\}$ is an orthogonal basis for a subspace W, then $\left\{\frac{v_1}{||v_1||}, \frac{v_2}{||v_2||}, \ldots \frac{v_n}{||v_n||}\right\}$ is an orthonormal basis for W and hence

$$proj_W(v) = \sum_{i=1}^{n} \left\langle v, \frac{v_i}{||v_i||}\right\rangle \frac{v_i}{||v_i||} = \sum_{i=1}^{n} \boxed{\frac{\langle v, v_i\rangle}{\langle v_i, v_i\rangle} v_i} = \sum_{i=1}^{n} proj_{v_i}(v).$$

Example

Let W be the 2-dimensional subspace of $V = \mathbb{R}^3$ spanned by the orthogonal vectors $v_1 = (1, 2, 1)$ and $v_2 = (1, -1, 1)$. What is the projection of $v = (-2, 2, 2)$ on W?

$$proj_{v_1} v = \frac{\langle v, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = \frac{4}{6}(1, 2, 1) = \frac{2}{3}(1, 2, 1).$$

$$proj_{v_2} v = \frac{\langle v, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 = -\frac{2}{3}(1, -1, 1).$$

$$\text{Hence} \quad proj_W(v) = Proj_{v_1}(v) + Proj_{v_2}(v)$$
$$= \frac{2}{3}(1, 2, 1) - \frac{2}{3}(1, -1, 1)$$
$$= (0, 2, 0).$$

Projection as a linear transformation

Let V be an inner product space and W be a subspace.

Then the projection of vectors in V to W is a linear transformation from V to V with image W.

Choose o.n. basis $\{w_1, \ldots, w_n\}$ for W.

$P_W(v) = proj_W(v)$.

$P_W(v_1 + v_2) = P_W(v_1) + P_W(v_2)$ & $P_W(cv) = c\,P_W(v)$

$$P_W(v_1+v_2) = proj_W(v_1+v_2) = \sum_{i=1}^{n} proj_{w_i}(v_1+v_2) = \sum_{i=1}^{n} \langle v_1+v_2, w_i \rangle w_i = \sum_{i=1}^{n} \langle v_1, w_i \rangle w_i + \sum_{i=1}^{n} \langle v_2, w_i \rangle w_i$$

$$P_W(cv) = proj_W(cv) = \sum_{i=1}^{n} \langle cv, w_i \rangle w_i = c \sum_{i=1}^{n} \langle v, w_i \rangle w_i$$

Denote this linear transformation as P_W.

Some properties of the projection Pw :

The linear transformation P_W has some interesting properties (some of which actually characterize it) :

$\{w_1, \ldots, w_n\}$ o.n. basis

$\sum_{i=1} \langle v, w_i \rangle w_i$

If $v \in W$ then $v = \sum c_i w_i$

i) $P_W(v) = v$, for all $v \in W$.

ii) Image ~~Range~~$(P_W) = W$.

iii) $W^{\perp} = \{v \mid v \in V, \text{ such that } \langle v, w \rangle = 0 \forall w \in W\}$ is the null space of P_W.

iv) $P_W^2 = P_W$. $\quad P_W \circ P_W \quad P_W(\underbrace{P_W(v)}_{\in W}) = P_W(v)$.

v) $||P_W(v)|| \leq ||v||$.

The Gram-Schmidt process :

In an inner product space

Any basis		Orthonormal basis
$x_1, x_2, \ldots, x_n$	$\xrightarrow[\text{process}]{\text{Gram-Schmidt}}$	$v_1, v_2, \ldots, v_n$

Example and intuition

Consider the basis $\beta = \{(1,2,2), (-1,0,2), (0,0,1)\}$ for $\mathbb{R}^3$. Can we use this to obtain an orthonormal basis for $\mathbb{R}^3$?

Let $v_1 = (1,2,2)$. We want a vector which is orthogonal to v_1, i.e. a vector in $\langle v_1 \rangle^{\perp}$, so we use the projection P_{v_1} to v_1.

$$\text{Define } v_2 = (-1,0,2) - P_{v_1}((-1,0,2)) = \underbrace{(I - P_{v_1})(-1,0,2)}_{\in \langle v_1 \rangle^{\perp}}$$

$$= (-1,0,2) - \frac{\langle (-1,0,2), (1,2,2) \rangle}{\langle (1,2,2), (1,2,2) \rangle}(1,2,2)$$

$$= \left(-\frac{4}{3}, -\frac{2}{3}, \frac{4}{3}\right).$$

$\boxed{\langle v_2, v_1 \rangle = 0}$

$W \quad W^{\perp} = \{v \mid \langle v, w \rangle = 0 \ \forall w \in W\}$ = Null space of P_W.

$P_W(v) = 0 \iff w \in W^{\perp}$.

$W^{\perp} \ni \boxed{(I - P_W)(v)} = v - P_W(v)$ & $P_W(v - P_W(v)) = 0$.

Example and intuition (contd.)

We want a vector which is orthogonal to both v_1 and v_2, i.e. a vector in $Span(\{v_1, v_2\})^{\perp}$, so we use the projection $P_{Span(\{v_1,v_2\})}$ to $Span(\{v_1, v_2\})$.

$$\begin{aligned}
\text{Define } v_3 &= (0,0,1) - P_{v_1}((0,0,1)) - P_{v_2}((0,0,1)) \\
&= (0,0,1) - \frac{\langle (0,0,1)(1,2,2) \rangle}{\langle (1,2,2),(1,2,2) \rangle}(1,2,2) \\
&\quad - \frac{\langle (0,0,1), (-\frac{4}{3}, -\frac{2}{3}, \frac{4}{3}) \rangle}{\langle (-\frac{4}{3}, -\frac{2}{3}, \frac{4}{3}), (-\frac{4}{3}, -\frac{2}{3}, \frac{4}{3}) \rangle}\left(-\frac{4}{3}, -\frac{2}{3}, \frac{4}{3}\right) \\
&= \left(\frac{2}{9}, -\frac{2}{9}, \frac{1}{9}\right).
\end{aligned}$$

Check $\langle v_1, v_2 \rangle = \langle v_1, v_3 \rangle = \langle v_2, v_3 \rangle = 0$.

Thus $\{v_1, v_2, v_3\}$ is an orthogonal basis and dividing each vector by its norm yields an orthonormal basis

$$\left\{ \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right), \left(-\frac{2}{3}, -\frac{1}{3}, \frac{2}{3}\right), \left(\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\right) \right\}.$$

The Gram-Schmidt process

Let V be an inner product space with a basis $\{x_1, x_2, \ldots, x_n\}$. Define the orthogonal basis $\{v_1, v_2, \ldots, v_n\}$ and the corresponding orthonormal basis $\{w_1, w_2, \ldots, w_n\}$ as follows :

$$\mathrm{Span}(\{x_1, x_2, \ldots, x_{i-1}\}) = \mathrm{Span}(\{v_1, v_2, \ldots, v_{i-1}\}) = \mathrm{Span}(\{w_1, w_2, \ldots, w_{i-1}\})$$

$$v_1 = x_1 \; ; \; w_1 = \frac{v_1}{||v_1||}$$

$$v_2 = x_2 - \langle x_2, w_1 \rangle w_1 \quad ; \quad w_2 = \frac{v_2}{||v_2||}$$

$$\vdots \quad \vdots \quad \vdots$$

$$v_i = x_i - \boxed{\langle x_i, w_1 \rangle w_1 - \langle x_i, w_2 \rangle w_2 - \ldots - \langle x_i, w_{i-1} \rangle w_{i-1}} \; ; \; w_i = \frac{v_i}{||v_i||}$$

$$= P_{\mathrm{Span}(\{w_1, \ldots, w_{i-1}\})}(x_i) = P_{\mathrm{Span}(\{x_1, \ldots, x_{i-1}\})}(x_i)$$

$$\vdots \quad \vdots \quad \vdots$$

$$v_n = x_n - \langle x_n, w_1 \rangle w_1 - \langle x_n, w_2 \rangle w_2 - \ldots$$

$$\ldots - \langle x_n, w_{n-1} \rangle w_{n-1} \quad ; \quad w_n = \frac{v_n}{||v_n||}$$

Theorem: Any finite-dimensional vector space with an inner product has an orthonormal basis.

Any basis can be changed to an orthonormal basis using the Gram-Schmidt process.

