

MA113 Notes

Brady Exoo

September 1, 2022 - November 17, 2022

Abstract

Notes for MA113 (Mutlivariable Calculus) taught by Dr. Holder.

Contents

1	Coordinate Systems and Polar Functions	4
1.1	Rectangular Coordinates	4
1.2	Polar Coordinates	4
1.3	Transformations	4
1.4	Polar Graphs	5
1.5	Differentiating Polar Functions	5
1.5.1	Chain Rule	6
1.5.2	Polar Derivatives	6
1.6	Arc Lengths	6
1.6.1	Review	6
1.6.2	Polar Arc Length	7
1.7	Polar Area	7
2	Vectors	7
2.1	Fundamentals	7
2.2	Norms	8
2.3	Dot Product	9
2.3.1	Angles in the Dot Product	9
2.3.2	Projections	9
2.4	Three Dimensions	10
2.4.1	Set Notation	10
2.5	Cross Product	10
2.5.1	Properties	11

2.6	Matrices	11
2.6.1	2 by 2	11
2.6.2	3 by 3	11
2.6.3	Calculating the Cross Product With Matrices	12
2.7	Parametric Equations	12
2.7.1	Some Notation	12
2.7.2	Crossing Sets	12
2.8	Lines in 3 Dimensions	13
2.9	Distance Between Line and Point	13
2.10	Planes	14
2.10.1	Given Three Points	14
2.10.2	Finding line Given Two Planes	14
2.10.3	Distance From Point To Plane	14
3	Parametric Equations	15
3.1	Example	15
3.2	Derivatives of Parametric Equations	15
3.3	Arclength	15
3.4	Constant Length	15
3.5	Special Unit Vectors	16
3.6	Curvature	16
3.7	Decomposing Acceleration	17
3.8	Integration	17
3.9	Projectile Motion	17
4	Functions of Several Variables	18
4.1	Limits	18
4.1.1	Delta Epsilon Definition of The Limit	18
4.2	Partial Derivatives	19
4.2.1	Linearity	19
4.2.2	Notation	20
4.3	Chain Rule	20
4.4	Level Curves	20
4.5	Directional Derivatives	21
4.6	Matrices	21
4.7	Hessian	21
4.8	Taylor Expansions	22
4.8.1	Example	22
4.8.2	Taylor In 2 Variables	22
4.9	Tangent Planes	23

4.9.1	Level Surfaces	23
5	Optimization	23
5.1	Unconstrained optimization	23
5.1.1	Classification	24

1 Coordinate Systems and Polar Functions

“This your last chance. After this there is no turning back. You take the blue pill, the story ends. You wake up in your bed and believe whatever you want to. You take the red pill, you stay in Wonderland, and I show you how deep the rabbit hole goes.”

Morpheus, *The Matrix*

1.1 Rectangular Coordinates

- (x, y)
- Every point has only one set of coordinates
- Unique representation!

1.2 Polar Coordinates

- (r, θ)
 - r is the distance from the origin
 - θ is the angle the radius faces
- Lacks unique representation
 - Can add 2π to any angle and get the same point
 - Can make radius negative and add π to the angle and get the same point

These coordinates will be helpful for setting up integrals later in the course. Often by a change of coordinates we turn a complicated integral into an elegant one!

1.3 Transformations

Definition. $\tan \theta$ is the distance from the intersection of the extended radius and the vertical tangent to the tangent point.

- $x^2 + y^2 = r^2$
- $\tan \theta = \frac{y}{x}$

- $x = 0 \implies \theta \in \{\frac{\pi}{2}, \frac{3\pi}{2}\}$
- $\theta = \tan^{-1} \frac{y}{x}$
 - $\tan \theta$ is the distance from the intersection of the extended radius and the vertical tangent to the tangent point
 - This is problematic! Range of arctan is $(-\frac{\pi}{2}, \frac{\pi}{2})$, so we do not get all 360 degrees.
- $\theta = \begin{cases} \tan^{-1} \frac{y}{x} & x > 0 \\ \frac{\pi}{2} & x = 0, y > 0 \\ \frac{3\pi}{2} & x = 0, y < 0 \\ \tan^{-1} \frac{y}{x} + \pi & x < 0 \end{cases}$
- $r \cos \theta = x$
- $r \sin \theta = y$

1.4 Polar Graphs

- Graphs are of the form $r(\theta)$, r is a function of θ .
- $r(\theta) = \cos \theta - 1$ is a cardioid.
- $r(\theta) = \cos(2\theta)$ is a 4-flower.
 - $\cos(n\theta)$ has $2n$ petals if n is even, n petals if n is odd.
 - $r = \theta$ is the Archimedes spiral.
 - * This can be generalized by $r = a\theta$.

