Multivariable Calculus

Azure

2022

Contents

Definition 0.0.1: Common Notation

The notes will be using common mathematical notations.

$\mathbb{N}$	The set of all natural numbers
$\mathbb{Z}$	The set of all integers
$\mathbb{Q}$	The set of all rational numbers
$\mathbb{R}$	The set of all real numbers
$\mathbb{C}$	The set of all complex numbers
$(x_1, ..., x_n)$	Ordered n-tupled with values $x_1, x_2, x_3, ..., x_n$
$\{x_1, ..., x_n\}$	Finite set with elements $x_1, x_2, x_3, ..., x_n$
$\emptyset$	Empty set
x $\in$ E	x is an element of set E
x $\notin$ E	x is not an element of set E
A $\subset$ B	Set A is a subset set E
A $\not\subset$ B	Set A is not a subset set E
f: A $\rightarrow$ B	Function f maps set A into set B

1 Vectors

1.1 Vectors

Definition 1.1.1: Vectors

A scalar c is a number in $\mathbb{R}$.

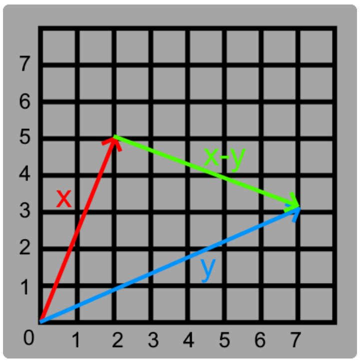
A vector x $\in \mathbb{R}^n$ is an ordered n-tuple of real numbers.

x = $(x_1, ..., x_n) = < x_1, ..., x_n >$ where each $x_i \in \mathbb{R}$

Let the zero vector 0 = (0,...,0).

If x,y $\in \mathbb{R}^n$ and c is a scalar:

Comparison:	x = y if $x_i = y_i$ for i = {1,...,n}
Vector Addition:	x+y = $(x_1 + y_1, ..., x_n + y_n)$
Scalar Multiplicaton:	cx = $(cx_1, ..., cx_n)$



Theorem 1.1.2: Vector Operations

(a) x+y = y+x

Proof

$\boxed{\text{x+y} = (x_1 + y_1, ..., x_n + y_n) = (y_1 + x_1, ..., y_n + x_n) = \text{y+x}}$

(b) x+(y+z) = (x+y)+z

Proof

$$\begin{aligned}\text{x+(y+z)} &= (x_1, ..., x_n) + (y_1 + z_1, ..., y_n + z_n) = (x_1 + y_1 + z_1, ..., x_n + y_n + z_n) \\ &= (x_1 + y_1, ..., x_n + y_n) + (z_1, ..., z_n) = \text{(x+y)+z}\end{aligned}$$

(c) x+0 = x

Proof

$\boxed{\text{x+0} = (x_1 + 0, ..., x_n + 0) = (x_1, ..., x_n) = \text{x}}$

(d) c(x+y) = cx + cy

Proof

$\boxed{\text{c(x+y)} = (c(x_1 + y_1), ..., c(x_n + y_n)) = (cx_1 + cy_1, ..., cx_n + cy_n) = \text{cx+cy}}$

(e) (c+k)v = cv + kv

Proof

$\boxed{\text{(c+k)v} = ((c + k)v_1, ..., (c + k)v_n) = (cv_1 + kv_1, ..., cv_n + kv_n) = \text{cv+kv}}$

(f) c(kx) = (ck)x = k(cx)

Proof

$$\begin{aligned} \text{c(kv)} &= (c(kx_1), ..., c(kx_n)) = (ckx_1, ..., ckx_n) = \text{(ck)x} = (kcx_1, ..., kcx_n) \\ &= (k(cx_1), ..., k(cx_n)) = \text{k(cx)} \end{aligned}$$

Definition 1.1.3: Standard Basis Vectors

The standard basis vectors for $\mathbb{R}^n$ are $e_1, ..., e_n$ where each i = {1,...,n}:

$$e_i = (\underset{1}{0}, \underset{2}{0}, ..., \underset{i-1}{0}, \underset{i}{1}, \underset{i+1}{0}, ..., \underset{n-1}{0}, \underset{n}{0},)$$

Thus, for any x $\in \mathbb{R}^n$:

$$\text{x} = (x_1, ..., x_n) = x_1e_1 + ... + x_ne_n$$

1.2 Dot Product

Definition 1.2.1: Dot Product, Norm, and Orthogonality

The dot product of x,y $\in \mathbb{R}^n$ is the sum of the products of their components:

$$x \cdot y = x_1y_1 + ... + x_ny_n$$

The length of x $\in \mathbb{R}^n$ is the norm:

$$||x|| = \sqrt{x_1^2 + ... + x_n^2} = \sqrt{x \cdot x} \qquad \Rightarrow \qquad x \cdot x = ||x||^2$$

Thus, $||cx|| = \sqrt{(cx_1)^2 + ... + (cx_n^2)} = |c|\sqrt{x_1^2 + ... + x_n^2} = |c|\ ||x||$.

Then, a unit vector (i.e. vector of length 1) in the direction of x is $\frac{x}{||x||}$.

x,y $\in \mathbb{R}^n$ are orthogonal (i.e. perpendicular) if:

$$x \cdot y = 0$$

Theorem 1.2.2: Properties of the Dot Product

(a) $x \cdot x \geq 0$

Proof

$\boxed{x \cdot x = x_1x_1 + ... + x_nx_n = x_1^2 + ... + x_n^2 \geq 0 + ... + 0 = 0}$

(b) $x \cdot x = 0$ if and only if x = 0

Proof

$x \cdot x = x_1x_1 + ... + x_nx_n = x_1^2 + ... + x_n^2$

Thus, $x \cdot x = 0$ if and only if each $x_i^2 = 0$ so each $x_i = 0$. Thus, x = 0.

(c) $x \cdot y = y \cdot x$

Proof

$\boxed{x \cdot y = x_1y_1 + ... + x_ny_n = y_1x_1 + ... + y_nx_n = y \cdot x}$

(d) $x \cdot (y+z) = x \cdot y + x \cdot z$
Proof

$$x \cdot (y+z) = x_1(y_1+z_1) + ... + x_n(y_n+z_n) = (x_1y_1 + x_1z_1) + ... + (x_ny_n + x_nz_n)$$
$$= (x_1y_1 + ... + x_ny_n) + (x_1z_1 + x_nz_n) = x \cdot y + x \cdot z$$

(e) $(x+y) \cdot z = x \cdot z + y \cdot z$
Proof

$$(x+y) \cdot z = (x_1+y_1)z_1 + ... + (x_n+y_n)z_n = (x_1z_1 + y_1z_1) + ... + (x_nz_n + y_nz_n)$$
$$= (x_1z_1 + ... + x_nz_n) + (y_1z_1 + ... + y_nz_n) = x \cdot z + y \cdot z$$

(f) $cx \cdot y = c(x \cdot y) = x \cdot cy$
Proof

$$cx \cdot y = (cx_1)y_1 + ... + (cx_n)y_n = c(x_1y_1) + ... + c(x_ny_n) = c(x \cdot y)$$
$$= x_1(cy_1) + ... + x_n(cy_n) = x \cdot cy$$

Theorem 1.2.3: $x \cdot y = ||x|| \, ||y|| \cos(\theta)$

For x,y $\in \mathbb{R}^n$:
$x \cdot y = ||x|| \, ||y|| \cos(\theta)$
where $\theta \in [0, \pi]$ is the angle between x and y

Proof

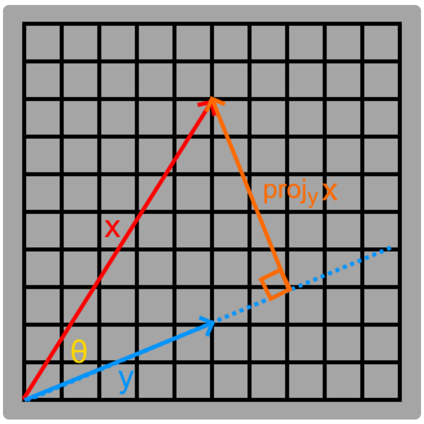
Since x, y, and x-y form a triangle, by the Law of Cosine:
$||x-y||^2 = ||x||^2 + ||y||^2 - 2||x|| \, ||y|| \cos(\theta)$
where $\theta \in [0, \pi]$ is the angle between x and y. Since:
$||x-y||^2 = (x-y) \cdot (x-y) = x \cdot x + y \cdot y - 2(x \cdot y) = ||x||^2 + ||y||^2 - 2(x \cdot y)$
then $x \cdot y = ||x|| \, ||y|| \cos(\theta)$.

Theorem 1.2.4: Vector Projection

The projection of x $\in \mathbb{R}^n$ onto y $\in \mathbb{R}^n$ is the component of x parallel to y:
$\text{proj}_y x = \frac{x \cdot y}{||y||^2} y$

Proof

Since $\text{proj}_y x$ is parallel to y, let $\text{proj}_y x$ = cy for some constant c $\in \mathbb{R}$.
Let $y^\perp$ be the orthogonal component of x to y. Thus, x = $\text{proj}_y x + y^\perp$ = cy + $y^\perp$.
Since $y^\perp$ is orthogonal to y, then:
$x \cdot y = (cy + y^\perp) \cdot y = cy \cdot y + y^\perp \cdot y = cy \cdot y = c||y||^2$
Thus, c = $\frac{x \cdot y}{||y||^2}$ so $\text{proj}_y x$ = cy = $\frac{x \cdot y}{||y||^2} y$.