Orthogonal transformations and rotations :

What are orthogonal transformations :

Let V be an inner product space and T be a linear transformation from V to V. T is said to be orthogonal transformation if

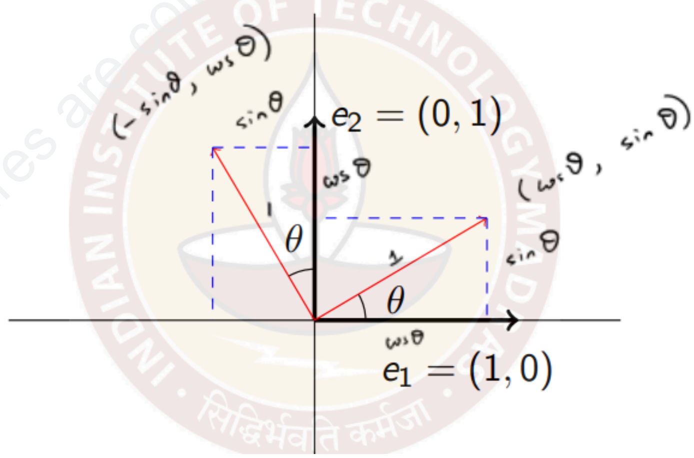
$$\langle Tv, Tw \rangle = \langle v, w \rangle \ \forall v, w \in V.$$

When $V = \mathbb{R}^n$ with the usual inner product, a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal if it preserves angles and lengths.

Fact : It is enough to demand that the linear transformation preserves lengths. In that case, angles automatically get preserved (think of triangle congruences).

Finding the rotation matrix in $\mathbb{R}^2$

Consider the standard basis$\{(1,0),(0,1)\}$ of $\mathbb{R}^2$. Rotate the plane by an angle θ. The vectors obtained after rotation tell us the matrix corresponding to this linear transformation.



Let T_θ be the corresponding linear transformation. Then

$$T_\theta(1,0) = (cos(\theta), sin(\theta)) \quad \text{and} \quad T_\theta(0,1) = (-sin(\theta), cos(\theta)).$$

Thus the matrix corresponding to this linear transformation is

$$R_\theta = \begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix}.$$

$$R_{-\theta} = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix} = R_\theta^T$$

Note that $R_\theta^T = R_{-\theta}$ and $\boxed{R_\theta^T R_\theta = R_\theta R_\theta^T = I.}$

Further note that since angles and lengths ~~and angles~~ are preserved and the standard basis is orthonormal, the rotated vectors are also orthonormal and therefore yield an orthonormal basis of $\mathbb{R}^2$.

Rotations in $\mathbb{R}^3$

Consider the rotations about the axes in $\mathbb{R}^3$. Since these clearly preserve angles and distances and are linear transformations, they are orthogonal transformations.

Rotations about the axes can be described by considering its effect on the standard basis $\{e_1, e_2, e_3\}$.

When considering the rotation about the Z-axis, e_3 remains unchanged and the XY-plane gets rotated exactly as in the previous case of $\mathbb{R}^2$. Therefore its matrix is

$$T_3(\theta) = \begin{pmatrix} cos(\theta) & -sin(\theta) & 0 \\ sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Similarly, the matrix corresponding to rotation about the X-axis is

$$T_1(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\theta) & -sin(\theta) \\ 0 & sin(\theta) & cos(\theta) \end{pmatrix}$$

and the matrix corresponding to rotation about the Y-axis is

$$T_2(\theta) = \begin{pmatrix} cos(\theta) & 0 & -sin(\theta) \\ 0 & 1 & 0 \\ sin(\theta) & 0 & cos(\theta) \end{pmatrix}.$$

Notice : $T_i(\theta)^T = T_i(-\theta)$ and $T_i(\theta)^T T_i(\theta) = T_i(\theta) T_i(\theta)^T = I$.

Another example of an orthogonal transformation

Let us define a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$, where

$$T(x_1, x_2, x_3) = \frac{1}{3}(x_1 - 2x_2 + 2x_3, 2x_1 - x_2 - 2x_3, 2x_1 + 2x_2 + x_3).$$

Then evaluating T on the standard basis $\{e_1, e_2, e_3\}$ yields :

$\|Tx\|$
$\|$
$\|x\|$

$$T(e_1) = v_1 = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$$
$$T(e_2) = v_2 = \left(-\frac{2}{3}, -\frac{1}{3}, \frac{2}{3}\right)$$
$$T(e_3) = v_3 = \left(\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\right).$$

Thus, the matrix corresponding to T is $A = \frac{1}{3}\begin{pmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{pmatrix}$.

Orthogonal matrices :

As $\{v_1, v_2, v_3\}$ is an ~~orthogonal~~ normal set, the linear transformation T is an orthogonal transformation.

Observe that the $AA^T = A^TA = I_3$.

A square matrix A is called an orthogonal matrix if $\boxed{AA^T = A^TA = I}$

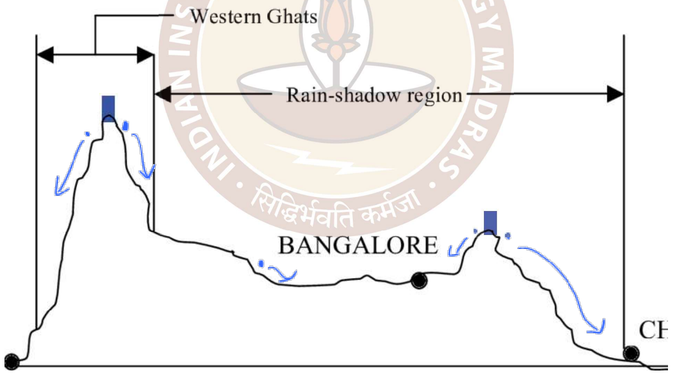
$$\|Tx\| = \langle Tx, Tx\rangle = (Tx)\cdot(Tx) = (Ax)^TAx = x^TA^TAx = x^Tx = \|x\|.$$

week- 10 ....

Tracing water flowing down a hill

For an electrification project, Mohan Bhargava is trying to understand how water flows down a particular hill.

How does water flow downhill? It will move in the direction where the altitude decreases most rapidly.



Tracing water flowing down a hill (contd.)

Mohan models the hill as in the previous picture using the graph of a function $h(x)$ where h is the altitude of a point.

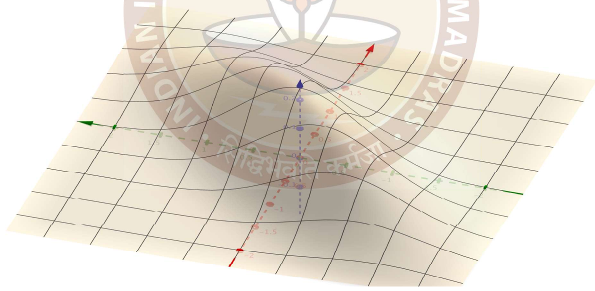
He then calculates the derivative $h'(x)$ and computes at which points it is negative, positive and 0.

- If the derivative is negative at a point, water flows to the right from that point.
- If the derivative is positive at a point, water flows to the right from that point.
- If the derivative is 0 at a point, water will remain stationary at that point.

Tracing water flowing down a hill (contd.)

Mohan's friend Gita points out that a one dimensional model (i.e. a cross section) will not be useful and that a two dimensional model will be more useful.

She uses a 2-dimensional function $h(x, y)$ to model the altitude h.



Tracing water flowing down a hill (contd.)

Mohan asks her how she will find out the direction in which the water will flow since there are now more than 2 directions to contend with.

Gita tells Mohan that water will flow in the direction in which the altitude decreases fastest, i.e. along the steepest slope downward.

She says that to compute it, we have to find the direction in which the function h decreases fastest or equivalently the rate of *decrease* of h is fastest.

This is same as finding the vector u in which the directional derivative h_u is largest in absolute value amongst those for which it has negative sign i.e. u such that

1. $h_u \leq h_v$ for all $v \in \mathbb{R}^2$ and
2. $h_u < 0$.

In what direction is the directional derivative minimized?

Let $f(x_1, x_2, \dots, x_n)$ be a function defined on a domain D in $\mathbb{R}^n$ containing some open ball around the point $\underset{\sim}{a}$.

Suppose ∇f exists and is continuous on some open ball around the point $\underset{\sim}{a}$.

$$f_u = \nabla f(\underset{\sim}{a}) \cdot u = \|\nabla f(\underset{\sim}{a})\| \, \|u\| \cos(\theta) = \|\nabla f(\underset{\sim}{a})\| \cos(\theta)$$

where θ is the angle between $\nabla f(\underset{\sim}{a})$ and u.

$\cos(\theta)$ is minimized when $\theta = \pi$ i.e. u is pointing in the direction opposite to $\nabla f(\underset{\sim}{a})$.

$\therefore$ The minimum value of f_u is attained when $u = -\nabla f(\underset{\sim}{a}) / \|\nabla f(\underset{\sim}{a})\|$ & is equal to $-\|\nabla f(\underset{\sim}{a})\|$.

