Contents

1 Week 1 **5**
1.1 Lines & Planes 5
1.1.1 Lines in 2-D 5
1.1.2 Planes in 3-D 5
1.1.3 Lines in 3D 5
1.2 Curves & surfaces in 3D 6
1.2.1 Classical Quadratic Surfaces 6
1.3 Coordinates & Points 6
1.4 Vectors 7
1.4.1 Vector Algebra 7
1.4.2 Norm of a Vector 7
1.4.3 Application of vectors in parameterization 8
1.4.4 Basis vectors 8

2 Week 2 **9**
2.1 The Dot Product 9
2.1.1 Orthogonality and the dot product 9
2.1.2 Orthogonal Projection 9
2.1.3 Implicit definition of hyperplanes using the dot product 10
2.2 The Cross Product 10
2.2.1 Properties of the cross product 10
2.2.2 Geometric understanding of the cross product 10
2.2.3 Distance to a line and the cross product 10
2.2.4 The Scalar Triple product 10
2.3 Intro to vector calculus 11
2.3.1 Derivative of a vector 11
2.3.2 Arclength of a parameterized curve 11
2.3.3 Rules of vector differentiation 11
2.4 Vectors & Physical Motion 12
2.4.1 Finding the velocity and position vector with acceleration 12
2.4.2 Unit tangent and normal vectors 12
2.4.3 The osculating circle 13
2.4.4 The unit binormal vector 13

3 Week 3 **14**
3.1 Matrix Equations 14
3.1.1 Linear System of Equations 14
3.2 Gaussian elimination, or row reduction 14
3.3 Inverse matrices 15
3.3.1 Inverse 2-by-2 matrices 15
3.3.2 Uniqueness of the inverse 15
3.3.3 Solving a system of linear equations with the inverse 15
3.3.4 Finding the inverse via Gaussian Elimination or row reduction 16
3.3.5 Finding the inverse of a block diagonal matrix 16

4 Week 4 **17**
4.1 Linear Transformations . . . 17
4.1.1 Diagonal matrices as rescaling function . . . 17
4.1.2 Shearing matrices . . . 17
4.1.3 Rotation matrices . . . 17
4.1.4 Matrix composition . . . 18
4.1.5 Example in Euler Angles, 3D, and 4D matrices . . . 18
4.2 Change in Basis . . . 18
4.2.1 Basis vectors . . . 18
4.2.2 Coordinate changes . . . 19
4.3 The Determinant . . . 19
4.3.1 Computing the determinant . . . 19
4.3.2 By formula . . . 19
4.3.3 By Minor Expansion . . . 20
4.3.4 By Row Operations . . . 20
4.3.5 By Block Diagonal Matrices . . . 20
4.4 Geometry of determinants . . . 20
4.4.1 Properties of the determinant . . . 21

5 Week 5 **22**
5.1 Multivariate functions . . . 22
5.1.1 Multivariate functions as non-linear transforms . . . 22
5.1.2 Multivariate functions example - Market Equilibrium . . . 22
5.2 Partial Derivatives . . . 23
5.2.1 Computing the partial derivative and implicit differentiation . . . 23
5.2.2 Partial derivatives and rates of change . . . 23
5.3 The Derivative . . . 23
5.3.1 A note on notation . . . 24
5.3.2 Computing inputs and outputs rates of change . . . 24
5.3.3 Definition of the derivative . . . 24
5.3.4 A Taylor Perspective . . . 25
5.3.5 Existence of the derivative . . . 25