Theorem 1.2.5: Cauchy-Schwarz Inequality

For x,y $\in \mathbb{R}^n$, $|x \cdot y| \leq ||x||\ ||y||$

Proof

Let y = proj$_x y$ + $x^{\perp}$ = cx + $x^{\perp}$ where $x^{\perp}$ is the orthogonal component of y to x and proj$_x y$ = cx is the parallel component to x for some c $\in \mathbb{R}$.

$$x \cdot y = x \cdot (cx + x^{\perp}) = c(x \cdot x) + x \cdot x^{\perp} = c||x||^2 + 0 = c||x||^2$$

Thus, c = $\frac{x \cdot y}{||x||^2}$. Then:

$$||y||^2 = ||cx + x^{\perp}||^2 = (cx + x^{\perp}) \cdot (cx + x^{\perp}) = cx \cdot cx + x^{\perp} \cdot x^{\perp} + 2(cx \cdot x^{\perp})$$
$$= c^2||x||^2 + ||x^{\perp}||^2 = (\tfrac{x \cdot y}{||x||^2})^2||x||^2 + ||x^{\perp}||^2$$
$$||x||^2||y||^2 = ||x||^2(\tfrac{x \cdot y}{||x||^2})^2||x||^2 + ||x||^2||x^{\perp}||^2 = (x \cdot y)^2 + ||x||^2||x^{\perp}||^2$$

Since $||x||^2||x^{\perp}||^2 \geq 0$, then $(x \cdot y)^2 \leq ||x||^2||y||^2$ so $|x \cdot y| \leq ||x||\ ||y||$.

Theorem 1.2.6: Triangle Inequality

For x,y $\in \mathbb{R}^n$, $||x + y|| \leq ||x|| + ||y||$

Proof

$$||x + y||^2 = (x + y) \cdot (x + y) = x \cdot x + y \cdot y + 2(x \cdot y) = ||x||^2 + ||y||^2 + 2(x \cdot y)$$
$$\leq ||x||^2 + ||y||^2 + 2|x \cdot y| \leq ||x||^2 + ||y||^2 + 2||x||\ ||y|| = (||x|| + ||y||)^2$$

1.3 Cross Product

Definition 1.3.1: Cross Product

The cross product of x,y $\in \mathbb{R}^3$ is the determinant of the standard basis, x, y:

$$x \times y = \det(\begin{bmatrix} e_1 & e_2 & e_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}) = \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} e_1 - \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} e_2 + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} e_3$$

Theorem 1.3.2: Properties of the Cross Product

(a) $x \times y = -(y \times x)$

Proof

$$x \times y = \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} e_1 - \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} e_2 + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} e_3$$
$$= -\begin{vmatrix} y_2 & y_3 \\ x_2 & x_3 \end{vmatrix} e_1 + \begin{vmatrix} y_1 & y_3 \\ x_1 & x_3 \end{vmatrix} e_2 - \begin{vmatrix} y_1 & y_2 \\ x_1 & x_2 \end{vmatrix} e_3 = -(y \times x)$$

(b) $x \times (y + z) = x \times y + x \times z$

Proof

$$x \times (y + z) = \begin{vmatrix} x_2 & x_3 \\ y_2 + z_2 & y_3 + z_3 \end{vmatrix} e_1 - \begin{vmatrix} x_1 & x_3 \\ y_1 + z_1 & y_3 + z_3 \end{vmatrix} e_2 + \begin{vmatrix} x_1 & x_2 \\ y_1 + z_1 & y_2 + z_2 \end{vmatrix} e_3$$
$$= (\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} + \begin{vmatrix} x_2 & x_3 \\ z_2 & z_3 \end{vmatrix})e_1 - (\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} + \begin{vmatrix} x_1 & x_3 \\ z_1 & z_3 \end{vmatrix})e_2 + (\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} + \begin{vmatrix} x_1 & x_2 \\ z_1 & z_2 \end{vmatrix})e_3$$
$$= \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} e_1 - \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} e_2 + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} e_3 + \begin{vmatrix} x_2 & x_3 \\ z_2 & z_3 \end{vmatrix} e_1 - \begin{vmatrix} x_1 & x_3 \\ z_1 & z_3 \end{vmatrix} e_2 + \begin{vmatrix} x_1 & x_2 \\ z_1 & z_2 \end{vmatrix} e_3$$
$$= x \times y + x \times z$$

(c) $(x+y) \times z = x \times z + y \times z$

Proof

$$(x+y) \times z = -[z \times (x+y)] = -[z \times x + z \times y]$$
$$= -[-(x \times z) + -(y \times z)] = x \times z + y \times z$$

(d) $c(x \times y) = cx \times y = x \times cy$

Proof

$$c(x \times y) = c\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} e_1 - c\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} e_2 + c\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} e_3$$
$$= \begin{vmatrix} cx_2 & cx_3 \\ y_2 & y_3 \end{vmatrix} e_1 - \begin{vmatrix} cx_1 & cx_3 \\ y_1 & y_3 \end{vmatrix} e_2 + \begin{vmatrix} cx_1 & cx_2 \\ y_1 & y_2 \end{vmatrix} e_3 = cx \times y$$
$$= \begin{vmatrix} x_2 & x_3 \\ cy_2 & cy_3 \end{vmatrix} e_1 - \begin{vmatrix} x_1 & x_3 \\ cy_1 & cy_3 \end{vmatrix} e_2 + \begin{vmatrix} x_1 & x_2 \\ cy_1 & cy_2 \end{vmatrix} e_3 = x \times cy$$

Theorem 1.3.3: Orthgonality of $x \times y$

$x \times y$ is orthogonal to x and y

Proof

$$x \times y \cdot \mathrm{x} = (\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}, -\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}, \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}) \cdot (x_1, x_2, x_3)$$
$$= (x_2y_3 - x_3y_2)x_1 - (x_1y_3 - x_3y_1)x_2 + (x_1y_2 - x_2y_1)x_3$$
$$= x_1x_2y_3 - x_1x_3y_2 - x_1x_2y_3 + x_2x_3y_1 + x_1x_3y_2 - x_2x_3y_1 = 0$$
$$x \times y \cdot \mathrm{y} = (\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}, -\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}, \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}) \cdot (y_1, y_2, y_3)$$
$$= (x_2y_3 - x_3y_2)y_1 - (x_1y_3 - x_3y_1)y_2 + (x_1y_2 - x_2y_1)y_3$$
$$= x_2y_1y_3 - x_3y_1y_2 - x_1y_2y_3 + x_3y_1y_2 + x_1y_2y_3 - x_2y_1y_3 = 0$$

Theorem 1.3.4: $||x \times y|| = ||x||\ ||y|| \sin(\theta)$

For x,y $\in \mathbb{R}^3$:

$$||x \times y|| = ||x||\ ||y|| \sin(\theta)$$

where $\theta \in [0, \pi]$ is the angle between x and y

Proof

By theorem 1.2.3, $x \cdot y = ||x||\ ||y|| \cos(\theta)$ where $\theta \in [0, \pi]$ is the angle between x,y.

$$||x||^2||y||^2 - (x \cdot y)^2 = ||x||^2||y||^2(1 - \cos^2(\theta)) = ||x||^2||y||^2 \sin^2(\theta)$$

Also:

$$||x||^2||y||^2 - (x \cdot y)^2 = (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2) - (x_1y_1 + x_2y_2 + x_3y_3)^2$$
$$= (x_1^2y_1^2 + x_2^2y_2^2 + x_3^2y_3^2 + x_1^2y_2^2 + x_1^2y_3^2 + x_2^2y_1^2 + x_2^2y_3^2 + x_3^2y_1^2 + x_3^2y_2^2$$
$$-x_1^2y_1^2 - x_2^2y_2^2 - x_3^2y_3^2 - 2x_1x_2y_1y_2 - 2x_1x_3y_1y_3 - 2x_2x_3y_2y_3)$$
$$= (x_2y_3 - x_3y_2)^2 + (x_3y_1 - x_1y_3)^2 + (x_1y_2 - x_2y_1)^2 = ||x \times y||^2$$

Thus, $||x \times y|| = ||x||\ ||y|| \sin(\theta)$.

Theorem 1.3.5: Area of Parallelogram

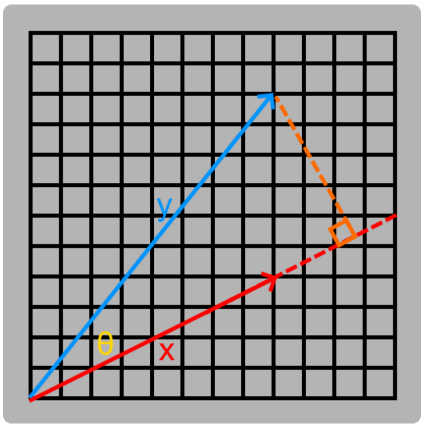
The area of a parallelogram P with sides x,y $\in \mathbb{R}^3$:

$\text{Vol}_2(P(x,y)) = ||x \times y||$

Proof

Since parallelogram P with sides x and y is two triangles with sides x and y, then:

$$\begin{aligned}\text{Vol}_2(P(x,y)) &= 2 * \text{Vol}_2(\text{Triangle}(x,y)) \\ &= 2 * \tfrac{1}{2} \text{ (base of triangle) } * \text{ (height of triangle)} \\ &= ||x|| * (||y|| \sin(\theta)) = ||x \times y||\end{aligned}$$



Theorem 1.3.6: Volume of Parallelepiped

The volume of a parallelepiped P with sides x,y,z $\in \mathbb{R}^n$:

$\text{Vol}_3(P(x,y,z)) = |(x \times y) \cdot z|$

Proof

Let sides x and y form a base for P.