Directions in which the directional derivative is maximized or remains unchanged

Assume the same hypothesis as the previous slide.

$$f_u = \|\nabla f(\underset{\sim}{a})\| \, \|u\| \cos(\theta) = \|\nabla f(\underset{\sim}{a})\| \cos(\theta)$$

It is maximized when $\theta = 0$, i.e. u is in the same direction as $\nabla f(\underset{\sim}{a})$ i.e. $u = \dfrac{\nabla f(\underset{\sim}{a})}{\|\nabla f(\underset{\sim}{a})\|}$.

It remains unchanged when $f_u = 0$ i.e. $\theta = \pi/2$ i.e. u is orthogonal / perpendicular to $\nabla f(\underset{\sim}{a})$.

Directions : steepest ascent, steepest descent, no change

Assume the same hypothesis as the previous slide.

Property	In terms of directional derivatives	Direction
Steepest ascent	f_u is positive and maximum	$u = \nabla f/\|\|\nabla f\|\|$
Steepest descent	f_u is negative and minimum	$u = -\nabla f/\|\|\nabla f\|\|$
No change	$f_u = 0$	u is orthogonal to ∇f

Examples

1. $f(x, y) = sin(xy)$

   $\nabla f(x,y) = (y\cos(xy),\ x\cos(xy))$

   At $(\pi,1)$ what is the direction of steepest descent on the graph of this function?

   $\nabla f(\pi,1) = (1 \times \cos(\pi),\ \pi \times \cos(\pi)) = (-1, -\pi)$.

   $u = \dfrac{-\nabla f(\pi,1)}{\|\nabla f(\pi,1)\|} = \dfrac{(1,\pi)}{\sqrt{1+\pi^2}}$.

2. $f(x, y, z) = x^2 + y^2 + z^2$

   $\nabla f(x,y,z) = (2x, 2y, 2z)$.

   At $(1,1,1)$ what is the direction in which the fn. increases fastest?

   $\nabla f(1,1,1) = (2,2,2)$. In the direction $u = \dfrac{(2,2,2)}{\sqrt{2^2+2^2+2^2}} = \dfrac{1}{\sqrt{3}}(1,1,1)$.

   In which direction does the fn. remain constant?

   e.g. $(1,-1,0)\frac{1}{\sqrt{2}}$, $\frac{1}{\sqrt{2}}(1,0,-1)$.

A cautionary tale

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

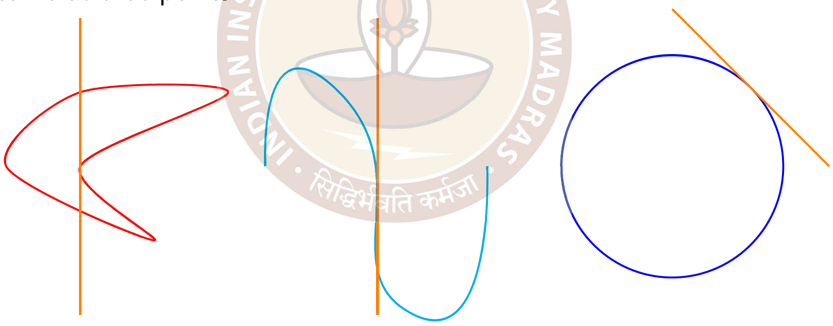
$\nabla f(0,0) = (0,0)$.

$f_u(0,0)$ DNE unless $u = e_1$ or e_2.

Recall : tangent lines to curves

A tangent line to a curve C at a point p (on C) is a line which represents the *instantaneous* direction in which the curve C moves at the point p.

Traditionally, it was thought of as a line which *just touches* the curve at that point.



The tangent line to a function of one variable

Let $f : D \to \mathbb{R}$ be a function where D is a subset of $\mathbb{R}$. Assume that $\Gamma(f)$, the graph of f is a curve. Let $x \in D$.

Then a tangent (line) to f at x is a tangent (line) to $\Gamma(f)$ at the point $(x, f(x))$.

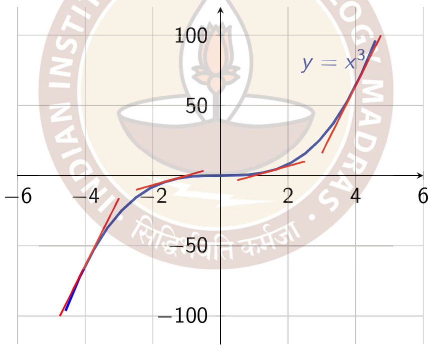


Figure: Tangent lines for $y = x^3$

The equations of the tangent line

$$x(t) = a + tu_1,\quad y(t) = b + tu_2,\quad z(t) = f(a,b) + t f_u(a,b).$$ Parametric eqns.

$$\frac{x-a}{u_1} = \frac{y-b}{u_2} = \frac{z-f(a,b)}{f_u(a,b)}$$ Symm. eqns.

$$(x(t), y(t), z(t)) = (a, b, f(a,b)) + t\boxed{(u_1, u_2, f_u(a,b))}.$$

vector form.

$u = (u_1, u_2)$ unit vector on L.
$\underset{\sim}{a} = (a, b)$ is the point.
$f_u(a,b)$ is the directional derivative at (a, b).

L: $z = 0$, $u_1(y-b) = u_2(x-a)$.

P: $u_1(y-b) = u_2(x-a)$.

$$x(t) = a + tu_1,\quad y(t) = b + tu_2,\quad z(t) = 0.$$

$$(x(t), y(t), z(t)) = (a, b, 0) + t(u_1, u_2, 0).$$

Examples

$f(x,y) = x + y$; tangent at $(1,1)$ in the direction of $(1,0)$

$f_u(1,1) = \frac{\partial f}{\partial x}(x,y) = 1.$

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

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

$f(x,y) = xy$; tangent at $(1,1)$ in the direction of $(3,4)$

$u = (\frac{3}{5}, 4/5).$ $f_u(1,1) = 1 \times \frac{3}{5} + 1 \times 4/5 = 7/5.$ $\nabla f(x,y) = (y,x)$; $\nabla f(1,1) = (1,1).$

$(x(t), y(t), z(t)) = (1,1,1) + t(3/5, 4/5, 7/5)$
$= (1 + 3t/5, 1 + 4t/5, 1 + 7t/5).$

$f(x,y) = sin(xy)$; tangent at $(\pi,1)$ in the direction of $(1,2)$

$u = \frac{1}{\sqrt{5}}(1,2).$ $\nabla f(x,y) = (y\cos(xy), x\cos(xy))$. $\therefore \nabla f(\pi,1) = (-1,-\pi).$

$f_u(\pi,1) = (-1) \times \frac{1}{\sqrt{5}} + (-\pi) \times \frac{2}{\sqrt{5}} = \frac{-2\pi-1}{\sqrt{5}}.$

$(x(t), y(t), z(t)) = (\pi, 1, 0) + t(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}, \frac{-2\pi-1}{\sqrt{5}}).$

Tangents for scalar-valued multivariable functions

Let $f(\underset{\sim}{x})$ be a function defined on a domain D in $\mathbb{R}^n$ containing some open ball around the point $\underset{\sim}{a}$.

Consider a line L in D passing through $\underset{\sim}{a}$ and restrict f to L.

Since it is now a function of one variable, we can consider the tangent to f at $\underset{\sim}{a}$ over L as before.

If the line L is in the direction of the unit vector u, then the tangent (if it exists) will be the line with slope $f_u(\underset{\sim}{a})$ and passing through the point $(\underset{\sim}{a}, f(\underset{\sim}{a}))$.

Parametric equations and an example

Similar to the two-variable case, we can deduce the equations of the tangent line as :

$\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$, z is the variable in which we are measuring the fn.

$\underset{\sim}{a} = (a_1, \ldots, a_n)$, $u = (u_1, \ldots, u_n)$.

Line through $\underset{\sim}{a}$ in the direction of u is $\boxed{(\underset{\sim}{x}(t), z(t)) = (\underset{\sim}{a}, 0) + t(u, 0).}$ $x_i(t) = a_i + tu_i$, $z = 0$. L

$\therefore$ The tangent line to f at $\underset{\sim}{a}$ above L is $\boxed{(\underset{\sim}{x}(t), z(t)) = (\underset{\sim}{a}, f(\underset{\sim}{a})) + t(u, f_u(\underset{\sim}{a})). \quad x_i(t) = a_i + tu_i, \ z(t) = f(\underset{\sim}{a}) + t f_u(\underset{\sim}{a}).}$

Example : $f(x,y) = xy + yz + zx$; tangent at $(1,1,1)$ in the direction $(-1,-2,2)$

$u = \frac{1}{3}(-1,-2,2).$ $\nabla f(x,y,z) = (y+z, z+x, x+y)$; $\nabla f(1,1,1) = (2,2,2).$

$f_u(1,1,1) = -2/3$

$(x(t), y(t), z(t), u(t)) = (1,1,1,3) + t(\frac{-1}{3}, \frac{-2}{3}, \frac{2}{3}, -2/3)$

$x(t) = 1 - t/3, \ y(t) = 1 - 2t/3, \ z(t) = 1 + 2t/3, \ u(t) = 3 - 2t/3.$

Caution : tangents need not always exist.