1.5 Differentiating Polar Functions

- We come to a problem, we wish to find $\frac{dy}{dx}$ but our functions are in terms of r and θ !
- $\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta}$ by the Chain Rule.
 - Not because we can cancel out the $d\theta$ terms!

1.5.1 Chain Rule

The default way the Chain Rule is portrayed is

$$\frac{d}{dx}f(g(x)) = f'(g(x)) \cdot g'(x).$$

Other notation is

$$\frac{df}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}.$$

The Chain Rule allows us to change variables which we do not wish to differentiate by.

1.5.2 Polar Derivatives

Now we can substitute y and x with our earlier transformations to get

$$\frac{dy}{dx} = \frac{\frac{d}{d\theta}r \sin \theta}{\frac{d}{d\theta}r \cos \theta}.$$

1.6 Arc Lengths

1.6.1 Review

In Calculus II, we learned the arclength of a function from a to b as

$$\int_a^b \sqrt{1 + (f'(x))^2} dx$$

This is somewhat related to

$$\frac{d}{dx} \langle x, f(x) \rangle = \langle 1, f'(x) \rangle$$

$$||\langle 1, f'(x) \rangle|| = \sqrt{1 + (f'(x))^2}.$$

Here we are essentially integrating the magnitude of the derivative of the vector, which gives us the arclength (think about this visually with the Pythagorean Theorem). We can also think of this as “vectorizing” the derivative.

1.6.2 Polar Arc Length

We want to find the arc length of a function in terms of $r(\theta)$.

$$\begin{aligned}
 s &= \int_{\theta_1}^{\theta_2} \left\| \frac{d}{d\theta} \langle r \cos(\theta), r \sin(\theta) \rangle \right\| d\theta \\
 &= \int_{\theta_1}^{\theta_2} \left\| \langle r' \cos(\theta) - r \sin(\theta), r' \sin(\theta) + r \cos(\theta) \rangle \right\| d\theta \\
 &= \int_{\theta_1}^{\theta_2} \sqrt{(r' \cos(\theta) - r \sin(\theta))^2 + (r' \sin(\theta) + r \cos(\theta))^2} d\theta \\
 &= \int_{\theta_1}^{\theta_2} \sqrt{r^2 + (r')^2} d\theta
 \end{aligned}$$

This is the same as the previous arc length formula, just with substitutions.

1.7 Polar Area

We integrate all the sectors of the function, each of which approximates the sector of a circle, each of which has angle $\Delta\theta$ and radius $r(\theta)$. The area of each sector is

$$(\pi r^2) \left(\frac{\Delta\theta}{2\pi} \right) = \frac{1}{2} r^2 \Delta\theta.$$

Thus, the total area is

$$\frac{1}{2} \int_{\theta_1}^{\theta_2} r(\theta)^2 d\theta.$$

We can subtract these integrals as needed to find areas between two curves:

$$\frac{1}{2} \int_{\theta_1}^{\theta_2} r_2(\theta)^2 - r_1(\theta)^2 d\theta.$$

2 Vectors

2.1 Fundamentals

- The point $A(1, 2)$ to the point $B(6, 5)$ is denoted as the vector $\overrightarrow{AB} = \langle 6 - 1, 5 - 2 \rangle = \langle 5, 3 \rangle$.
- Vectors have magnitude and direction.
- Vectors in angle bracket notation are typically assumed to start at the origin.