$\text{Vol}_3(P(x,y,z)) = \text{(Area of base) } * \text{ (height)} = ||x \times y|| * (||z|| \cos(\theta))$

where $\theta \in [0, \pi]$ is the angle between $x \times y$ and z. By theorem 1.2.3:

$\text{Vol}_3(P(x,y,z)) = (x \times y) \cdot z$

Since -1 $\leq \cos(\theta) \leq$ 1 for $\theta \in [0, 2\pi]$, then $(x \times y) \cdot z$ can be negative. Thus:

$\text{Vol}_3(P(x,y,z)) = |(x \times y) \cdot z|$

1.4 Distances and Planes

Theorem 1.4.1: Equation of a Plane: Method #1: Point and Normal Vector

A plane in $\mathbb{R}^3$ through a point p = (p_x, p_y, y_z) and orthogonal to a vector called a normal vector n = (a, b, c) has an equation of the form:

n $\times$ $[(x,y,z) - p] = a(x - p_x) + b(y - p_y) + c(z - p_z) = 0$

Proof

Let (x,y,z) be any point in the plane. Then (x,y,z) - p = $(x - p_x, y - p_y, z - p_z)$ is a vector parallel to the plane. Since the plane is orthogonal to vector n, then any vector parallel to the plane is orthogonal to n. Thus:

$n \cdot (x - p_x, y - p_y, z - p_z) = 0$

$a(x - p_x) + b(y - p_y) + c(z - p_z) = 0$

Theorem 1.4.2: Equation of a Plane: Method #2: 3 Points

A plane in $\mathbb{R}^3$ through points $p_1 = (x_1, y_1, z_1)$, $p_2 = (x_2, y_2, z_2)$, and $p_3 = (x_3, y_3, z_3)$ has an equation of the form:

$[(p_2 - p_1) \times (p_3 - p_1)] \cdot [(x, y, z) - p_1] = 0$

Proof

Since p_1, p_2, and p_3 are on the plane, then $p_2 - p_1$ and $p_3 - p_1$ are vectors on the plane and thus, parallel to the plane. Since $(p_2 - p_1) \times (p_3 - p_1)$ is orthgonal to $(p_2 - p_1)$ and $(p_3 - p_1)$, then $(p_2 - p_1) \times (p_3 - p_1)$ is orthogonal to the plane and thus, a normal vector. By theorem 1.4.1, then:

$[(p_2 - p_1) \times (p_3 - p_1)] \cdot [(x, y, z) - p_1] = 0$

Theorem 1.4.3: Distance: Point + Line or 2 Parallel Lines

The distance from line L(t) = tv+x_0 to point p $\in \mathbb{R}^3$ where t $\in \mathbb{R}$, v,$x_0 \in \mathbb{R}^3$:

$\frac{||v \times (p - x_0)||}{||v||}$

If line $L_2(t)$ is parallel to L(t), choose a point on $L_2(t)$ and apply formula above to get the distance between two parallel lines.

Proof

Since x_0 is a point on L(t), then $p - x_0$ is a vector from line L(t) to p.
Let θ be the angle between $p - x_0$ and L(t). Thus:

$\sin(\theta) = \frac{d}{||p - x_0||} \quad \Rightarrow \quad \text{d} = ||p - x_0|| * \sin(\theta) = \frac{||v|| * ||p - x_0|| * \sin(\theta)}{||v||} = \frac{||v \times (p - x_0)||}{||v||}$

Theorem 1.4.4: Distance: Parallel Planes

The distance between parallel planes P_1: $a(x - x_1) + b(y - y_1) + c(z - z_1) = 0$ and P_2: $a(x - x_2) + b(y - y_2) + c(z - z_2) = 0$:

$\text{d} = \frac{|(a,b,c) \cdot (x_2 - x_1, y_2 - y_1, z_2 - z_1)|}{\sqrt{a^2 + b^2 + c^2}}$

Proof

Planes P_1 and P_2 are parallel since they both have the normal vector n = (a,b,c). Since (x_1, y_1, z_1) is a point on P_1 and (x_2, y_2, z_2) is a point on P_2, then $(x_2, y_2, z_2) - (x_1, y_1, z_1)$ is a vector from P_1 to P_2.
Then the distance is the norm of the orthogonal component of $(x_2, y_2, z_2) - (x_1, y_1, z_1)$ to P_1, P_2. Since normal vector n is orthogonal to both planes, then the orthogonal component of $(x_2, y_2, z_2) - (x_1, y_1, z_1)$ and n are parallel.
Thus, by theorem 1.2.4:

$\text{d} = ||\text{proj}_n[(x_2, y_2, z_2) - (x_1, y_1, z_1)]|| = ||\frac{[(x_2, y_2, z_2) - (x_1, y_1, z_1)] \cdot (a,b,c)}{||(a,b,c)||^2}(a, b, c)||$

$\text{d} = \frac{|(x_2 - x_1, y_2 - y_1, z_2 - z_1) \cdot (a,b,c)|}{||(a,b,c)||}$

Theorem 1.4.5: Distance: Skew Lines

Lines $L_1, L_2 \in \mathbb{R}^3$ are skewed if they are neither parallel or intersecting.
Let $L_1(t) = tv_1 + x_1$ and $L_1(t) = tv_2 + x_2$. The distance between L_1 and L_2:

$\text{d} = \frac{|(v_2 \times v_1) \cdot (x_2 - x_1)|}{||v_2 \times v_1||}$

Proof

Let L_1, L_2 be in two parallel planes. Note the distance between L_1 and L_2 is the distance between the two planes.
Since $v_2 \times v_1$ is orthogonal to v_1, v_2 and v_1, v_2 are vectors parallel to each plane, then $v_2 \times v_1$ is orthogonal to each plane and thus, a normal vector. By theorem 1.4.4:

$\text{d} = \frac{|(v_2 \times v_1) \cdot (x_2 - x_1)|}{||v_2 \times v_1||}$

1.5 Matrices

Definition 1.5.1: Matrix

A m by n matrix $M_{m\times n}(\mathbb{R})$:

$$M = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \qquad \text{where each } a_{ij} \in \mathbb{R}$$

A row vector is a 1 by n matrix:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \end{bmatrix}$$

A column vector is a m by 1 matrix:

$$\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{bmatrix}$$

A zero matrix $0 \in M_{m\times n}(\mathbb{R})$:

$$0 = \begin{bmatrix} 0_{11} & 0_{12} & a_{13} & \dots & 0_{1n} \\ 0_{21} & 0_{22} & a_{23} & \dots & 0_{2n} \\ 0_{31} & 0_{32} & a_{33} & \dots & 0_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{m1} & 0_{m2} & a_{m3} & \dots & 0_{mn} \end{bmatrix}$$

Theorem 1.5.2: Matrix Operations

(a) Addition

For A,B $\in M_{m\times n}(\mathbb{R})$, then A+B $\in M_{m\times n}(\mathbb{R})$ where each a_{ij}, b_{ij}:

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

(b) Scalar Multiplication

For A $\in M_{m\times n}(\mathbb{R})$, then cA $\in \in M_{m\times n}(\mathbb{R})$ where each a_{ij}, b_{ij}:

$$c\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{bmatrix}$$

(c) Multiplication

For A $\in M_{m\times n}(\mathbb{R})$, B $\in M_{n\times k}(\mathbb{R})$, then AB $\in M_{m\times k}(\mathbb{R})$ where each a_{ij}, b_{ij}:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nk} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n a_{1i}b_{i1} & \sum_{i=1}^n a_{1i}b_{i2} & \dots & \sum_{i=1}^n a_{1i}b_{ik} \\ \sum_{i=1}^n a_{2i}b_{i1} & \sum_{i=1}^n a_{2i}b_{i2} & \dots & \sum_{i=1}^n a_{2i}b_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n a_{mi}b_{i1} & \sum_{i=1}^n a_{mi}b_{i2} & \dots & \sum_{i=1}^n a_{mi}b_{ik} \end{bmatrix}$$

Theorem 1.5.3: Properties of Matrix Operations

(a) A+B = B+A

Proof

$[A + B]_{ij} = a_{ij} + b_{ij} = b_{ij} + a_{ij} = [B + A]_{ij}$

(b) A+(B+C) = (A+B)+C

Proof

$[A + (B + C)]_{ij} = a_{ij} + (b_{ij} + c_{ij}) = (a_{ij} + b_{ij}) + c_{ij} = [(A + B) + C]_{ij}$

(c) A+0 = A

Proof

$[A + 0]_{ij} = a_{ij} + 0_{ij} = a_{ij} = [A]_{ij}$

(d) (c+k)A = cA + kA

Proof

$[(c + k)A]_{ij} = (c + k)a_{ij} = ca_{ij} + ka_{ij} = [cA]_{ij} + [kA]_{ij} = [cA + kA]_{ij}$

(e) c(A+B) = cA + cB

Proof

$[c(A + B)]_{ij} = c(a_{ij} + b_{ij}) = ca_{ij} + cb_{ij} = [cA]_{ij} + [cB]_{ij} = [cA + cB]_{ij}$

(f) c(kA) = (ck)A = k(cA)

Proof

$[c(kA)]_{ij} = c(ka_{ij}) = (ck)a_{ij} = [(ck)A]_{ij} = k(ca_{ij}) = [k(cA)]_{ij}$

(g) A(BC) = (AB)C

Proof

Let A $\in M_{m\times n}(\mathbb{R})$, B $\in M_{n\times k}(\mathbb{R})$, and C $\in M_{k\times p}(\mathbb{R})$.
For u $\in$ {1,...,n} and v $\in$ {1,...,p}, then $[\text{BC}]_{uv} = \sum_{s=1}^{k} b_{us}c_{sv}$.
Thus, for i $\in$ {1,...,m} and j $\in$ {1,...,p}:

$[A(BC)]_{ij} = \sum_{t=1}^{n} a_{it}[BC]_{tj} = \sum_{t=1}^{n}[a_{it} \sum_{s=1}^{k} b_{ts}c_{sj}] = \sum_{t=1}^{n} \sum_{s=1}^{k} a_{it}b_{ts}c_{sj}$
$= \sum_{s=1}^{k} \sum_{t=1}^{n} a_{it}b_{ts}c_{sj} = \sum_{s=1}^{k}[\sum_{t=1}^{n} a_{it}b_{ts}]c_{sj} = \sum_{s=1}^{k}[AB]_{is}c_{sj} = [(AB)C]_{ij}$

(h) c(AB) = (cA)B = A(cB)

Proof

Let A $\in M_{m\times n}(\mathbb{R})$ and B $\in M_{n\times k}(\mathbb{R})$. For i $\in$ {1,...,m} and j $\in$ {1,...,k}:

$[c(AB)]_{ij} = c\sum_{t=1}^{n} a_{it}b_{tj} = \sum_{t=1}^{n} ca_{it}b_{tj} = \sum_{t=1}^{n}(ca_{it})b_{tj} = [(cA)B]_{ij}$
$[c(AB)]_{ij} = c\sum_{t=1}^{n} a_{it}b_{tj} = \sum_{t=1}^{n} a_{it}cb_{tj} = \sum_{t=1}^{n} a_{it}(cb_{tj}) = [A(cB)]_{ij}$

(i) A(B+C) = AB + AC

Proof

Let A $\in M_{m\times n}(\mathbb{R})$ and B,C $\in M_{n\times k}(\mathbb{R})$. For i $\in$ {1,...,m} and j $\in$ {1,...,k}:

$[A(B+C)]_{ij} = \sum_{t=1}^{n} a_{it}[B+C]_{tj} = \sum_{t=1}^{n} a_{it}(b_{tj}+c_{tj}) = \sum_{t=1}^{n} a_{it}b_{tj}+a_{it}c_{tj}$
$= \sum_{t=1}^{n} a_{it}b_{tj} + \sum_{t=1}^{n} a_{it}c_{tj} = [AB]_{ij} + [AC]_{ij} = [AB + AC]_{ij}$

(j) (A+B)C = AC + BC

Proof

Let A,B $\in M_{m\times n}(\mathbb{R})$ and C $\in M_{n\times k}(\mathbb{R})$.
For i $\in$ {1,...,m} and j $\in$ {1,...,k}, the ij-th entry for (A+B)C:

$[(A+B)C]_{ij} = \sum_{t=1}^{n}[A+B]_{it}c_{tj} = \sum_{t=1}^{n}(a_{it} + b_{it})c_{tj} = \sum_{t=1}^{n} a_{it}c_{tj} + b_{it}c_{tj}$
$= \sum_{t=1}^{n} a_{it}c_{tj} + \sum_{t=1}^{n} b_{it}c_{tj} = [AC]_{ij} + [BC]_{ij} = [AC + BC]_{ij}$

Definition 1.5.4: Transpose

For matrix A $\in M_{m\times n}(\mathbb{R})$:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

then the transpose, $A^T \in M_{n\times m}(\mathbb{R})$:

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{m1} \\ a_{12} & a_{22} & a_{32} & \dots & a_{m2} \\ a_{13} & a_{23} & a_{33} & \dots & a_{m3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \dots & a_{mn} \end{bmatrix}$$

Theorem 1.5.5: Properties of the Transpose

(a) $(A^T)^T = A$

Proof

$[(A^T)^T]_{ij} = [A^T]_{ji} = [A]_{ij}$

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

Proof

Let A $\in M_{m\times n}(\mathbb{R})$ and B $\in M_{n\times k}(\mathbb{R})$. For i = {1,...,k} and j = {1,...,m}:

$$[(AB)^T]_{ij} = [AB]_{ji} = \sum_{t=1}^{n} a_{jt} b_{ti} = \sum_{t=1}^{n} b_{ti} a_{jt} = \sum_{t=1}^{n} b_{it}^T a_{tj}^T = [B^T A^T]_{ij}$$

(c) $x \cdot y = x^T y$

Proof

$x \cdot y = x_1 y_1 + x_2 y_2 + ... + x_n y_n = [x_1 \ x_2 \ ... \ x_n] y = x^T y$

Definition 1.5.6: Determinant

For A $\in M_{n\times n}(\mathbb{R})$, let prod(A) = $a_{1,j_1} * a_{2,j_2} * ... * a_{n,j_n}$ such that for any two a_{k,j_k},a_{p,j_p} where k < p, then $j_k \neq j_p$. Let prod(A) be unique in the sense that no two prod(A) have exactly the same $\{a_{1,j_1}, a_{2,j_2}, ..., a_{n,j_n}\}$.

Also, for any two such a_{k,j_k},a_{p,j_p}, let an inversion be 1 if $j_k < j_p$ and 0 if $j_k > j_p$.

Then for any prod(A), associate a sign(A) = $(-1)^{\text{total number of inversions in prod(A)}}$.

Then the determinant of A:

$$\det(A) = \sum_{\text{all prod(A)}} \text{prod}(A) * \text{sign}(A)$$

Example

Let A = $\begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 1 \\ 5 & -2 & 3 \end{bmatrix}$.

det(A) = (1*1*3)$(-1)^0$ + (1*-2*1)$(-1)^1$ + (-1*2*3)$(-1)^1$ + (-1*-2*3)$(-1)^2$ + (5*2*1)$(-1)^2$ + (5*1*3)$(-1)^3$ = 12

Theorem 1.5.7: Cofactor Expansion

Let A $\in M_{n\times n}(\mathbb{R})$. Let A_{ij} be A, but the i-th row and j-th column removed.
Then for a fixed i $\in$ {1,...,n}:

$$\det(A) = (-1)^{i+1}a_{i1}\det(A_{i1}) + (-1)^{i+2}a_{i2}\det(A_{i2}) + ... + (-1)^{i+n}a_{in}\det(A_{in})$$

Or for a fixed j $\in$ {1,...,n}:

$$\det(A) = (-1)^{1+j}a_{1j}\det(A_{1j}) + (-1)^{2+j}a_{2j}\det(A_{2j}) + ... + (-1)^{n+j}a_{nj}\det(A_{nj})$$

Proof

For any n by n matrix A, each prod(A) must contain n a_{ij} where each a_{ij}'s i,j is different from another a_{ij}'s i,j. Thus, each prod(A) must contain only one a_{ij} in each row and column.
There are n possibles a_{ij} choices in the first column and by choosing any such one, then that row is eliminated for choice in the following columns. Thus, there are n-1 possible a_{ij} choices in the second column and by choosing any such one, then that row is also eliminated for choice in the following columns. Repeating the pattern, then there are n*(n-1)*(n-2)*...*1 = n! total unique prod(A) combinations. In the cofactor expansion, let choose a fixed i. The case for a fixed j is analogous. For a fixed i, the cofactor expansion iterates through each of the n columns in row i so there are n unique a_{ij}. For each a_{ij}, the A_{ij} has the i-th row and j-th column removed so A_{ij} is a (n-1) by (n-1) matrix and thus, there are (n-1)! unique prod(A_{ij}) combinations as proved earlier. Since each A_{ij} removes a different j-th column, then each prod(A_{ij}) from different columns are unique. Thus, the n unique a_{ij} has (n-1)! unique prod(A_{ij}) combinations so there are n*(n-1)! = n! unique prod(A) combinations. Thus, the prod(A) combinations in the cofactor expansion must be equivalent to the prod(A) combinations in the original determinant.
For the fixed i, let fixed j $\in$ {1,...,n}:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,j-1} & a_{1,j} & a_{1,j+1} & \cdots & a_{1,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & a_{i-1,2} & a_{i-1,3} & \cdots & a_{i-1,j-1} & a_{i-1,j} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i,1} & a_{i,2} & a_{i,3} & \cdots & a_{i,j-1} & a_{i,j} & a_{i,j+1} & \cdots & a_{i,n} \\ a_{i+1,1} & a_{i+1,2} & a_{i+1,3} & \cdots & a_{i+1,j-1} & a_{i+1,j} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,j-1} & a_{n,j} & a_{n,j+1} & \cdots & a_{n,n} \end{bmatrix}$$

In the original determinant, each prod(A) associates sign(A) = $(-1)^{\#\text{ inversions in prod(A)}}$. As proven earlier, each prod(A) is expressed in the coefactor expansion. So for any prod(A) that contains a_{ij} with the fixed i,j, then from the $a_{ij}\det(A_{ij})$ in the cofactor expansion, the $\det(A_{ij})$ consists of the other a_{ij} in the prod(A) since none of the other a_{ij} can exist in row i or column j by definition of the determinant and thus, $\det(A_{ij})$ must account for all the inversions exclusively between the other a_{ij}. To account for the inversions between the other a_{ij} and the fixed a_{ij}, refer to the matrix above. The only a_{ij} which contributes an inversion with the fixed a_{ij} must be in the lower left and upper right of the matrix by defintion of the determinant. Let A = $\#a_{ij}$ in upper left, B = $\#a_{ij}$ in upper right, C = $\#a_{ij}$ in lower left, and D = $\#a_{ij}$ in lower right.
Sinch each prod(A) must have a a_{ij} in each row and column, then:

A+B = i-1 A+C = j-1 $\Rightarrow$ B+C = i+j-2-2A

Thus, sign(A) = $(-1)^{B+C} = (-1)^{i+j-2-2A} = (-1)^{i+j}(-1)^{-2}(-1)^{-2A} = (-1)^{i+j}$ which is the coefficient in the cofactor expansion and thus, the cofactor expansion is calculated in the same way as the original determinant and thus, have the same value.

1.6 Different Coordinate Systems

Definition 1.6.1: Polar Coordinates

Thus far, all vectors has been in the Cartesian (i.e. rectangular (x,y)) System. However, vectors can also be expressed in the Polar (i.e. circular) System.

For any point (x,y), a right triangle can be drawn by adding a perpendicular line from the x-axis to (x,y). Thus:

$\mathrm{r} = \sqrt{x^2+y^2}$ $\quad \mathrm{x} = \mathrm{r}\cos(\theta)$ $\quad \mathrm{y} = \mathrm{r}\sin(\theta)$

Thus, the polar coordinates can express points as (r, θ).

To convert from polar to rectangular:

$\mathrm{x} = \mathrm{r}\cos(\theta)$ $\quad \mathrm{y} = \mathrm{r}\sin(\theta)$

To convert from rectangular to polar:

$r^2 = x^2 + y^2$ $\quad \tan(\theta) = \frac{y}{x}$

Definition 1.6.2: Cylindrical Coordiantes

While polar coordinates are the circular equivalent to $\mathbb{R}^2$, cylindrical coordinates are the circular equivalent to $\mathbb{R}^3$.