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

$$f_u(0,0) = \begin{cases} 0 & \text{if } (u_1,u_2) = \pm e_1 \text{ or } \pm e_2 \\ \text{DNE} & \text{o.w.} \end{cases}$$

The tangent lines in all directions other than along the x or y-axes at (0,0) DNE.

$$f(x,y) = |x| + |y|$$

For many directions at many points, the tangent line will not exist.

When do all the tangents exist?

This is equivalent to asking when do all the directional derivatives exist.

Let $f(x_1, x_2, \ldots, x_n)$ be a function defined on a domain D in $\mathbb{R}^n$ containing some open ball around the point $\underset{\sim}{a}$.

Theorem

Suppose ∇f exists and is continuous on some open ball around the point $\underset{\sim}{a}$. Then for every unit vector u, the directional derivative $f_u(\underset{\sim}{a})$ exists and equals $\nabla f(\underset{\sim}{a}) \cdot u$.

Conclusion : All the tangents at a point $\underset{\sim}{a}$ exist when ∇f exists and is continuous on some open ball around the point $\underset{\sim}{a}$.

Recall : Tangent lines for $f(x,y)$

Let $f(x,y)$ be a function defined on a domain D in $\mathbb{R}^2$ containing some open ball around the point $\underset{\sim}{a}$.

The tangent line above L passing through $\underset{\sim}{a}$ is the tangent to the function f obtained by restricting f to L and considering its tangent as a function of one variable.

If u is a unit vector in the direction of the line L, then the tangent (if it exists) will be the line with slope $f_u(\underset{\sim}{a})$ passing through the point $(\underset{\sim}{a}, f(\underset{\sim}{a}))$ and so its parametric equation is :

$$x(t) = a + tu_1, \quad y(t) = b + tu_2, \quad z(t) = f(a,b) + t f_u(a,b)$$

$u = (u_1, u_2)$

$\underset{\sim}{a} = (a, b).$

The collection of all tangents

Let $f(x,y)$ be a function defined on a domain D in $\mathbb{R}^2$ containing some open ball around the point (a,b).

Suppose ∇f exists and is continuous on some open ball around the point (a,b).

Then all the tangent lines at the point (a,b) exist and we can rewrite the equation of a tangent line in the direction of the unit vector u as :

$$f_u(a,b) = \nabla f(a,b)\cdot u = \frac{\partial f}{\partial x}(a,b)u_1 + \frac{\partial f}{\partial y}(a,b)u_2.$$

$$x(t) = a + u_1 t, \quad y(t) = b + u_2 t, \quad z(t) = f(a,b) + f_u(a,b)\,t = f(a,b) + \left(\frac{\partial f}{\partial x}(a,b)u_1 + \frac{\partial f}{\partial y}(a,b)u_2\right)t$$

$$z(t) - f(a,b) = \frac{\partial f}{\partial x}(a,b)(x(t)-a) + \frac{\partial f}{\partial y}(a,b)(y(t)-b)$$

Tangent lines in terms of linear algebra for $f(x,y)$

$$(x(t), y(t), z(t)) = (a,b,f(a,b)) + t\,(u_1, u_2, f_u(a,b)).$$

Tangent line to f at (a,b) in the direction of u $= (a,b,f(a,b)) + W_u$

W_u: Line passing through the vector $(u_1, u_2, f_u(a,b))$.

$$f_u(a,b) = u_1\frac{\partial f}{\partial x}(a,b) + u_2\frac{\partial f}{\partial y}(a,b)$$

The lines W_u all lie on the plane P: $z = \frac{\partial f}{\partial x}(a,b)\,x + \frac{\partial f}{\partial y}(a,b)\,y$.

Tangent plane of f at (a,b) $= (a,b,f(a,b)) + P$.

The equation of the tangent plane

Let $f(x,y)$ be a function defined on a domain D in $\mathbb{R}^2$ containing some open ball around the point (a,b).

Suppose ∇f exists and is continuous on some open ball around the point (a,b).

Then the equation of the tangent plane to f at (a,b) is given by :

$$z = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b).$$

$$\frac{\partial f}{\partial x}(a,b)\,x + \frac{\partial f}{\partial y}(a,b)\,y - z = \frac{\partial f}{\partial x}(a,b)\,a + \frac{\partial f}{\partial y}(a,b)\,b.$$

Examples

$f(x,y) = x + y$; tangent at $(1,1)$ $\nabla f(1,1) = (1,1)$.

$$z = 2 + 1(x-1) + 1(y-1)$$
$$z = 2 + x - 1 + y - 1 = x + y.$$
$$z = x + y.$$

$f(x,y) = xy$; tangent at $(1,1)$ $\nabla f(1,1) = (1,1)$.

$$z = 1 + 1(x-1) + 1(y-1)$$
$$= x + y - 1.$$

$f(x,y) = sin(xy)$; tangent at $(1,0)$ $\nabla f(1,0) = (0,1)$.

$$z = 0 + 0(x-1) + 1(y-0)$$
$$= y.$$

Eqn. : $z = y$.

The tangent hyperplane

Let $f(\underset{\sim}{x})$ be a function defined on a domain D in $\mathbb{R}^n$ containing some open ball around the point $\underset{\sim}{a}$.

Suppose ∇f exists and is continuous on some open ball around the point $\underset{\sim}{a}$.

$$(\underset{\sim}{x}(t), z(t)) = (\underset{\sim}{a}, f(\underset{\sim}{a})) + t(u, \boxed{f_u(\underset{\sim}{a})})$$

$$\sum u_i \frac{\partial f}{\partial x_i}(\underset{\sim}{a})$$

$$z = \sum x_i \frac{\partial f}{\partial x_i}(\underset{\sim}{a})$$

Then the equation of the tangent hyperplane to f at (a,b) is given by :

$$z = f(\underset{\sim}{a}) + \sum \frac{\partial f}{\partial x_i}(\underset{\sim}{a})(x_i - a_i)$$
$$= f(\underset{\sim}{a}) + \nabla f(\underset{\sim}{a}).(\underset{\sim}{x} - \underset{\sim}{a}).$$

Examples

$f(x,y,z) = xy + yz + zx$; tangent at $(1,1,1)$

$$\nabla f(x,y,z) = (x+y, y+z, z+x)$$
$$\nabla f(1,1,1) = (2,2,2).$$

Tangent hyperplane eqn. is :

$$u = f(1,1,1) + \nabla f(1,1,1) . (x-1, y-1, z-1)$$
$$= 3 + (2,2,2) . (x-1, y-1, z-1)$$

Eqn. is $u = 3 + 2(x-1) + 2(y-1) + 2(z-1)$.

$f(x,y,z) = x^2 + y^2 + z^2$; tangent at $(2,3,-1)$

$$\nabla f = (2x, 2y, 2z)$$
$$\nabla f(2,3,-1) = (4,6,-2).$$

$$u = f(2,3,-1) + \nabla f(2,3,-1) . (x-2, y-3, z+1)$$
$$= 14 + (4,6,-2) . (x-2, y-3, z+1)$$

Eqn. is : $u = 14 + 4(x-2) + 6(y-3) - 2(z+1)$.

Linear approximation

Let $f(\underset{\sim}{x})$ be a function defined on a domain D in $\mathbb{R}^n$ containing some open ball around the point $\underset{\sim}{a}$.

Suppose ∇f exists and is continuous on some open ball around the point $\underset{\sim}{a}$.

Then the function $L_f(\underset{\sim}{x}) = f(\underset{\sim}{a}) + \nabla f(\underset{\sim}{a}) \cdot (\underset{\sim}{x} - \underset{\sim}{a})$ is the best linear approximation for the function f close to $\underset{\sim}{a}$.

Examples

Linear approximation to $f(x, y) = xy$ at $(1, 1)$

$\nabla f(x,y) = (y,x)$

$\nabla f(1,1) = (1,1)$.

$$L_f(x,y) = f(1,1) + \nabla f(1,1)\cdot(x-1,y-1)$$
$$= 1 + (1,1)\cdot(x-1,y-1)$$
$$= 1 + x-1+y-1 = x+y-1.$$

is the best linear approx. to f close to $(1,1)$.