- Arrow hats represent vectors, some books boldface instead.
 - Column vectors are also used: $\begin{bmatrix} 5 \\ 3 \end{bmatrix}$
- Adding vectors is the same as adding their components.
 - Geometrically, we add tail of one vector to the head of another.
 - Can only add vectors in the same number of dimensions.
 - $\langle 1, 2, 0 \rangle + \langle -1, 4, 8 \rangle = \langle 0, 6, 3 \rangle$
- In subtraction we add the negative of the vector.
- Scalar multiplication of a vector is the same as multiplying each component by that scalar.
 - $3 \langle 1, 2, 0 \rangle = \langle 3, 6, 0 \rangle$

2.2 Norms

- Norms are the length or magnitude of a vector
 - Notated as $\|\vec{v}\|$.
- $\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.
- $\langle 1, 3, 2 \rangle = \sqrt{14} \left\langle \frac{1}{\sqrt{14}}, \frac{3}{\sqrt{14}}, \frac{2}{\sqrt{14}} \right\rangle$.
 - Generally, this is useful because we separate the norm from the direction (unit vector).
 - General form is $\|\vec{v}\| \left(\frac{\vec{v}}{\|\vec{v}\|} \right)$.
 - This is called normalizing a vector.
- Special Vectors
 - $\vec{i} = \langle 1, 0, 0 \rangle$
 - $\vec{j} = \langle 0, 1, 0 \rangle$
 - $\vec{k} = \langle 0, 0, 1 \rangle$
 - Always assume we are in three dimensions (every vector has 0 \vec{k}).

2.3 Dot Product

The dot product is defined as

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

This has a lot of cool properties, for one:

$$\vec{v} \cdot \vec{w} = \|\vec{v}\| \|\vec{w}\| \cos \theta.$$

Also, the dot product is commutative and can be distributed.

Example 2.1. $\langle 1, 3, 2 \rangle \cdot \langle 3, 0, 2 \rangle = 3 + 0 + 2 = 7$.

2.3.1 Angles in the Dot Product

Additionally, by the law of cosines on our vector subtraction example, we get:

$$\begin{aligned} \|\vec{v}\|^2 + \|\vec{u}\|^2 - 2\|\vec{u}\|\|\vec{v}\|\cos \theta &= \|\vec{v} - \vec{u}\|^2 \\ &= (\vec{v} - \vec{u}) \cdot (\vec{v} - \vec{u}) \\ &= \vec{v} \cdot \vec{v} - 2\vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{u} \end{aligned}$$

Simplifying, we get

$$\cos \theta = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|}$$

This may also be represented as

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

So, the dot product gives us:

- Norms
- Angles
- Distance (metric)

2.3.2 Projections

We wish to find the component of one vector \vec{u} onto another vector \vec{v} . This is written as $\text{proj}_{\vec{v}} \vec{u}$, which is the projection of \vec{u} onto \vec{v} . This projection

will be a scalar multiple of \vec{v} (the vector we project onto), which can be represented as $\text{proj}_{\vec{v}} \vec{u} = \alpha \vec{v}$. Recall that

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

In our case $\theta = 90^\circ$, so $\vec{u} \cdot \vec{v} = 0$. Note that $\vec{u} \perp \vec{v} \iff \vec{u} \cdot \vec{v} = 0$. We know that

$$\begin{aligned} \alpha \vec{v} \cdot (\vec{u} - \alpha \vec{v}) &= 0 \\ \vec{u} &= \alpha \vec{v} \cdot \vec{v} \\ \alpha &= \frac{\vec{v} \cdot \vec{u}}{\vec{v} \cdot \vec{v}}. \end{aligned}$$

Therefore,

$$\text{proj}_{\vec{v}} \vec{u} = \left(\frac{\vec{v} \cdot \vec{u}}{\vec{v} \cdot \vec{v}} \right) \vec{v}.$$

A way to remember this is that \vec{v} is the vector we are projecting onto.

2.4 Three Dimensions

We orient our axes by the right hand rule, where the x axis is coming towards us.

2.4.1 Set Notation

- A sphere would be $\{\vec{u} : \|\vec{u}\| = k\}$.
 - $x^2 + y^2 + z^2 = k^2$
- A sphere centered at another point would be $\{\vec{u} : \|\vec{u} - \vec{w}\| = k\}$, where \vec{w} is the vector from the origin to the center of the sphere.

$$\circ (x - a)^2 + (y - b)^2 + (z - c)^2 = k^2$$

2.5 Cross Product

Definition. $\vec{v} \times \vec{w} = \|\vec{v}\| \|\vec{w}\| \sin \theta \vec{n}$, where \vec{n} is a unit vector which satisfies the right-hand rule.