Cylindrical coordinates are expressed as (r, θ, z) where:

$\mathrm{x} = \mathrm{r}\cos(\theta)$ $\quad \mathrm{y} = \mathrm{r}\sin(\theta)$ $\quad \mathrm{z} = \mathrm{z}$

The standard basis vectors for cylindrical coordinates:

$e_r = \frac{xe_1+ye_2}{\sqrt{x^2+y^2}} = \cos(\theta)e_i + \sin(\theta)e_2$

$e_z = e_3$

$e_\theta = e_z \times e_r = -\sin(\theta)e_1 + \cos(\theta)e_2$

Definition 1.6.3: Spherical Coordiantes

Although way to express coordiantes in $\mathbb{R}^3$ is spherical coordiantes.

Spherical coordinates are expressed as (p, θ, ϕ) where:

$\mathrm{x} = p\sin(\phi)\cos(\theta)$ $\quad \mathrm{y} = p\sin(\phi)\sin(\theta)$ $\quad \mathrm{z} = p\cos(\phi)$

To convert from rectangular to spherical:

$p^2 = x^2 + y^2 + z^2$ $\quad \tan(\phi) = \frac{\sqrt{x^2+y^2}}{z}$ $\quad \tan(\theta) = \frac{y}{x}$

To convert from cylindrical to spherical:

$p^2 = r^2 + z^2$ $\quad \tan(\phi) = \frac{r}{z}$ $\quad \theta = \theta$

The standard basis vectors for spherical coordinates:

$e_p = \frac{xe_1+ye_2+ze_3}{\sqrt{x^2+y^2+z^2}} = \sin(\phi)\cos(\theta)e_1 + \sin(\phi)\sin(\theta)e_2 + \cos(\phi)e_3$

$e_\theta = -\sin(\theta)e_1 + \cos(\theta)e_2$

$e_\phi = e_\theta \times e_p = \cos(\phi)\cos(\theta)e_1 + \cos(\phi)\sin(\theta)e_2 - \sin(\phi)e_3$

2 Differentiation

2.1 Limits & Continuity

Definition 2.1.1: Limit

For f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$, let a $\in$ X.

If for every $\epsilon > 0$, there is a $\delta > 0$ such that for all x $\in$ X where $||x-a|| < \delta$:

$||f(x) - L|| < \epsilon$

Then the limit of f(x) as x approaches a is $\lim_{x\to a}$ f(x) = L.

Example

Let f(x,y) = $2x^2 + xy$. Find f(x,y) as (x,y) $\to$ (-1,1).

L = f(-1,1) = 1. Let $\sqrt{(x+1)^2 + (y-1)^2} < \delta$ so $|x+1| < \delta$ and $|y-1| < \delta$. Thus:

$$\begin{aligned}|f(x,y) - L| &= |2x^2 + xy - 1| = |2x^2 - 2 + xy + 1| \\ &= |2(x+1)(x-1) + (x+1)(y+1) - (x+1+y-1)| \\ &\leq 2|x+1| * |x-1| + |x+1| * |y+1| + |x+1| + |y-1| \\ &< 2\delta(\delta+2) + \delta(\delta+2) + 2\delta = 3\delta^2 + 8\delta\end{aligned}$$

Since $\min(3\delta^2 + 8\delta) = \frac{-16}{3} < 0$, then for any $\epsilon > 0$, there is a δ where $3\delta^2 + 8\delta < \epsilon$.
Thus, $|f(x) - L| < 3\delta^2 + 8\delta < \epsilon$.

Theorem 2.1.2: Limits are Unique

If $\lim_{x\to a}$ f(x) = L_1 and $\lim_{x\to a}$ f(x) = L_2, then $L_1 = L_2$.

Proof

Since $\lim_{x\to a}$ f(x) = L_1, there is a δ_1 where for $||x-a|| < \delta_1$, then $||f(x) - L_1|| < \frac{\epsilon}{2}$.
Since $\lim_{x\to a}$ f(x) = L_2, there is a δ_2 where for $||x-a|| < \delta_2$, then $||f(x) - L_2|| < \frac{\epsilon}{2}$.
Let $\delta = \min(\delta_1, \delta_2)$. Then for $||x-a|| < \delta$:

$$||L_1 - L_2|| \leq ||L_1 - f(x)|| + ||f(x) - L_2|| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Theorem 2.1.3: Properties of the Limit

(a) For f,g: $\mathbb{R}^n \to \mathbb{R}^m$, if $\lim_{x\to a}$ f(x) = A and $\lim_{x\to a}$ g(x) = B, then:

$\lim_{x\to a}$ (f+g)(x) = A+B

Proof

Since $\lim_{x\to a}$ f(x) = A, there is a δ_1 where for $||x-a|| < \delta_1$, then:

$||f(x) - A|| < \frac{\epsilon}{2}$

Since $\lim_{x\to a}$ g(x) = B, there is a δ_2 where for $||x-a|| < \delta_2$, then:

$||g(x) - B|| < \frac{\epsilon}{2}$

Let $\delta = \min(\delta_1, \delta_2)$. Then for $||x-a|| < \delta$:

$||(f+g)(x) - (A+B)|| = ||f(x) + g(x) - A - B||$

$\leq ||f(x) - A|| + ||g(x) - B|| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$

(b) For f: $\mathbb{R}^n \to \mathbb{R}^m$, if $\lim_{x\to a}$ f(x) = A and scalar c $\in \mathbb{R}$, then:

$\lim_{x\to a}$ cf(x) = cA

Proof

Since $\lim_{x\to a}$ f(x) = A, there is a δ where for $||x-a|| < \delta$, then:

$||f(x) - A|| < \frac{\epsilon}{c}$

Then, $||cf(x) - cA|| = c||f(x) - A|| < c\frac{\epsilon}{c} = \epsilon$.

(c) For f,g: $\mathbb{R}^n \to \mathbb{R}$, if $\lim_{x\to a}$ f(x) = A and $\lim_{x\to a}$ g(x) = B, then:
$\lim_{x\to a}$ (fg)(x) = AB

Proof

Note 4fg $= (f+g)^2 - (f-g)^2$.
By part (a), there is a δ where for $||x-a|| < \delta$:
$|(f+g)(x) - (A+B)| < \epsilon$
Then as x $\to$ a:
$|[(f+g)(x)]^2 - [A+B]^2| = |[(f+g)(x) - (A+B)][(f+g)(x) + (A+B)]|$
$= |(f+g)(x) - (A+B)| * |(f+g)(x) + (A+B)| = \epsilon(2(A+B))$
Thus, $\lim_{x\to a}$ $(f+g)^2(x) = (A+B)^2$.
The proof for $\lim_{x\to a}$ $(f-g)^2(x) = (A-B)^2$ is analogous. Thus:
$\lim_{x\to a}$ (fg)(x) = $\lim_{x\to a}$ $\frac{1}{4}[(f+g)^2(x)$-$(f-g)^2(x)]$
$= \frac{1}{4}[(A+B)^2 - (A-B)^2] = \frac{1}{4}4AB =$ AB

(d) For f,g: $\mathbb{R}^n \to \mathbb{R}$, if $\lim_{x\to a}$ f(x) = A and $\lim_{x\to a}$ g(x) = B $\neq$ 0, then:
$\lim_{x\to a}$ $(\frac{\text{f}}{\text{g}})$(x) = $\frac{\text{A}}{\text{B}}$

Proof

Since $\lim_{x\to a}$ g(x) = B, then there is a δ where for $||x-a|| < \delta$:
$|g(x) - B| < \epsilon$
Thus, as x $\to$ a:
$|\frac{1}{g(x)} - \frac{1}{B}| = |\frac{B-g(x)}{Bg(x)}| = |B - g(x)| * |\frac{1}{Bg(x)}| < \epsilon\frac{1}{B^2}$
Thus, $\lim_{x\to a}$ $\frac{1}{g(x)} = \frac{1}{B}$. By part (c), then $\lim_{x\to a}$ $(\frac{\text{f}}{\text{g}})$(x) = $\frac{\text{A}}{\text{B}}$.

Theorem 2.1.4: Components of Limits

For f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$, let f(x) = $(f_1(x), ..., f_m(x))$. Then for i = {1,...,m}:
$\lim_{x\to a}$ f(x) = L = $(L_1, ..., L_m)$ if and only if each $\lim_{x\to a}$ $f_i(x) = L_i$

Proof

If $\lim_{x\to a}$ f(x) = L = $(L_1, ..., L_m)$, then there is a δ such that for $||x-a|| < \delta$:
$||f(x) - L|| < \epsilon$
$||(f_1(x), ..., f_m(x)) - (L_1, ..., L_m)|| = \sqrt{(f_1(x) - L_1)^2 + ... + (f_m(x) - L_m)^2} < \epsilon$
Thus, each $|f_i(x) - L_i| < \epsilon$ for $||x-a|| < \delta$ so $\lim_{x\to a}$ $f_i(x) = L_i$.

If each $\lim_{x\to a}$ $f_i(x) = L_i$, then there are δ_i such that for $||x-a|| < \delta_i$:
$|f_i(x) - L_i| < \frac{\epsilon}{\sqrt{m}}$
Let $\delta = \min(\delta_1, ..., \delta_m)$. Then for $||x-a|| < \delta$:
$||f(x) - L|| = ||(f_1(x), ..., f_m(x)) - (L_1, ..., L_m)||$
$= \sqrt{(f_1(x) - L_1)^2 + ... + (f_m(x) - L_m)^2} < \sqrt{\sum_{i=1}^m (\frac{\epsilon}{\sqrt{m}})^2} = \sqrt{\epsilon^2} = \epsilon$

Definition 2.1.5: Continuity

For f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$, let a $\in$ X.
Then f is continuous at a if $\lim_{x\to a}$ f(x) = f(a).
If f is continuous at every x $\in$ X, then f is continuous on X.

Theorem 2.1.6: Properties of Continuity

(a) If f,g: X $\subset \mathbb{R}^n \to \mathbb{R}^m$ are continuous at a $\in$ X, then f+g is continuous at a

Proof

> Since $\lim_{x \to a}$ f(x) = f(a) and $\lim_{x \to a}$ g(x) = g(a), by theorem 2.1.3(a), then A = f(a) and B = g(a). Thus, $\lim_{x \to a}$ (f+g)(x) = f(a)+f(b).