Linear approximation to $f(x, y, z) = x^2 + y^2 + z^2$ at $(2, 3, -1)$

$\nabla f(2,3,-1) = (4,6,-2)$.

$$L_f(x,y,z) = 3 + (4,6,-2)\cdot(x-2,\, y-3,\, z+1)$$
$$= 3 + 4(x-2) + 6(y-3) - 2(z+1)$$
$$= 4x + 6y - 2z - 29.$$

is the best linear approx. to f close to $(2,3,-1)$.

Recall : Critical points for functions of one variable

A point a is called a critical point of a function $f(x)$ if either f is not differentiable at a or $f'(a) = 0$.

If f is differentiable at a point a of local extremum, it satisfies $f'(a) = 0$ and so every point of local extremum is a critical point.

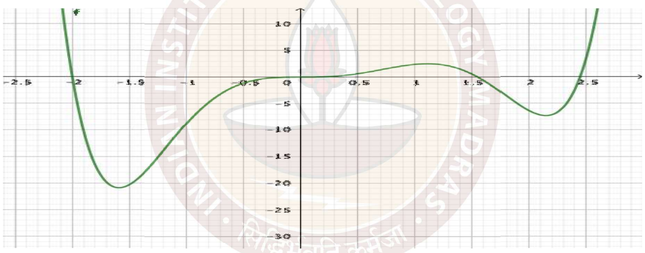


Figure: $f(x) = (x^2 - 4x + 3.8)(x + 2)x^3$

Not every critical point is a point of local extremum. A saddle point is a critical point which is not a point of local extremum.

Points of local extrema for multivariable functions

Let $f(\underset{\sim}{x})$ be a function defined on a domain D in $\mathbb{R}^n$ and suppose $\underset{\sim}{a} \in D$.

The point $\underset{\sim}{a}$ is a local maximum (or point of local maximum) of f if for some open ball B containing $\underset{\sim}{a}$, $f(\underset{\sim}{x}) \le f(\underset{\sim}{a})$ whenever $\underset{\sim}{x} \in B \cap D$.

The point $\underset{\sim}{a}$ is a local minimum (or point of local minimum) of f if for some open ball B containing $\underset{\sim}{a}$, $f(\underset{\sim}{a}) \le f(\underset{\sim}{x})$ whenever $\underset{\sim}{x} \in B \cap D$.

A local extremum (or point of local extremum) of f is either a local maximum or a local minimum of f.

The gradient vector at points of local extrema

Let $f(\underset{\sim}{x})$ be a function defined on a domain D in $\mathbb{R}^n$ containing some open ball around a point $\underset{\sim}{a}$ of local extremum.

Restrict f to a line L passing through $\underset{\sim}{a}$ and view it as a function of one variable on L.

Then $\underset{\sim}{a}$ is a local extremum for the restricted function on L and hence the directional directive of f in the direction of the line L (if it exists) at $\underset{\sim}{a}$ is 0.

In particular, those partial derivatives which exist at $\underset{\sim}{a}$ must be 0.

If $\nabla f(\underset{\sim}{a})$ exists for a local extremum $\underset{\sim}{a}$, then $\nabla f(\underset{\sim}{a}) = 0$. ← vector

Critical points

A point $\underset{\sim}{a}$ is called a critical point of a function $f(\underset{\sim}{x})$ if either $\nabla f(\underset{\sim}{a})$ does not exist or $\nabla f(\underset{\sim}{a})$ exists and $\nabla f(\underset{\sim}{a}) = 0$.

Example : Critical points of $f(x, y) = x^2 + 6xy + 4y^2 + 2x - 4y$.

$$\frac{\partial f}{\partial x} = 2x + 6y + 2. \qquad \frac{\partial f}{\partial y} = 6x + 8y - 4.$$

$$\nabla f(x, y) = (2x + 6y + 2, 6x + 8y - 4).$$

Set $\nabla f = 0$ i.e. $(2x + 6y + 2, 6x + 8y - 4) = (0, 0)$.

$$2x + 6y + 2 = 0$$
$$6x + 8y - 4 = 0$$

$$\left[\begin{array}{cc|c} 2 & 6 & -2 \\ 6 & 8 & 4 \end{array}\right] \xrightarrow{R_1/2} \left[\begin{array}{cc|c} 1 & 3 & -1 \\ 6 & 8 & 4 \end{array}\right] \xrightarrow{R_2 - 6R_1} \left[\begin{array}{cc|c} 1 & 3 & -1 \\ 0 & -10 & 10 \end{array}\right] \xrightarrow{R_2/-10} \left[\begin{array}{cc|c} 1 & 3 & -1 \\ 0 & 1 & -1 \end{array}\right] \xrightarrow{R_1 - 3R_2} \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & -1 \end{array}\right]$$

$\Rightarrow x = 2, y = -1$.

$\therefore$ Critical pt. of f is $(2, -1)$.

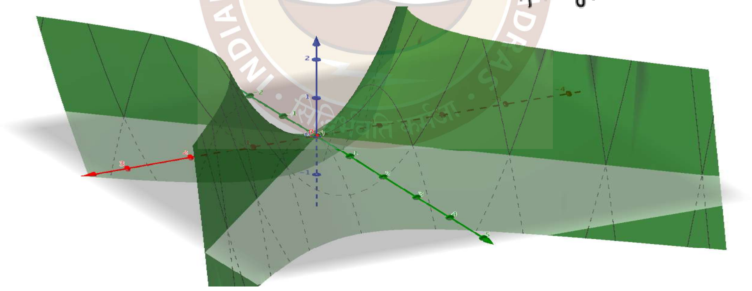
Saddle points

Every local extremum is a critical point. Unfortunately, not all critical points are local extrema.

Example : $f(x, y) = x^3$.

A saddle point is a critical point $\underset{\sim}{a}$ such that $\nabla f(\underset{\sim}{a})$ exists and $\nabla f(\underset{\sim}{a}) = 0$ but $\underset{\sim}{a}$ is not a local extremum.

$f(x,y) = x^2 - y^2$



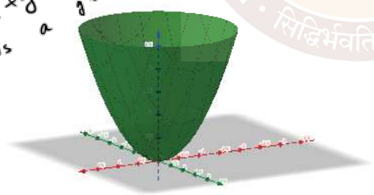
Absolute (or global) extrema

Let $f(\underset{\sim}{x})$ be a function defined on a domain D in $\mathbb{R}^n$ and suppose $\underset{\sim}{a} \in D$.

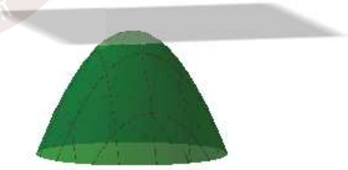
The point $\underset{\sim}{a}$ is an absolute maximum (or global maximum) of f if $f(\underset{\sim}{x}) \leq f(\underset{\sim}{a})$ for all $\underset{\sim}{x} \in D$.

The point $\underset{\sim}{a}$ is an absolute minimum (or global minimum) of f if $f(\underset{\sim}{a}) \leq f(\underset{\sim}{x})$ for all $\underset{\sim}{x} \in D$.

x^2+y^2 $(0,0)$ is a global minimum



$-x^2-y^2$ $(0,0)$ is a global maximum.



Existence of absolute maximum/minimum

A domain D in $\mathbb{R}^n$ is called closed if it contains all its boundary points. A domain D in $\mathbb{R}^n$ is called bounded if it is contained inside a ball around 0 with finite radius.

Fact : If the domain D is closed and bounded and f is continuous on D, then the global maximum and minimum must exist.

Note that the global maximum and minimum are in particular local maxima or local minima unless they are on boundary points.

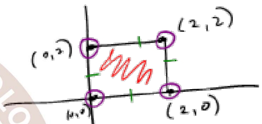
Thus to find the global maximum and minimum, we find the critical points

- inside the domain D
- on the boundary of D
- on the boundary of the boundary of D

. . . and check the value of f on all of them.

Example

Find the absolute maximum and minimum of the function $f(x,y) = x^3 + y^3 - 3x - 3y^2 + 1$ over the square with vertices $(0,0), (2,0), (2,2), (0,2)$.



$\nabla f = (3x^2 - 3,\ 3y^2 - 6y).$

Set to 0. $3x^2 - 3 = 0,\ 3y^2 - 6y = 0.$

$x^2 = 1,\ y(y-2) = 0.$

$x = \pm 1,\ y = 0$ or $2.$

$(1,0),\ (1,2),\ (0,0),\ (0,2),\ (2,2),\ (2,0)$

$-1 \quad -5 \quad 1 \quad -3 \quad -1 \quad 3$

$f(x,0) = x^3 - 3x + 1.\ :\ 0 < x < 2$

$f'(x) = 0 \Rightarrow 3x^2 - 3 = 0 \Rightarrow x = \pm 1 \Rightarrow (1,0).$

$f(x,2) = x^3 + 8 - 3x - 12 + 1 = x^3 - 3x - 3.$

$f'(x) = 0 \Rightarrow$ ✗

$0 \le y \le 2$

$f(0,y) = y^3 - 3y^2 + 1.$

$f'(y) = 0$

$3y^2 - 6y = 0$

$y(y-2) = 0.$

$(0,0), (0,2).$

$f(2,y) = y^3 - 3y^2 + c.$

Abs. max. occurs at $(2,0)$

" min. " " $(1,2)$

week-11 ....