n must be perpendicular to both vectors \vec{v} and \vec{w} . Alternatively,

$$\|\vec{v} \times \vec{w}\| = \|\vec{v}\| \|\vec{w}\| |\sin \theta|$$

2.5.1 Properties

- Not commutative, can be clearly seen by the right-hand rule, which gives opposite values if we reverse the order.

$$\circ \vec{v} \times \vec{w} = -(\vec{w} \times \vec{v})$$

- The magnitude of the cross product is the area of the parallelogram formed by the two vectors.
- $\vec{v} \times (\alpha \vec{v}) = \langle 0, 0, 0 \rangle$
- $\|\vec{v}\| \|\vec{w}\| = \|\vec{v} \times \vec{w}\| \iff \vec{v} \perp \vec{w}$.
 - This means $\vec{v} \perp \vec{w} \iff \vec{v} \cdot \vec{w} = 0$.

2.6 Matrices

2.6.1 2 by 2

Given a matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

The determinant of the matrix is $AD - BC$. This is the area of the parallelogram formed by the vectors $\langle A, B \rangle$ and $\langle C, D \rangle$.

2.6.2 3 by 3

Given a matrix

$$\begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix}$$

We first take A and multiply it by the determinant of the matrix formed by crossing out its row and column (any place a rook could be placed and not hit A).

$$A(EI - FH)$$

Every other term is negative, so we subtract the B term and add the C term, giving us

$$A(EI - FH) - B(DI - FG) + C(DH - GE)^1$$

¹We can calculate this easier by expanding by a row (or column) with 0s, we don't always need to do the top row! The process will be the same, and the +s and -s form a checkerboard pattern.

as the determinant of this matrix. This is the volume of the parallelepiped formed by the vectors $\langle A, B, C \rangle$, $\langle D, E, F \rangle$ and $\langle G, H, I \rangle$. This is also known as the triple product, but it is really just

$$\langle A, B, C \rangle \times \langle D, E, F \rangle \cdot \langle G, H, I \rangle.$$

2.6.3 Calculating the Cross Product With Matrices

To calculate the cross product of $\vec{v} = \langle 0, 1, 2 \rangle$ and $\vec{w} = \langle 1, 1, 3 \rangle$ we use the determinant.

$$\vec{v} \times \vec{w} = \det^2 \left(\begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ 0 & 1 & 2 \\ 1 & 1 & 3 \end{bmatrix} \right)$$

We can dot the resulting vector with any of our original vectors and we should get 0, this verifies that our calculation is correct.

2.7 Parametric Equations

Generally, a parametric function is one that takes a variable t and maps it to a point (x, y, z) . They are usually used to model the position of moving objects. It can also be thought of as mapping a straight line onto a flexible line in space.

2.7.1 Some Notation

Usually we write functions as $y(x) = x^2$, however a more specific notation would be

$$y : \mathbb{R} \rightarrow \mathbb{R}_+ : x \mapsto x^2$$

This means we have a function y such that it maps real numbers onto positive real numbers, and it maps a value x onto x^2 . In this way, we can write a parametric function as

$$\vec{r} : \mathbb{R} \rightarrow \mathbb{R}^3 : t \mapsto \langle x(t), y(t), z(t) \rangle$$

2.7.2 Crossing Sets

We take the Cartesian product (similar to the cross product) of sets like this.

$$\{1, 2, 3\} \times \{a, b\} = \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}$$

²This is slightly a misuse of notation, usually we should not have a vector in this matrix.

In this same way,

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}.$$

Also,

$$\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}.$$

2.8 Lines in 3 Dimensions

Defining lines as $y = mx + b$ does not work in three dimensions, since a line in three dimensions does not have a slope. Instead, we define lines with a point and a direction.

$$\vec{r}(t) = \vec{A} + t\vec{v}$$

where \vec{A} is the origin point of the line and \vec{v} is the direction of the line, which we can think of as taking the place of the slope. Usually we combine these two terms into a single vector, like

$$\langle 2 - t, 1 - 2t, 4 - 4t \rangle.$$

Alternatively, we may write this as

$$\begin{aligned} x(t) &= 2 - t \\ y(t) &= 1 - 2t \\ z(t) &= 4 - 4t. \end{aligned}$$

The former is the vector form, and the latter equations are parametric equations.