(b) If f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$ is continuous at a $\in$ X and scalar c $\in \mathbb{R}$, then cf is continuous at a

Proof

> Since $\lim_{x \to a}$ f(x) = f(a), by theorem 2.1.3(b), then A = f(a).
> Thus, $\lim_{x \to a}$ cf(x) = cf(a).

(c) If f,g: X $\subset \mathbb{R}^n \to \mathbb{R}$ are continuous at a $\in$ X, then fg is continuous at a

Proof

> Since $\lim_{x \to a}$ f(x) = f(a) and $\lim_{x \to a}$ g(x) = g(a), by theorem 2.1.3(c), then A = f(a) and B = g(a). Thus, $\lim_{x \to a}$ (fg)(x) = f(a)f(b).

(d) If f,g: X $\subset \mathbb{R}^n \to \mathbb{R}$ are continuous at a $\in$ X where g(x) $\neq$ 0, then $\frac{\mathrm{f}}{\mathrm{g}}$ is continuous at a

Proof

> Since $\lim_{x \to a}$ f(x) = f(a) and $\lim_{x \to a}$ g(x) = g(a), by theorem 2.1.3(d), then A = f(a) and B = g(a) $\neq$ 0. Thus, $\lim_{x \to a}$ $(\frac{\mathrm{f}}{\mathrm{g}})$(x) = $\frac{\mathrm{f(a)}}{\mathrm{g(b)}}$.

Theorem 2.1.7: Components of Continuity

For f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$, let f(x) = $(f_1(x), ..., f_m(x))$. Then for i = {1,...,m}:

f is continuous at a $\in$ X if and only if each f_i is continuous at a

Proof

> If f is continuous at a, then $\lim_{x \to a}$ f(x) = f(a) = $(f_1(a), ..., f_m(a))$. By theorem 2.1.4, then L = $(f_1(a), ..., f_m(a))$ so each $L_i = f_i(a)$. Thus, for each i = {1,...,m}:
> $\lim_{x \to a} f_i(x) = L_i = f_i(a)$
>
> ---
>
> If each f_i is continuous at a, then for i = {1,...,m}, $\lim_{x \to a} f_i(x) = f_i(a)$. By theorem 2.1.4, then L = $(f_1(a), ..., f_m(a))$. Thus:
> $\lim_{x \to a}$ f(x) = L = $(f_1(a), ..., f_m(a))$ = f(a)

Theorem 2.1.8: Composite of Continuous functions are Continuous

If f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$ and g: Y $\subset \mathbb{R}^m \to \mathbb{R}^k$ are continuous where f(X) $\subset$ Y, then g $\circ$ f = g(f): X $\subset \mathbb{R}^n \to \mathbb{R}^k$ is continuous

Proof

> For any a $\in$ X and any $\delta > 0$, there is a $\eta > 0$ such that for $||x - a|| < \eta$:
> $||f(x) - f(a)|| < \delta$
> Since f(X) $\subset$ Y, then for any x $\in$ X, then f(x) $\in$ Y.
> For any f(a) $\in$ Y and any $\epsilon > 0$, there is a $\delta > 0$ such that for $||y - f(a)|| < \delta$:
> $||g(y) - g(f(a))|| < \epsilon$
> Thus, for $||x - a|| < \eta$, then $||g(f(x)) - g(f(a))|| < \epsilon$.

2.2 Differentiability

Definition 2.2.1: Partial Derivative

For f: X $\subset \mathbb{R}^n \to \mathbb{R}$, let x = $(x_1, ..., x_n) \in$ X.

For i = {1,...,n}, the partial derivative of f with respect to x_i:

$D_i f = \frac{\partial f}{\partial x_i} = f_{x_i}(x) = \lim_{h \to 0} \frac{f(x_1,...,x_i+h,...,x_n) - f(x_1,...,x_n)}{h}$

Theorem 2.2.2: Tangent Plane

For f: X $\subset \mathbb{R}^2 \to \mathbb{R}$, let z = f(x,y).

The tangent plane at (a,b,f(a,b)) has an equation of the form:

z = f(a,b) + $f_x(a,b)(x-a)$ + $f_y(a,b)(y-b)$

Proof

Since $f_x(a,b)$ which is the change in z for every change in x is a tangent vector to f in direction of x at (a,b), then (1,0,$f_x(a,b)$) is parallel to the tangent plane. Similarily, (0,1,$f_y(a,b)$) is parallel to the tangent plane.

Thus, (1,0,$f_x(a,b)$) $\times$ (0,1,$f_y(a,b)$) = $(-f_x(a,b), -f_y(a,b), 1)$ is orthogonal to the tangent plane. Thus, for any (x,y,z) in the plane:

$(-f_x(a,b), -f_y(a,b), 1) \cdot$ [(x,y,z) - (a,b,f(a,b))] = 0

$-f_x(a,b)(x-a) - f_y(a,b)(y-b) + z - f(a,b) = 0$

z = f(a,b) + $f_x(a,b)(x-a)$ + $f_y(a,b)(y-b)$

Definition 2.2.3: Differentiability in $\mathbb{R}^2 \to \mathbb{R}$

f: X $\subset \mathbb{R}^2 \to \mathbb{R}$ is differentiable at x $\in$ X if there is an A $\in M_{1\times 2}(\mathbb{R})$ such that for h $\in$ X:

$\lim_{h\to 0} \frac{|f(x+h)-f(x)-Ah|}{||h||} = 0$

Then, the derivative of f at x is Df(x) = A = $\begin{bmatrix} \frac{\partial f}{\partial x}(x,y) & \frac{\partial f}{\partial y}(x,y) \end{bmatrix}$.

If f is differentiable at every x $\in$ X, then f is differentiable on X.

Theorem 2.2.4: Continuous partials imply Differentiability

If f: X $\subset \mathbb{R}^2 \to \mathbb{R}$ has continuous partial derivatives at (a,b),

then f is differentiable at (a,b)

Proof

Since $f_x(x,y), f_y(x,y)$ is continuous at (a,b), then for $\epsilon > 0$, there is a $\delta > 0$ where for $||(x,y)-(a,b)|| < \delta$:

$|f_x(x,y) - f_x(a,b)| < \epsilon \qquad |f_y(x,y) - f_y(a,b)| < \epsilon$

Then for h = $h_1 e_1 + h_2 e_2$:

$\lim_{h\to 0} \frac{|f(a+h_1,b+h_2)-f(a,b)-[f_x(a,b)h_1+f_y(a,b)h_2]|}{||h||}$

$= \lim_{h\to 0} \frac{|f(a+h_1,b+h_2)-f(a+h_1,b)+f(a+h_1,b)-f(a,b)-[f_x(a,b)h_1+f_y(a,b)h_2]|}{||h||}$

Since $f_x(x,y), f_y(x,y)$ exist, then by the Mean Value Theorem, there are $t_1 \in (0,h_1)$ and $t_2 \in (0,h_2)$ such that:

$f(a+h_1,b) - f(a,b) = h_1 * f_x(a+t_1,b)$

$f(a+h_1,b+h_2) - f(a+h_1,b) = h_2 * f_y(a+h,b+t_2)$

Thus, for $||h-(a,b)|| < \delta$:

$\lim_{h\to 0} \frac{|f(a+h_1,b+h_2)-f(a,b)-[f_x(a,b)h_1+f_y(a,b)h_2]|}{||h||}$

$= \lim_{h\to 0} \frac{|h_2 * f_y(a+h,b+t_2)+h_1 * f_x(a+t_1,b)-[f_x(a,b)h_1+f_y(a,b)h_2]|}{||h||}$

$= \lim_{h\to 0} \frac{|h_2 * [f_y(a+h,b+t_2)-f_y(a,b)]+h_1 * [f_x(a+t_1,b)-f_x(a,b)]|}{||h||} < \lim_{h\to 0} \frac{||h||\epsilon + ||h||\epsilon}{||h||} = 2\epsilon$

Theorem 2.2.5: Differentiability implies Continuity

If f: X $\subset \mathbb{R}^2 \to \mathbb{R}$ is differentiable at (a,b), then f is continuous at (a,b)

Proof

If f is differentiable at (a,b), then $\lim_{h\to 0} \frac{|f((a,b)+h)-f(a,b)-Ah|}{||h||} = 0$.
Thus, as h $\to$ 0, then A $= \frac{f((a,b)+h)-f(a,b)}{||h||}$. So:

$$f((a,b)+h)-f(a,b) = [f((a,b)+h)-f(a,b)]\frac{||h||}{||h||} = A||h|| \to 0$$

Thus, f is continuous at (a,b).

2.3 Differentiability in Higher Dimensions

Definition 2.3.1: Differentiability in $\mathbb{R}^n \to \mathbb{R}$

Differentiability can be extended for $\mathbb{R}^n$.
f: X $\subset \mathbb{R}^n \to \mathbb{R}$ is differentiable at x $\in$ X if there is an A $\in M_{1\times n}(\mathbb{R})$ such that for h $\in$ X:

$$\lim_{h\to 0} \frac{|f(x+h)-f(x)-Ah|}{||h||} = 0$$

Then, the derivative of f at x is Df(x) = A = $\begin{bmatrix} \frac{\partial f}{\partial x_1}(x) & \frac{\partial f}{\partial x_2}(x) & \dots & \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$.
If f is differentiable at every x $\in$ X, then f is differentiable on X.

The gradient of f:

$$\nabla f(x) = (\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)) = [\text{Df(x)}]^T$$

Definition 2.3.2: Differentiability in $\mathbb{R}^n \to \mathbb{R}^m$

Differentiability can be extended into $\mathbb{R}^m$.
f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$ where f = $(f_1, \dots, f_m)$ is differentiable at x $\in$ X if there is an A $\in M_{m\times n}(\mathbb{R})$ such that for h $\in$ X:

$$\lim_{h\to 0} \frac{|f(x+h)-f(x)-Ah|}{||h||} = 0$$

Then, the derivative of f at x is Df(x) = A = $\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \dots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}$.
If f is differentiable at every x $\in$ X, then f is differentiable on X.