Recall : Partial derivatives

Let $f(x_1, x_2, \ldots, x_n)$ be a scalar-valued multivariable function defined on a domain D in $\mathbb{R}^n$.

The partial derivative of f w.r.t. x_i is the function denoted by $f_{x_i}\left(\underset{\sim}{x}\right)$ or $\frac{\partial f}{\partial x_i}\left(\underset{\sim}{x}\right)$ and defined as

$$\frac{\partial f}{\partial x_i}\left(\underset{\sim}{x}\right) = \lim_{h \to 0} \frac{f\left(\underset{\sim}{x} + he_i\right) - f\left(\underset{\sim}{x}\right)}{h} .$$

Its domain consists of those points of D at which the limits exists.

The partial derivative of f w.r.t. x_i at a point $\underset{\sim}{a}$ measures the rate of change of f at $\underset{\sim}{a}$ in the direction of the standard basis vector e_i (i.e. w.r.t. the variable x_i).

Second order partial derivatives for $f(x, y)$

Let $f(x, y)$ be a function defined on a domain D in $\mathbb{R}^2$.

Then the second order partial derivatives of f are the partial derivatives of the partial derivatives.

Notation :

- $f_{xx} = (f_x)_x$ or $\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)$
- $f_{yy} = (f_y)_y$ or $\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)$
- $\underline{\underline{f_{xy}}} = (f_x)_y$ or $\underline{\underline{\frac{\partial^2 f}{\partial y \partial x}}} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$
- $f_{yx} = (f_y)_x$ or $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$

Mixed partial derivatives

Examples

$f(x, y) = x + y$ $\qquad \frac{\partial f}{\partial x} = 1 \qquad \frac{\partial f}{\partial y} = 1.$

$\frac{\partial^2 f}{\partial x^2} = 0, \quad \frac{\partial^2 f}{\partial y^2} = 0, \quad \frac{\partial^2 f}{\partial x \partial y} = 0, \quad \frac{\partial^2 f}{\partial y \partial x} = 0.$

$f(x, y) = sin(xy)$ $\qquad \frac{\partial f}{\partial x} = y \cos(xy), \quad \frac{\partial f}{\partial y} = x \cos(xy).$

$\frac{\partial^2 f}{\partial x^2} = y \times \{-\sin(xy)\, y\} = -y^2 \sin(xy), \quad \frac{\partial^2 f}{\partial y^2} = -x^2 \sin(xy).$

$\frac{\partial^2 f}{\partial x \partial y} = 1 \times \cos(xy) + x \times \{y \sin(xy)\} = \cos(xy) - xy \sin(xy).$

$\frac{\partial^2 f}{\partial y \partial x} = 1 \times \cos(xy) + y \times \{-x \sin(xy)\} = \cos(xy) - xy \sin(xy)$

$=$

Clairaut's Theorem about mixed partials

Theorem (Clairaut's theorem)

Let $f(x, y)$ be a function defined on a domain D in $\mathbb{R}^2$ containing a point $\underset{\sim}{a}$ and an open ball around it.

Clairaut's Theorem about mixed partials

Theorem (Clairaut's theorem)

Let $f(x, y)$ be a function defined on a domain D in $\mathbb{R}^2$ containing a point $\underset{\sim}{a}$ and an open ball around it.

If the second order mixed partial derivatives f_{xy} and f_{yx} are continuous in an open ball around $\underset{\sim}{a}$, then $f_{xy}\left(\underset{\sim}{a}\right) = f_{yx}\left(\underset{\sim}{a}\right)$.

Example advising caution

$$f(x, y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h \to 0} \frac{f(0+h,0) - f(0,0)}{h} = \lim_{h \to 0} \frac{0-0}{h} = 0.$$

$$\frac{\partial f}{\partial y}(0,0) = 0.$$

$$\frac{\partial f}{\partial x}(x,y) = \frac{(x^2+y^2)\{y(x^2-y^2) + xy(2x)\} - xy(x^2-y^2)2x}{(x^2+y^2)^2} = \frac{-y^5 + x^4y + 4x^2y^3}{(x^2+y^2)^2}$$

if $(x,y) \neq (0,0)$.

$$\frac{\partial f}{\partial y}(x,y) = \frac{x^5 - xy^4 - 4x^3y^2}{(x^2+y^2)^2}$$

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial y}(h,0) - \frac{\partial f}{\partial y}(0,0)}{h} = \lim_{h \to 0} \frac{h^5/(h^2+0^2)^2 - 0}{h} = \lim_{h \to 0} \frac{h-0}{h} = 1.$$

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{h \to 0} \frac{\partial f/\partial x(0,h) - \frac{\partial f}{\partial x}(0,0)}{h} = \lim_{h \to 0} \frac{-h^5/(0^2+h^2)^2 - 0}{h} = -1.$$

Second order partial derivatives

Let $f(x_1, x_2, \ldots, x_n)$ be a function defined on a domain D in $\mathbb{R}^n$. Then the second order partial derivatives of f are defined analogously as the partial derivatives of the partial derivatives.

$$f_{x_i x_i} = (f_{x_i})_{x_i} \quad \text{or} \quad \frac{\partial^2 f}{\partial x_i^2} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_i}\right)$$

$$f_{x_i x_j} = (f_{x_i})_{x_j} \quad \text{or} \quad \frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right).$$

Example : $f(x, y) = xy + yz + zx$

$$\frac{\partial f}{\partial x} = y+z, \quad \frac{\partial f}{\partial y} = x+z, \quad \frac{\partial f}{\partial z} = x+y.$$

$$\frac{\partial^2 f}{\partial x^2} = 0, \quad \frac{\partial^2 f}{\partial y \partial x} = 1. \quad \frac{\partial^2 f}{\partial z \partial x} = 1.$$

$$\frac{\partial^2 f}{\partial z \partial y} = 1.$$

$$\frac{\partial^2 f}{\partial x \partial y} = 1. \quad \frac{\partial^2 f}{\partial y^2} = 0 \quad \frac{\partial^2 f}{\partial z^2} = 0.$$

$$\frac{\partial^2 f}{\partial x \partial z} = 1 \quad \frac{\partial^2 f}{\partial y \partial z} = 1.$$

Higher order partial derivatives

Let $f(x_1, x_2, \ldots, x_n)$ be a function defined on a domain D in $\mathbb{R}^n$. Then the higher order partial derivatives of f are defined analogously by taking successive partial derivatives.

$$f_{x_{i_1} x_{i_2} \ldots x_{i_k}} = \left(\left((f_{x_{i_1}})_{x_{i_2}}\right)_{x_{i_3}} \cdots\right)_{x_{i_k}}$$

$$\text{or} \quad \frac{\partial^k f}{\partial x_{i_k} \ldots \partial x_{i_2} \partial x_{i_1}} = \frac{\partial}{\partial x_{i_k}}\left(\frac{\partial}{\partial x_{i_{k-1}}}\left(\frac{\partial}{\partial x_{i_{k-2}}}\left(\cdots\left(\frac{\partial f}{\partial x_{i_1}}\right)\right)\right)\right)$$

An appropriately modified statement of Clairaut's theorem holds.

Under suitable hypothesis

$$f_{x_{i_1} x_{i_2} \cdots x_{i_k}} = f_{x_{j_1} x_{j_2} \cdots x_{j_k}}.$$

The Hessian matrix

Let $f(x_1, x_2, \ldots, x_n)$ be a function defined on a domain D in $\mathbb{R}^n$.

Then the Hessian matrix of f is defined as :

$$Hf = \begin{array}{c} \begin{array}{ccccc} & & j^{\text{th}}\text{ column} & & \end{array} \\ \left[\begin{array}{ccccc}
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_j} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_j} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\
\cdots & \cdots & \vdots & \cdots & \cdots & \cdots \\
\frac{\partial^2 f}{\partial x_i \partial x_1} & \frac{\partial^2 f}{\partial x_i \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_i \partial x_j} & \cdots & \frac{\partial^2 f}{\partial x_i \partial x_n} \\
\cdots & \cdots & \vdots & \cdots & \cdots & \cdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_j} & \cdots & \frac{\partial^2 f}{\partial x_n^2}
\end{array}\right] \leftarrow i^{\text{th}}\text{ row} \end{array}$$

The Hessian matrix

Let $f(x_1, x_2, \ldots, x_n)$ be a function defined on a domain D in $\mathbb{R}^n$.