2.9 Distance Between Line and Point

We wish to find the distance d between a line and a point. The line is defined by a point A and a vector \vec{v} , and the point is B . Let θ be the angle between the vectors \vec{v} and \vec{AB} . We quickly see that

$$\sin \theta = \frac{d}{\|\vec{AB}\|}.$$

Now we may see hints of the cross product in here.

$$\begin{aligned} d &= \frac{\|\vec{v}\| \|\vec{AB}\| \sin \theta}{\|\vec{v}\|} \\ &= \boxed{\frac{\|\vec{AB} \times \vec{v}\|}{\|\vec{v}\|}}. \end{aligned}$$

2.10 Planes

A plane is defined as, given a point A and a vector \vec{v} rooted at A , the set of all points P such that $\overrightarrow{AP} \perp \vec{v}$, or $\overrightarrow{AP} \cdot \vec{v} = 0$. If we write A and P as vectors, this is equivalent to

$$\vec{v} \cdot (\vec{P} - \vec{A}) = 0$$

Usually, this is multiplied out, and the equation of the plane is written in the form

$$ax + by + cz = d$$

In this case, the normal vector turns out to be $\langle a, b, c \rangle$.

2.10.1 Given Three Points

When given three points A, B , and C , we first need to find the normal vector. This can be done by taking $\overrightarrow{AB} \times \overrightarrow{AC}$. Then, the rest of the calculation proceeds as shown prior.

2.10.2 Finding line Given Two Planes

To find the line of intersection given two planes, first find a point on the line. Then find the normal vectors \vec{n} and \vec{m} , and take the cross product, giving a line which lies on both planes.

2.10.3 Distance From Point To Plane

We have a plane defined by a point A and a normal vector \vec{n} , and a point B . Let d be the length of the projection of \overrightarrow{AB} onto \vec{n} . Note that d will be the desired length, as we can move B horizontally parallel to the plane onto \vec{n} and conserve the distance to the plane. Then, by simple trigonometry and substitutions we get

$$\begin{aligned} d &= \frac{\|\vec{n}\| \|\overrightarrow{AB}\| |\cos \theta|}{\|\vec{n}\|} \\ &= \boxed{\frac{|\vec{n} \cdot \overrightarrow{AB}|}{\|\vec{n}\|}} \end{aligned}$$

as desired.

3 Parametric Equations

3.1 Example

Assume we have a function

$$\vec{r}(t) = \langle \cos t, \sin t, t \rangle.$$

The first two coordinates we recognize as the equation for a circle, and the third is simply linear, so the resulting graph is a spiral in the $+z$ direction.

3.2 Derivatives of Parametric Equations

The derivative of $r(t) = \langle x(t), y(t), z(t) \rangle$ is defined

$$\begin{aligned} \frac{d}{dt} \vec{r}(t) &= \lim_{h \rightarrow 0} \frac{\vec{r}(t+h) - \vec{r}(t)}{h} \\ &= \lim_{h \rightarrow 0} \left\langle \frac{x(t+h) - x(t)}{h}, \frac{y(t+h) - y(t)}{h}, \frac{z(t+h) - z(t)}{h} \right\rangle \\ &= \langle x'(t), y'(t), z'(t) \rangle \end{aligned}$$

If $\vec{r}(t)$ is position, then $\vec{r}'(t)$ is velocity and $\vec{r}''(t)$ is acceleration. The magnitude of velocity is speed.

3.3 Arclength

Let $s(t)$ be the function for the arclength of a parametric function. Simply,

$$s(t) = \int_{t_1}^t \|\vec{v}(\tau)\| d\tau.$$

Additionally,

$$\begin{aligned} s'(t) &= \frac{d}{dt} \int_{t_1}^t \|\vec{v}(\tau)\| d\tau \\ &= \|\vec{v}(t)\|. \end{aligned}$$

3.4 Constant Length

If $\vec{r}(\alpha)$ has the property that $\|\vec{r}'(\alpha)\| = k$, then

$$\vec{r} \cdot \vec{r}' = k^2.$$

Differentiating, we get

$$\vec{r}' \cdot \vec{r} + \vec{r} \cdot \vec{r}' = 0$$

so

$$2\vec{r} \cdot \vec{r}' = 0$$

and

$$\vec{r} \perp \vec{r}'.$$

3.5 Special Unit Vectors

We define the normal tangent vector as

$$\mathbf{T} = \frac{\vec{r}'}{\|\vec{r}'\|}$$

and the unit normal vector as

$$\mathbf{N} = \frac{dT/d\alpha}{\|dT/d\alpha\|}.$$

The binormal vector is defined as

$$\mathbf{B} = \mathbf{T} \times \mathbf{N}.$$

These vectors are all perpendicular to one another, and they create a coordinate system for us to describe motion along a curve.