Theorem 2.3.3: Differentiability implies Continuity in Higher Dimensions

If f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a $\in$ X, then f is continuous at a

Proof

Analogous to theorem 2.2.5. Replace (a,b) with a = $(a_1, \dots, a_n)$.

Theorem 2.3.4: Continuous partials imply differentiability in Higher Dimensions

If f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$ has continuous partial derivatives, $\frac{\partial f_i}{\partial x_j}$, at a $\in$ X for j = {1,...,n} and i = {1,...,m}, then f is differentiable at a

Proof

Analogous to theorem 2.2.4. Instead, h = $h_1 e_1 + ... + h_n e_n$ where:

$$\lim_{h\to 0} \frac{|f(x+h)-f(x)-Ah|}{||h||} = \lim_{h\to 0} \sum_{i=1}^m \frac{|f_i(x+h)-f_i(x)-[\sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(x)h_j]|}{||h||}$$

and add each $f_i(x+h_1e_1+...+h_ke_k)$ and apply Mean Value Theorem and continuity of partial derivatives analogously as performed in theorem 2.2.4.

Theorem 2.3.5: Components of Differentiability

f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$ where f = $(f_1, ..., f_m)$ is differentiable at a $\in$ X if and only if each f_i is differentiable at a for i = {1,...,m}

Proof

Note $\lim_{h\to 0} \frac{|f(x+h)-f(x)-Ah|}{||h||} = \lim_{h\to 0} \sum_{i=1}^m \frac{|f_i(x+h)-f_i(x)-[\sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(x)h_j]|}{||h||}$.

If f is differentiable at a, then for any $\epsilon > 0$:

$$\lim_{h\to 0} \frac{|f(x+h)-f(x)-Ah|}{||h||} < \epsilon$$

So $\lim_{h\to 0} \frac{|f_i(x+h)-f_i(x)-[\sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(x)h_j]|}{||h||} < \epsilon$ for each i = {1,...,m} and thus, each f_i is differentiable at a.

If each $\lim_{h\to 0} \frac{|f_i(x+h)-f_i(x)-[\sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(x)h_j]|}{||h||} < \frac{\epsilon}{m}$ for i = {1,...,m}, then:

$$\lim_{h\to 0} \frac{|f(x+h)-f(x)-Ah|}{||h||} < \lim_{h\to 0} \sum_{i=1}^m \frac{\epsilon}{m} = \epsilon$$

Thus, f is differentiable at a.

Theorem 2.3.6: Properties of Differentiability

(a) For f,g: X $\subset \mathbb{R}^n \to \mathbb{R}^m$, if f,g are differentiable at a $\in$ X, then:
f+g is differentiable at a where D(f+g)(a) = Df(a) + Dg(a)

Proof

Since f,g are differentiable at a $\in$ X, by theorem 2.3.5, then for i = {1,...,m}:

$$\mathrm{D}(f+g)_i(\mathrm{a}) = \begin{bmatrix} D_1(f_i+g_i)(a) & D_2(f_i+g_i)(a) & \dots & D_n(f_i+g_i)(a) \end{bmatrix}$$
$$= \begin{bmatrix} D_1 f_i(a) & D_2 f_i(a) & \dots & D_n f_i(a) \end{bmatrix} + \begin{bmatrix} D_1 g_i(a) & D_2 g_i(a) & \dots & D_n g_i(a) \end{bmatrix}$$
$$= \mathrm{D}f_i(\mathrm{a}) + \mathrm{D}g_i(\mathrm{a})$$

(b) For f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$, if f is differentiable at a $\in$ X and scalar c $\in \mathbb{R}$, then:
cf is differentiable at a where D(cf)(a) = cDf(a)

Proof

Since f is differentiable at a $\in$ X, by theorem 2.3.5, then for i = {1,...,m}:

$$\mathrm{D}(cf)_i(\mathrm{a}) = \begin{bmatrix} D_1(cf_i)(a) & D_2(cf_i)(a) & \dots & D_n(cf_i)(a) \end{bmatrix}$$
$$= \begin{bmatrix} cD_1 f_i(a) & cD_2 f_i(a) & \dots & cD_n f_i(a) \end{bmatrix} = \mathrm{cD}f_i(\mathrm{a})$$

(c) For f,g: X $\subset \mathbb{R}^n \to \mathbb{R}$, if f,g are differentiable at a $\in$ X, then:
fg is differentiable at a where D(fg)(a) = Df(a)g(a) + f(a)Dg(a)

Proof

Since f,g are differentiable at a $\in$ X:

$$D(fg)(a) = \begin{bmatrix} D_1(fg)(a) & D_2(fg)(a) & ... & D_n(fg)(a) \end{bmatrix}$$
$$= \begin{bmatrix} D_1f(a)g(a)+f(a)D_1g(a) & D_2f(a)g(a)+f(a)D_2g(a) & ... & D_nf(a)g(a)+f(a)D_ng(a) \end{bmatrix}$$
$$= \begin{bmatrix} D_1f(a)g(a) & D_2f(a)g(a) & ... & D_nf(a)g(a) \end{bmatrix} + \begin{bmatrix} f(a)D_1g(a) & f(a)D_2g(a) & ... & f(a)D_ng(a) \end{bmatrix}$$
$$= Df(a)g(a) + f(a)Dg(a)$$

(d) For f,g: X $\subset \mathbb{R}^n \to \mathbb{R}$, if f,g are differentiable at a $\in$ X where g(a) $\neq$ 0, then:
$\frac{f}{g}$ is differentiable at a where $D(\frac{f}{g})(a) = \frac{Df(a)g(a)-f(a)Dg(a)}{[g(a)]^2}$

Proof

Since f,g are differentiable at a $\in$ X:

$$D(\tfrac{f}{g})(a) = \begin{bmatrix} D_1(\frac{f}{g})(a) & D_2(\frac{f}{g})(a) & ... & D_n(\frac{f}{g})(a) \end{bmatrix}$$
$$= \begin{bmatrix} \frac{D_1f(a)g(a)-f(a)D_1g(a)}{[g(a)]^2} & \frac{D_2f(a)g(a)-f(a)D_2g(a)}{[g(a)]^2} & ... & \frac{D_nf(a)g(a)-f(a)D_ng(a)}{[g(a)]^2} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{D_1f(a)g(a)}{[g(a)]^2} & \frac{D_2f(a)g(a)}{[g(a)]^2} & ... & \frac{D_nf(a)g(a)}{[g(a)]^2} \end{bmatrix} - \begin{bmatrix} \frac{f(a)D_1g(a)}{[g(a)]^2} & \frac{f(a)D_1g(a)}{[g(a)]^2} & ... & \frac{f(a)D_1g(a)}{[g(a)]^2} \end{bmatrix}$$
$$= Df(a)\frac{g(a)}{[g(a)]^2} - \frac{f(a)}{[g(a)]^2}Dg(a) = \frac{Df(a)g(a)-f(a)Dg(a)}{[g(a)]^2}$$

Definition 2.3.7: Partial Derivatives of Higher Orders

The second order partial derivative of f in respect to x_i:

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_i}) = f_{x_i x_i}(x) = \lim_{h\to 0} \frac{f_{x_i}(x_1,...,x_i+h,...,x_n)-f_{x_i}(x_1,...,x_n)}{h}$$

The mixed partial derivative of f in respect to first x_i, then x_j:

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j}(\frac{\partial f}{\partial x_i}) = f_{x_i x_j}(x) = \lim_{h\to 0} \frac{f_{x_i}(x_1,...,x_j+h,...,x_n)-f_{x_i}(x_1,...,x_n)}{h}$$

In general, for f: X $\subset \mathbb{R}^n \to \mathbb{R}$, the k-th order partial derivative of f in respect to $x_{i_1}, ..., x_{i_k}$ in such order for k = {1,...,n}:

$$\frac{\partial^k f}{\partial x_{i_k}...\partial x_{i_1}} = \frac{\partial}{\partial x_{i_k}}...\frac{\partial f}{\partial x_{i_1}} = f_{x_{i_1}...x_{i_k}}(x)$$
$$= \lim_{h\to 0} \frac{f_{x_{i_1}...x_{i_{k-1}}}(x_1,...,x_{i_k}+h,...,x_n)-f_{x_{i_1}...x_{i_{k-1}}}(x_1,...,x_n)}{h}$$

Definition 2.3.8: Smoothness

For k = {1,...,n}, f: X $\subset \mathbb{R}^n \to \mathbb{R}$ is C^k if all partial derivatives of order 1 to k exist and are continuous on X.

If f has continuous partial derivatives of all order, then f is smooth (i.e C^∞).

For f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$ where f = $(f_1, ..., f_m)$, then f is C^k if each f_i is C^k for i = {1,...,m}.