Then the Hessian matrix of f is defined as :

$$Hf = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_j} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_j} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_i \partial x_1} & \frac{\partial^2 f}{\partial x_i \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_i \partial x_j} & \cdots & \frac{\partial^2 f}{\partial x_i \partial x_n} \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_j} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Examples

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

$$Hf = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}_{2\times 2}$$

$f(x, y) = sin(xy)$

$$Hf = \begin{bmatrix} -y^2 \sin(xy) & \cos(xy) - xy \sin(xy) \\ \cos(xy) - xy \sin(xy) & -x^2 \sin(xy) \end{bmatrix}$$

$f(x, y, z) = xy + yz + zx$

$$Hf = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}_{3\times 3}$$

The Hessian matrix and local extrema for $f(x, y)$

Sarang S. Sane

Recall : the second derivative test

Let $f : D \to \mathbb{R}$ be a function of one variable on the domain D.

A point $a \in D$ is a critical point if either f is not differentiable at a or $f'(a) = 0$.

Suppose f is twice differentiable at a. Then the second derivative test can be applied to check the nature of the critical points.

1. If a is a critical point and $f''(a) > 0$, then a is a local minimum.
2. If a is a critical point and $f''(a) < 0$, then a is a local maximum.
3. If a is a critical point and $f''(a) = 0$, then the test is inconclusive.

Recall : critical points for multivariable functions

Let $f(x_1, x_2, \ldots, x_n)$ be a scalar-valued multivariable function defined on a domain D in $\mathbb{R}^n$.

A point $\underset{\sim}{a} \in D$ is called a critical point of a function $f(\underset{\sim}{x})$ if either $\nabla f(\underset{\sim}{a})$ does not exist or $\nabla f(\underset{\sim}{a})$ exists and $\nabla f(\underset{\sim}{a}) = 0$.

Every local extremum is a critical point.

Unfortunately, not all critical points are local extrema.

A saddle point is a critical point $\underset{\sim}{a}$ such that $\nabla f(\underset{\sim}{a})$ exists and $\nabla f(\underset{\sim}{a}) = 0$ but $\underset{\sim}{a}$ is not a local extremum.

The Hessian test : Classifying critical points of $f(x, y)$

Let $f(x, y)$ be a function defined on a domain D in $\mathbb{R}^2$.

Let $\underset{\sim}{a}$ be a critical point of f such that the first and second order partial derivatives are continuous in an open ball around $\underset{\sim}{a}$.

Then the Hessian test can be applied to check the nature of the critical point $\underset{\sim}{a}$.

1. If $det(Hf(\underset{\sim}{a})) > 0$ and $f_{xx}(\underset{\sim}{a}) > 0$ then $\underset{\sim}{a}$ is a local minimum.
2. If $det(Hf(\underset{\sim}{a})) > 0$ and $f_{xx}(\underset{\sim}{a}) < 0$ then $\underset{\sim}{a}$ is a local maximum.
3. If $det(Hf(\underset{\sim}{a})) < 0$ then $\underset{\sim}{a}$ is a saddle point.
4. If $det(Hf(\underset{\sim}{a})) = 0$ then the test is inconclusive.

Examples

$f(x,y) = x^2 + y^2$

$\nabla f = (2x, 2y)$. Critical pt.: $(0,0)$.

$Hf = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. $Hf(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

$\det(Hf(0,0)) = 4 > 0$.

$f_{xx}(0,0) = 2 > 0$.

$\therefore (0,0)$ is a local minimum.

$f(x,y) = -x^2 - y^2$

$\nabla f = (-2x, -2y)$. Critical pt.: $(0,0)$

$Hf = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$ $Hf(0,0) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$.

$\det(Hf(0,0)) = 4 > 0$

$f_{xx}(0,0) = -2 < 0$.

$\therefore (0,0)$ is a local maximum.

$f(x,y) = x^2 - y^2$

$\nabla f = (2x, -2y)$. Critical pt.: $(0,0)$.

$Hf = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$. $Hf(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$.

$\det(Hf(0,0)) = -4 < 0$.

$\therefore (0,0)$ is a saddle pt.

$f(x,y) = x^4 + y^4$

$\nabla f = (4x^3, 4y^3)$. Critical pt. $(0,0)$.

$Hf = \begin{bmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{bmatrix}$. $Hf(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

$\det(Hf(0,0)) = 0$.

$\therefore$ The test is inconclusive.

Examples (contd.)

$f(x,y) = x^2 + 6xy + 4y^2 + 2x - 4y$

$\nabla f = (2x + 6y + 2, \ 6x + 8y - 4)$. Equating to 0, we get

the critical pt. $(2,-1)$.

$Hf = \begin{bmatrix} 2 & 6 \\ 6 & 8 \end{bmatrix} = Hf(2,-1)$.

$\det(Hf(2,-1)) = 16 - 36 = -20 < 0$.

$\therefore (2,-1)$ is a saddle point of $f(x,y)$.

$f(x,y) = xy - x^3 - y^2$

$\nabla f = (y - 3x^2, x - 2y)$. Equating to 0, we get:

$y = 3x^2, x = 2y \Rightarrow y = 3(2y)^2 = 12y^2$

$\Rightarrow y(1 - 12y) = 0$.

Critical pts.: $(0,0)$, $\left(\frac{1}{6}, \frac{1}{12}\right)$

$Hf(0,0) = \begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix}$ $\det(Hf(0,0)) = -1 < 0$.

$\therefore (0,0)$ is a saddle pt.

$Hf = \begin{bmatrix} -6x & 1 \\ 1 & -2 \end{bmatrix}$.

$Hf\left(\frac{1}{6}, \frac{1}{12}\right) = \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix}$ $\det\left(Hf\left(\frac{1}{6}, \frac{1}{12}\right)\right) = 2 - 1 = 1 > 0$.

$f_{xx}\left(\frac{1}{6}, \frac{1}{12}\right) = -1 < 0$.

$\therefore \left(\frac{1}{6}, \frac{1}{12}\right)$ is a local max.

Examples (contd.)

$f(x,y) = sin(xy)$

$\nabla f = (y \cos(xy), \ x \cos(xy))$.

$Hf = \begin{bmatrix} -y^2 \sin(xy) & \cos(xy) - xy\sin(xy) \\ \cos(xy) - xy\sin(xy) & -x^2\sin(xy) \end{bmatrix}$

Equating ∇f to 0, we get: (1) $\cos(xy) = 0$ or (2) $x = y = 0$.

$\Updownarrow$

$\sin(xy) = \pm 1$.

$Hf(0,0) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \Rightarrow \det(Hf(0,0)) = -1 < 0$.

$\Rightarrow (0,0)$ is a saddle pt. for f.

For pts. such that $\cos(xy) = 0$

$Hf(x,y) = \begin{bmatrix} -y^2 & -xy \\ -xy & -x^2 \end{bmatrix}$ if $\sin(xy) = 1$ or $Hf(x,y) = \begin{bmatrix} y^2 & xy \\ xy & x^2 \end{bmatrix}$ if $\sin(xy) = -1$

$\det(Hf(x,y)) = 0$. $\det(Hf(x,y)) = 0$.

Recall : The Hessian test for $f(x,y)$

Let $f(x,y)$ be a function defined on a domain D in $\mathbb{R}^2$.

Let $\underset{\sim}{a}$ be a critical point of f such that the first and second order partial derivatives are continuous in an open ball around $\underset{\sim}{a}$.

Then the Hessian test can be applied to check the nature of the critical point $\underset{\sim}{a}$.

1. If $det(Hf(\underset{\sim}{a})) > 0$ and $f_{xx}(\underset{\sim}{a}) > 0$ then $\underset{\sim}{a}$ is a local minimum.
2. If $det(Hf(\underset{\sim}{a})) > 0$ and $f_{xx}(\underset{\sim}{a}) < 0$ then $\underset{\sim}{a}$ is a local maximum.
3. If $det(Hf(\underset{\sim}{a})) < 0$ then $\underset{\sim}{a}$ is a saddle point.
4. If $det(Hf(\underset{\sim}{a})) = 0$ then the test is inconclusive.

The Hessian test : Classifying critical points of $f(x,y,z)$

Let $f(x,y,z)$ be a function defined on a domain D in $\mathbb{R}^3$.

Let $\underset{\sim}{a}$ be a critical point of f such that the first and second order partial derivatives are continuous in an open ball around $\underset{\sim}{a}$.

Then the Hessian test can be applied to check the nature of the critical point $\underset{\sim}{a}$.