3.6 Curvature

We define curvature as the change in the tangent vector per unit length, or:

$$\kappa = \left\| \frac{dT}{ds} \right\|.$$

This is intuitively simple, but deceptively hard to calculate! Thankfully, we have the chain rule:

$$\begin{aligned} \kappa &= \left\| \frac{dT/d\alpha}{ds/d\alpha} \right\| \\ &= \left\| \frac{dT/d\alpha}{\|\vec{v}(\alpha)\|} \right\| \end{aligned}$$

So, to calculate the curvature we have to parameterize both the tangent vector and the speed, making differentiation simple. Some notable examples are:

- The curvature of a circle is $\frac{1}{r}$.

3.7 Decomposing Acceleration

We can parameterize acceleration as

$$\begin{aligned}
 \vec{a} &= a_t T + a_n N \\
 &= (\vec{a} \cdot \mathbf{T})T + (\vec{a} \cdot N)N \\
 &= \left(\frac{d}{dt} \|v\| \right) T + (\kappa \|v\|^2) N \\
 &= (\text{proj}_T \vec{a})T + (\text{proj}_N \vec{a})N
 \end{aligned}$$

a_T is called the scalar component of acceleration in the tangential direction, and a_N is called the normal component of acceleration in the tangential direction. When $a_N = 0$, the acceleration is entirely along the axis of motion. When $a_T = 0$, the acceleration is orthogonal to the the direction of motion.

3.8 Integration

When integrating any parametric or vector function, we take the antiderivative of each component individually.

$$\begin{aligned}
 \int \langle t, \sin t, e^{2t} \rangle dt &= \left\langle \frac{1}{2}t^2, -\cos t, \frac{1}{2}e^{2t} + \vec{c} \right\rangle \\
 &= \left\langle \frac{1}{2}t^2 + c_1, -\cos t + c_2, \frac{1}{2}e^{2t} + c_3 \right\rangle
 \end{aligned}$$

3.9 Projectile Motion

Instead of using kinematic equations, we can solve projectile motion through vectors and integration. As always, we have

$$\vec{a} = \langle 0, -9.8 \rangle$$

Then by integrating we get

$$\begin{aligned}
 \vec{v} &= \int \vec{a} dt = \langle 0, -9.8t \rangle + \vec{v}_0 \\
 &= \langle v_{0x}, -9.8t + v_{0y} \rangle
 \end{aligned}$$

Integrating again, we get

$$\vec{r} = \langle v_{0x}t, -4.9t^2 + v_{0y}t \rangle + \vec{r}_0$$

4 Functions of Several Variables

We will now study functions of the type

$$y : \mathbb{R}^2 \rightarrow \mathbb{R} \text{ or } y : \mathbb{R}^3 \rightarrow \mathbb{R}.$$

Take for example the function $f(x, y) = x^2 + y^2$. This function maps every point on the plane to another point on a number line. This function can also be visualized as $z = x^2 + y^2$, now it is a graph in 3 dimensions which forms a bowl.

4.1 Limits

$\frac{x}{x} \neq 1$ but $\lim_{x \rightarrow 0} \frac{x}{x} = \lim_{x \rightarrow 0} 1$, remember your notation! Limits are different in two dimensions, there is no left hand side or right hand side. This makes showing the limit exists more difficult. Limits in two dimensions are denoted as

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x, y).$$

Often we prohibit a certain line in these two dimensions, such as in the case

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - y^2}{x - y}$$

we exclude the line $y = x$.

4.1.1 Delta Epsilon Definition of The Limit

$$\lim_{x \rightarrow x_0} f(x) = L$$

means that

$$\forall \epsilon > 0, \exists \delta > 0 \ni,$$

$$0 < |x - x_0| < \delta \implies |f(x) - L| < \epsilon$$

In English, this says that for all $\epsilon > 0$, we can choose a delta such that if the left bottom statement is true the right bottom statement is also true.