Theorem 2.3.9: Clairaut's Theorem

If f: X $\subset \mathbb{R}^n \to \mathbb{R}$ is C^k, then:

$\frac{\partial^k f}{\partial x_{i_1}...\partial x_{i_k}} = \frac{\partial^k f}{\partial x_{j_1}...\partial x_{j_k}}$

Proof

If the claim holds true for C^2, then replace f with $f_{x_{i_p}}$ for any p = {1,...,k} and since f is C^k, then $f_{x_{i_p}}$ is C^{k-1} and apply the theorem again. Repeating the process k times, the result holds true by induction. Now the proof for C^2:

Since f is C^2, then f_x, f_y, f_{xy}, f_{yx} exist and are continuous.
Let d(x,y) = $f(x+h_1, y+h_2) - f(x, y+h_2) - f(x+h_1, y) + f(x,y)$.
Since f_x exist, then by the Mean Value Theorem, there is a $t_1 \in (0, h_1)$ where:
d(x,y) = $h_1 * (f_x(x+t_1, y+h_2) - f_x(x+t_1, y))$
Since f_y exist, then by the Mean Value Theorem, there is a $t_2 \in (0, h_2)$ where:
d(x,y) = $h_1 * h_2 * f_{xy}(x+t_1, y+t_2)$
Since f_{xy} is continuous, then since $(t_1, t_2) \to (0,0)$ as $(h_1, h_2) \to (0,0)$:
$f_{xy}(x,y) = \lim_{(h_1,h_2)\to(0,0)} f_{xy}(x+h_1, x+h_2)$
$= \lim_{(h_1,h_2)\to(0,0)} f_{xy}(x+t_1, x+t_2) = \lim_{(h_1,h_2)\to(0,0)} \frac{d(x,y)}{h_1 h_2}$
Rearrange d(x,y) = $f(x+h_1, y+h_2) - f(x+h_1, y) - f(x, y+h_2) + f(x,y)$.
Since f_y exist, by the Mean Value Theorem, there is a $s_2 \in (0, h_2)$ where:
d(x,y) = $h_2 * (f_y(x+h_1, y+s_2) - f_y(x, y+s_2))$
Since f_x exist, by the Mean Value Theorem, there is a $s_1 \in (0, h_1)$ where:
d(x,y) = $h_2 * h_1 * f_{yx}(x+s_1, y+s_2)$
Since f_{yx} is continuous, then since $(s_1, s_2) \to (0,0)$ as $(h_1, h_2) \to (0,0)$:
$f_{yx}(x,y) = \lim_{(h_1,h_2)\to(0,0)} f_{yx}(x+h_1, x+h_2)$
$= \lim_{(h_1,h_2)\to(0,0)} f_{xy}(x+s_1, x+s_2) = \lim_{(h_1,h_2)\to(0,0)} \frac{d(x,y)}{h_1 h_2}$
Thus, $f_{xy}(x,y) = f_{yx}(x,y)$.

Theorem 2.3.10: Chain Rule

Let f: X $\subset \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $x_0 \in$ X and g: f(X) $\subset$ Y $\subset \mathbb{R}^m \to \mathbb{R}^k$ be differentiable at f(x_0).
Then g $\circ$ f = g(f): X $\subset \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at x_0 such that:
D[g(f(x_0))] = Dg(f(x_0)) Df(x_0)

Proof

Since f is differentiable at x_0 and g is differentiable at f(x_0), then there is a A = Df(x_0) and B = Dg(f(x_0)) such that:
f(x_0+h) - f(x_0) = Ah + $r_A(h)$ where $\lim_{h\to 0} \frac{|r_A(h)|}{|h|} = 0$
g(f(x_0)+k) - g(f(x_0)) = Bk + $r_B(k)$ where $\lim_{k\to 0} \frac{|r_B(k)|}{|k|} = 0$
Let k = f(x_0+h) - f(x_0). Thus:
g(f(x_0+h)) - g(f(x_0)) - BAh
= g(f(x_0)+k) - g(f(x_0)) - BAh = Bk + $r_B(k)$ - BAh = B(k - Ah) + $r_B(k)$
= B(f(x_0+h) - f(x_0) - Ah) + $r_B(k)$ = B$r_A(h)$ + $r_B(k)$
Since f is differentiable at x_0, then f is continuous at x_0 and thus, $\lim_{h\to 0}$ k = 0.
Since $\lim_{h\to 0} \frac{|r_A(h)|}{|h|} = 0$ and $\lim_{k\to 0} \frac{|r_A(k)|}{|k|} = 0$, then:
$\lim_{h\to 0} \frac{|g(f(x_0+h)) - g(f(x_0)) - BAh|}{|h|} \leq \lim_{h\to 0} (||B|| \frac{|r_A(h)|}{|h|} + \frac{|r_B(k)|}{|h|}) = 0 + 0 = 0$
Thus, D[g(f(x_0))] = BA = Dg(f(x_0)) Df(x_0).

Theorem 2.3.11: Relationship between rectangular and polar partials

For (x,y) = $(r\cos(\theta), r\sin(\theta))$:

$\frac{\partial}{\partial r} = \cos(\theta)\frac{\partial}{\partial x} + \sin(\theta)\frac{\partial}{\partial y}$

$\frac{\partial}{\partial \theta} = -r\sin(\theta)\frac{\partial}{\partial x} + r\cos(\theta)\frac{\partial}{\partial y}$

Thus:

$\frac{\partial}{\partial x} = \cos(\theta)\frac{\partial}{\partial r} - \frac{\sin(\theta)}{r}\frac{\partial}{\partial \theta}$

$\frac{\partial}{\partial y} = \sin(\theta)\frac{\partial}{\partial r} + \frac{\cos(\theta)}{r}\frac{\partial}{\partial \theta}$

Proof

Let z = g(r,θ). Then let z = f(x,y) such that $(r\cos(\theta), r\sin(\theta))$ = (x,y).

By theorem 2.3.10:

D[g(r,θ)] = D(f(x,y)) D(x(r,θ),y(r,θ))

$$\begin{bmatrix} \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{bmatrix} = \begin{bmatrix} \cos(\theta)\frac{z\partial}{\partial x} + \sin(\theta)\frac{z\partial}{\partial y} \\ -r\sin(\theta)\frac{z\partial}{\partial x} + r\cos(\theta)\frac{z\partial}{\partial y} \end{bmatrix}$$

Thus:

$\frac{\partial}{\partial r} = \cos(\theta)\frac{\partial}{\partial x} + \sin(\theta)\frac{\partial}{\partial y}$

$\frac{\partial}{\partial \theta} = -r\sin(\theta)\frac{\partial}{\partial x} + r\cos(\theta)\frac{\partial}{\partial y}$

Then:

$-r\cos(\theta)\frac{\partial}{\partial r} + \sin(\theta)\frac{\partial}{\partial \theta} = -r\cos^2(\theta)\frac{\partial}{\partial x} - r\sin^2(\theta)\frac{\partial}{\partial x} = -r\frac{\partial}{\partial x}$

$r\sin(\theta)\frac{\partial}{\partial r} + \cos(\theta)\frac{\partial}{\partial \theta} = r\sin^2(\theta)\frac{\partial}{\partial y} + r\cos^2(\theta)\frac{\partial}{\partial y} = r\frac{\partial}{\partial y}$

2.4 Directional Derivative

Definition 2.4.1: Directional Derivative

Let f: X $\subset \mathbb{R}^n \to \mathbb{R}$ be differentiable at a $\in$ X. Then the directional derivative of f at a in the direction of vector v $\in \mathbb{R}^n$:

$D_v f(a) = \lim_{h\to 0} \frac{f(a+hv)-f(a)}{||hv||}$

Theorem 2.4.2: Relationship between Directional Derivative and Gradient

Let f: X $\subset \mathbb{R}^n \to \mathbb{R}$ be differentiable at a $\in$ X. Then the directional derivative of f at a in the direction of vector v $\in \mathbb{R}^n$:

$D_v f(a) = \nabla f(a) \cdot \frac{v}{||v||}$

If v is a unit vector, then $D_v f(a) = \nabla f(a) \cdot$ v.

Proof

Let y(t) = a+tv for t $\in (-\infty, \infty)$. Then by theorem 2.3.10:

$D_v f(a) = \lim_{h\to 0} \frac{f(a+hv)-f(a)}{||hv||} = \lim_{h\to 0} \frac{f(y(h))-f(y(0))}{|h|}\frac{1}{||v||}$ = Df(y(0)) Dy(0) $\frac{1}{||v||}$

= Df(a)) $\frac{v}{||v||}$ = [Df(a))]$^T \cdot \frac{v}{||v||} = \nabla f(a) \cdot \frac{v}{||v||}$

Theorem 2.4.3: Direction of Steepest Ascent

The directional derivative $D_v f(a) = \nabla f(a) \cdot \frac{v}{||v||}$ is:

Maximized when v is in the same direction as $\nabla f(a)$ with value $||\nabla f(a)||$

Minimized when v is in the opposite direction of $\nabla f(a)$ with value $-||\nabla f(a)||$

Proof

By theorem 1.2.3, $D_v f(a) = \nabla f(a) \cdot \frac{v}{||v||} = ||\nabla f(a)||\ ||\frac{v}{||v||}|| \cos(\theta) = ||\nabla f(a)|| \cos(\theta)$.
where $\theta \in [0, \pi]$ is the angle between $\nabla f(a)$ and $|\frac{v}{||v||}|$.
Since $D_v f(a)$ is maximized at $||\nabla f(a)||$ when $\theta = 0$, then $\nabla f(a)$ and v points in the same direction. Also, $D_v f(a)$ is minimized at $-||\nabla f(a)||$ when $\theta = \pi$, then $\nabla f(a)$ and v points in opposite directions.

Theorem 2.4.4: Gradient is orthogonal to the surface

If f: $X \subset \mathbb{R}^n \to \mathbb{R}$ is C^1, then for any x_0 where f(x_0) = c for constant c $\in \mathbb{R}$, then $\nabla f(x_0)$ is orthogonal to the surface f(x) = c at x_0.

Proof

For surface f(x) = c, let curve C(t) = $(x_1(t), ..., x_n(t))$ where C(0) = x_0 be defined such that f(C(t)) = c. Thus:

$\frac{d}{dt}$ f(C(t)) = $\frac{d}{dt}$ c = 0

Let v be the tangent vector to C(t) at x_0. Then by theorem 2.3.10, for t = 0:

$\frac{d}{dt}$ f(C(0)) = Df(C(0)) C'(0) = $\nabla f(C(0)) \cdot$ C'(0) = $\nabla f(x_0) \cdot$ v

Since $\nabla f(x_0) \cdot$ v = 0 where v is tangent to C(t) which lies on surface f(x) = c and thus, is tangent is f(x) = c, then $\nabla f(x_0)$ is orthogonal to f(x) = c at x_0.

3 Vector-Valued Functions

References

[1] Susan Jane Colley, *Vector Calculus (4th Edition)*, ISBN-13: 978-0321780652