1. If $f_{xx} > 0, (f_{xx}f_{yy} - f_{xy}^2)(\underset{\sim}{a}) > 0, det(Hf(\underset{\sim}{a})) > 0$ then $\underset{\sim}{a}$ is a local minimum.
2. If $f_{xx} < 0, (f_{xx}f_{yy} - f_{xy}^2)(\underset{\sim}{a}) < 0, det(Hf(\underset{\sim}{a})) < 0$ then $\underset{\sim}{a}$ is a local maximum.
3. If $det(Hf(\underset{\sim}{a})) \neq 0$ and cases 1 or 3 do not occur, then $\underset{\sim}{a}$ is a saddle point.

The Hessian test : Classifying critical points of $f(x,y,z)$

Let $f(x,y,z)$ be a function defined on a domain D in $\mathbb{R}^3$.

Let $\underset{\sim}{a}$ be a critical point of f such that the first and second order partial derivatives are continuous in an open ball around $\underset{\sim}{a}$.

Then the Hessian test can be applied to check the nature of the critical point $\underset{\sim}{a}$.

1. If $f_{xx} > 0, (f_{xx}f_{yy} - f_{xy}^2)(\underset{\sim}{a}) > 0, det(Hf(\underset{\sim}{a})) > 0$ then $\underset{\sim}{a}$ is a local minimum.
2. If $f_{xx} < 0, (f_{xx}f_{yy} - f_{xy}^2)(\underset{\sim}{a}) < 0, det(Hf(\underset{\sim}{a})) < 0$ then $\underset{\sim}{a}$ is a local maximum.
3. If $det(Hf(\underset{\sim}{a})) \neq 0$ and cases 1 or 3 do not occur, then $\underset{\sim}{a}$ is a saddle point.
4. If $det(Hf(\underset{\sim}{a})) = 0$ then the test is inconclusive.

Understanding the terms better

The terms involved in the test are : f_{xx}, $\boxed{(f_{xx}f_{yy} - f_{xy}^2)}(\underset{\sim}{a})$ and $det(Hf(\underset{\sim}{a}))$.

$$Hf(\underset{\sim}{a}) \quad \begin{bmatrix} f_{xx}(\underset{\sim}{a}) & f_{xy}(\underset{\sim}{a}) & f_{xz}(\underset{\sim}{a}) \\ f_{xy}(\underset{\sim}{a}) & f_{yy}(\underset{\sim}{a}) & f_{yz}(\underset{\sim}{a}) \\ f_{xz}(\underset{\sim}{a}) & f_{yz}(\underset{\sim}{a}) & f_{zz}(\underset{\sim}{a}) \end{bmatrix}$$

1×1	2×2	3×3	
+	+	+	→ local min.
−	+	−	→ local max.

all other non-degenerate cases i.e. $det(Hf(\underset{\sim}{a})) \neq 0$ → saddle point

degenerate case $det(Hf(\underset{\sim}{a})) = 0$ Inconclusive.

Examples

$f(x, y, z) = x^2 + y^2 + z^2$

$\nabla f = (2x, 2y, 2z)$ Critical pt. : $(0,0,0)$

$Hf = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = Hf(0,0,0)$.

$det(Hf(0,0,0)) = 8 > 0$

$det\left(\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right) = 4 > 0$

$f_{xx}(0,0,0) = 2 > 0$.

$\therefore (0,0,0)$ is a local min.

$f(x, y, z) = -x^2 - y^2 - z^2$

$\nabla f = (-2x, -2y, -2z)$ Critical pt. $(0,0,0)$.

$Hf = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} = Hf(0,0,0)$.

$det(Hf(0,0,0)) = -8 < 0$. −

$det\left(\begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}\right) = 4 > 0$ +

$f_{xx}(0,0,0) = -2 < 0$ −

$\therefore (0,0,0)$ is a local max.

$f(x, y, z) = x^2 - y^2 + z^2$

$(0,0,0)$

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

$det(Hf) = -8$ −

$det\left(\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}\right) = -4$ −

$f_{xx}(0,0,0) = 2$ +

$\therefore (0,0,0)$ is a saddle pt.

$f(x, y, z) = x^4 + y^4 + z^4$

$\nabla f = (4x^3, 4y^3, 4z^3)$. Critical pt. $(0,0,0)$.

$Hf = \begin{bmatrix} 12x^2 & 0 & 0 \\ 0 & 12y^2 & 0 \\ 0 & 0 & 12z^2 \end{bmatrix}$.

$Hf(0,0,0) = O_{3\times 3}$.

Inconclusive.

Example

$f(x, y) = xy + yz + zx$

$\nabla f = (y+z, z+x, x+y)$.

Equating to 0, we get $\left.\begin{matrix} x = -y = z \\ x = -z \end{matrix}\right\} \Rightarrow x = y = z = 0.$

Critical pt. $(0,0,0)$.

$Hf = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = Hf(0,0,0)$.

$det(Hf(0,0,0)) = 0 \times det\begin{bmatrix} & \end{bmatrix} - 1 \times det\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} + 1 \times det\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

$= 1 + 1 = 2 > 0$.

$det\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1 < 0$.

$\therefore (0,0,0)$ is a saddle point.

Differentiability for functions of one variable

Definition

Let f be a function defined on an open interval around a. Then f is differentiable at a if $\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$ exists.

For a function $f(x)$ its derivative function, $f'(x)$ or $\frac{df}{dx}(x)$ is

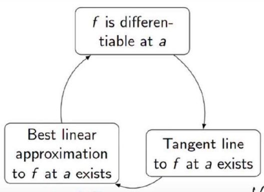
$$f'(x) = \frac{df}{dx}(x) = \boxed{\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}}.$$

Its domain consists of those points at which the function $f(x)$ is differentiable.

MORE VIDEOS

If f is differentiable at a, then it is continuous at a.

Differentiability, tangents and best linear approximation



$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - f'(a)h}{h} = 0.$$

$$\frac{f(a+h) - \boxed{(f(a) + f'(a)h)}}{h} = 0.$$

$$y = \underbrace{f(a) + f'(a)(x-a)}$$

$$L_f(x) = f(a) + f'(a)(x-a)$$

$$f(x) \approx L_f(x)$$

when x is near a.

MORE VIDEOS

5:51 / 14:42

Differentiability for scalar-valued multivariable functions

Definition

Let f be a scalar-valued multivarable function defined on a domain D in $\mathbb{R}^n$ containing an open ball around a point $\underset{\sim}{a}$.

Then f is differentiable at $\underset{\sim}{a}$ if

$$\lim_{\underset{\sim}{h} \to 0} \frac{f(\underset{\sim}{a} + \underset{\sim}{h}) - f(\underset{\sim}{a}) - \underset{\sim}{h} \cdot \nabla f(\underset{\sim}{a})}{||\underset{\sim}{h}||} = 0.$$

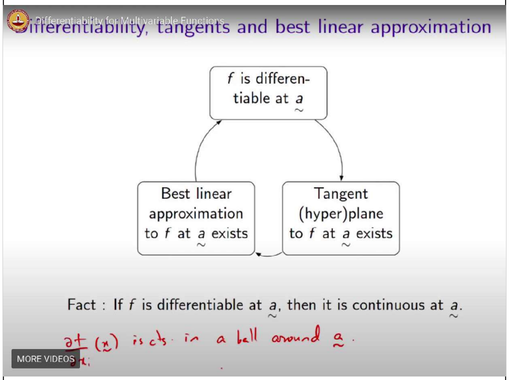
2-vars. $h = (h_1, h_2)$

$$\lim_{(h_1,h_2) \to (0,0)} \frac{f((a,b) + (h_1,h_2)) - f(a,b) - h_1 f_x(a,b) - h_2 f_y(a,b)}{\sqrt{h_1^2 + h_2^2}} = 0$$

$$f(\underset{\sim}{a} + \underset{\sim}{h}) - \underbrace{\{f(\underset{\sim}{a}) + \underset{\sim}{h} \cdot \nabla f(\underset{\sim}{a})\}}$$

$$z = f(\underset{\sim}{a}) + \nabla f(\underset{\sim}{a}) \cdot (\underset{\sim}{x} - \underset{\sim}{a})$$

MORE VIDEOS


Differentiability for Multivariable Functions
Differentiability, tangents and best linear approximation
f is differentiable at $\underset{\sim}{a}$
Best linear approximation to f at $\underset{\sim}{a}$ exists
Tangent (hyper)plane to f at $\underset{\sim}{a}$ exists
Fact : If f is differentiable at $\underset{\sim}{a}$, then it is continuous at $\underset{\sim}{a}$.
$\frac{\partial f}{\partial x_i}(\underset{\sim}{x})$ is c'ts. in a ball around $\underset{\sim}{a}$.
MORE VIDEOS