4.2 Partial Derivatives

Assume we have a function $f(x, y)$ and we wish to find its derivative. Derivative usually tell us the ratio of the rate of change of f to the rate of change of x , our input variable. However, in this case we have multiple input variables. To take the derivative of just one variable we have partial derivatives, which are defined in our case as

$$\frac{\partial f}{\partial x} = \lim_{h \rightarrow 0} \frac{f(x + h, y) - f(x, y)}{h}$$

$$\frac{\partial f}{\partial y} = \lim_{h \rightarrow 0} \frac{f(x, y + h) - f(x, y)}{h}$$

Second partial derivatives are written as

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left(\frac{\partial}{\partial x} (f) \right)$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial}{\partial x} (f) \right)$$

Also,

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$$

4.2.1 Linearity

A function f is linear if the following are true:

- $f(kx) = kf(x)$
- $f(x + y) = f(x) + f(y)$.

Therefore not all lines are linear! The derivative operator is linear because

$$\frac{d}{dx}(f(x) + g(x)) = \frac{df}{dx}(x) + \frac{dg}{dx}(x)$$

$$\frac{d}{dx}kf(x) = k\frac{df}{dx}(x).$$

4.2.2 Notation

- $\frac{\partial f}{\partial x} = f_x$
- $\frac{\partial f}{\partial y} = f_y$
- $\frac{\partial^2 f}{\partial y \partial x} = f_{xy}$
- $\frac{\partial^4 f}{\partial z \partial w^2 \partial x} = f_{xw^2wz}$

4.3 Chain Rule

Assume we have a function $f(x(t), y(t)) = z$. Note that f can be written as either a function of just t or also a function of $x(t)$ and $y(t)$. Then the Chain Rule states that

$$\begin{aligned} \frac{df}{dt} &= \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} \\ &= \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \cdot \left\langle \frac{dx}{dt}, \frac{dy}{dt} \right\rangle \\ &= \frac{d}{dt} f(\vec{r}(t)) \end{aligned}$$

$\left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle = \nabla f(x, y)$, or the gradient (more about this later), so the second line becomes

$$\frac{df}{dt} = \nabla f(x, y) \cdot r'.$$

If we have $f(x(u, w), y(u, w), z(u, w))$, to take the derivative we do

$$\begin{aligned} \frac{\partial f}{\partial u} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial u} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial u} \\ &= \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle \cdot \left\langle \frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u} \right\rangle \\ &= \nabla f(x, y, z) \cdot \left\langle \frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u} \right\rangle \end{aligned}$$

4.4 Level Curves

The level curve for $f(x, y) = k$ is the set of all solutions to this equation. They can be visualized in 2 variables as a plane slicing the function at $z = k$. If

we let $f(\vec{r}(t)) = k$ where $\vec{r}(t)$ describes all point on the level curve, then differentiating gives us

$$\nabla f(\vec{r}(t)) \cdot \vec{r}'(t) = 0$$

by the chain rule. This means that these two vectors are orthogonal. Therefore, **the gradient at a point is always perpendicular to the tangent line of the level curve at that point.**

4.5 Directional Derivatives

We wish to find the rate of change of some function $f(x, y)$ in a particular direction $\vec{d} = \langle d_1, d_2 \rangle$. Using the limit definition of the derivative, this turns out to be

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{f(x + hd_1, y + hd_2) - f(x, y)}{h} &= \frac{d}{dh} f(\langle x, y \rangle + h\vec{d}) \Big|_{h=0} \\ &= \nabla f(\langle x, y \rangle + h\vec{d}) \cdot \vec{d} \Big|_{h=0} \\ &= \nabla f(x, y) \cdot \vec{d} \end{aligned}$$

This is called the directional derivative. This also means that the gradient is the direction of steepest increase, since the above function is maximized when \vec{d} is in the same direction as the gradient. As the function changes faster, the gradient increases.

4.6 Matrices

When multiplying matrices, we multiply rows by columns. For example:

$$\begin{bmatrix} 2 & -1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \cdot 3 + -1 \cdot 5 \\ 4 \cdot 3 + 1 \cdot 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 17 \end{bmatrix}$$

To multiply two matrices, the first must have the same number of columns as the second matrix has rows.

4.7 Hessian

The Hessian of a function is denoted as $\nabla^2 f$. or H_f The Hessian of a function is

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}.$$

Recall that the gradient is

$$\left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle.$$

We obtain the Hessian by taking the partial derivatives of x and y of each component of the gradient. The Hessian gives us all the second order information about a function.

4.8 Taylor Expansions

Recall from Calculus II that the Taylor expansion is

$$f(x) \approx \overbrace{f(x_0) + f'(x_0)(x - x_0)}^{\text{Function of } x, y=T(x)} + \frac{1}{2}f''(x_0)(x - x_0)^2 + \dots$$

First order Taylor polynomial

This says that we can perfectly approximate any function given perfect information about a single point and all the derivatives at that point.

4.8.1 Example

We can use the Taylor expansion to gain intuition on some elementary derivatives:

$$\begin{aligned} \frac{d}{dx}e^x &= \frac{d}{dx} \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \right) \\ &= 0 + 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \\ &= e^x \end{aligned}$$

4.8.2 Taylor In 2 Variables

In 2 dimensions, we have

$$f(x, y) = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot \langle x - x_0, y - y_0 \rangle + \frac{1}{2} \underbrace{\langle x - x_0, y - y_0 \rangle}_{1 \times 2 \text{ matrix}} \nabla^2 f(x_0, y_0) \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} + \dots$$

This is as far as we will go. Note that the matrix multiplication of the last term all cancels out and we are left with a scalar term.

4.9 Tangent Planes

We know from our Taylor expansion that the tangent plane at a point should be

$$z = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot \langle x - x_0, y - y_0 \rangle.$$

We can now rewrite this as

$$\begin{aligned} 0 &= -(z - z_0) + \langle f_x, f_y \rangle \cdot \langle x - x_0, y - y_0 \rangle \\ 0 &= \langle f_x, f_y, -1 \rangle \cdot \langle x - x_0, y - y_0, z - z_0 \rangle \end{aligned}$$

Note that the vector $\langle f_x, f_y, -1 \rangle$ is the normal vector to the plane, and therefore also the coefficients of the tangent plane in standard form.

4.9.1 Level Surfaces

The tangent plane to a level surface $f(x, y, z) = k$ is

$$\frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0) + \frac{\partial f}{\partial z}(z - z_0) = 0$$

or alternatively,

$$\left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle \cdot (\langle x, y, z \rangle - \langle x_0, y_0, z_0 \rangle) = 0$$

5 Optimization

There are two cases of optimization:

- (Easier) Unconstrained
- (Harder) Constrained

5.1 Unconstrained optimization

Given a function $z = f(x, y)$ we see that if we construct a tangent plane at the local maximums and minimums, the normal vectors at those points will be $\langle 0, 0, -1 \rangle$. Therefore, local max/min $\implies \nabla f(x, y) = \langle 0, 0 \rangle$. Critical points are solutions to the equation $\nabla f(x, y) = \langle 0, 0 \rangle$. These can either be local minimums, local maximums, or saddle points.

5.1.1 Classification

We can classify critical points based on the Hessian ($\nabla^2 f$) at that point. Assume the Hessian is diagonalizable into

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

λ_1 and λ_2 are called eigenvalues. Then,

- Both positive \implies local minimum
- Both negative \implies local maximum
- Split signs \implies saddle point
- $\lambda_1 \cdot \lambda_2 = 0 \implies$ indeterminate

Alternatively, we can classify them by the determinant of the Hessian (note that this will give mostly the same results as using the eigenvalues because the determinant of the diagonalized Hessian is the product of the eigenvalues.)

- $\det(\nabla^2 f(x^*, y^*)) < 0 \implies$ saddle point.
- $\det(\nabla^2 f(x^*, y^*)) > 0$ and $\frac{\partial^2 f}{\partial x^2} > 0 \implies$ local minimum.
- $\det(\nabla^2 f(x^*, y^*)) > 0$ and $\frac{\partial^2 f}{\partial x^2} < 0 \implies$ local maximum.
- $\det(\nabla^2 f(x^*, y^*)) = 0 \implies$ inconclusive.