6 Week 6 **26**
6.1 The Chain Rule . . . 26
6.1.1 The classical chain rule . . . 26
6.1.2 The multivariate chain rule . . . 26
6.1.3 Note on notation . . . 26
6.1.4 Classical chain rules as a variation on the ONE single chain rule . . . 27
6.2 Differentiation rules: Linearity of the derivative . . . 28
6.3 Differentiation rules: Product Rule . . . 28
6.3.1 The Material Derivative . . . 29
6.4 The Inverse Function Theorem . . . 29
6.4.1 Inverse multivariate functions . . . 30
6.4.2 Linearization of nonlinear functions . . . 30
6.4.3 The Inverse Rule . . . 30
6.4.4 The Inverse Function Theorem . . . 30
6.5 The Implicit Function theorem . . . 31
6.5.1 Implicit differentiation and implicit functions . . . 31
6.5.2 The Implicit Function Theorem . . . 31

7 Week 7 **32**
7.1 Gradient . . . 32
7.1.1 The level set . . . 32
7.1.2 Definition of the gradient . . . 32
7.1.3 Gradients and level sets . . . 32
7.1.4 Directional Derivatives . . . 33
7.2 Tangent Spaces . . . 33
7.2.1 The tangent space of an implicit surface . . . 33

7.2.2 The tangent space of an parameterized curve . . . 33
7.2.3 The tangent space of a parameterized surface . . . 33
7.2.4 Images, Kernels, and tangent spaces in higher dimensions . . . 34
7.3 Linearization . . . 34
7.3.1 Differentials vs derivatives . . . 34
7.3.2 Relative rates of change and percentage errors . . . 34
7.4 Taylor Series . . . 35
7.4.1 Single variable case . . . 35
7.4.2 Multivariable case . . . 35
7.5 Computing Taylor Series . . . 36
7.5.1 Commonly used Taylor Series . . . 36
7.5.2 Computing the Taylor Expansion of a function with 2 inputs and 1 output . . . 37
7.5.3 Taylor Series and the Chain Rule . . . 37
7.5.4 Using Taylor Series for local analysis . . . 37
7.5.5 Taylor Series up to the second order . . . 37
7.5.6 The Hessian . . . 37

8 Week 8 **38**
8.1 Critical points and extrema . . . 38
8.1.1 Critical points . . . 38
8.1.2 Trace-Determinant Method . . . 38
8.2 Optimization & Linear regression . . . 39
8.2.1 Least-squares method . . . 39
8.2.2 Optimization & Nash Equilibria . . . 40
8.2.3 The Minimax Theorem . . . 41
8.2.4 Constrained Optimization . . . 41
8.2.5 The Lagrange Multiplier . . . 41

9 Integration **42**
9.1 Computing integrals and the Fubini theorem . . . 43
9.2 Double & Triple Integrals . . . 43
9.2.1 Finding the area in 2 dimensions . . . 43
9.2.2 Changing the limits . . . 43
9.3 Averages . . . 44

10 Week 10 **45**
10.1 Centroids & Centers of mass . . . 45
10.1.1 Centroids in 2D . . . 45
10.1.2 Centroids in 3D . . . 45
10.1.3 Center of Mass . . . 45
10.2 Moments of inertia . . . 46
10.2.1 Parallel Axis Theorem . . . 46
10.2.2 Radius of Gyration . . . 46
10.3 Inertia Matrix . . . 46
10.4 Solid Body Mechanics . . . 46
10.5 Probability & Integration . . . 46
10.6 Independence & Covariance . . . 47
10.7 Covariance Matrices . . . 47

11 Week 11 **48**
11.1 Polar & Cylindrical coordinates . . . 48
11.1.1 Polar Coordinates . . . 48
11.1.2 Cylindrical Coordinates . . . 48
11.2 Spherical coordinates . . . 48
11.3 Change in Variables Theorem . . . 49
11.3.1 Linear transformations and the volume element . . . 49
11.3.2 Non-linear transformations and the volume element . . . 49
11.3.3 Change of variables theorem . . . 49
11.4 Surface Integrals . . . 50

11.4.1 Parameterized surfaces . . . 50

12 Week 12 **51**

12.1 Fields . . . 51

12.1.1 Vector Fields . . . 51

12.1.2 Other fields . . . 51

12.2 Path Integrals . . . 51

12.2.1 Integrating a scalar field . . . 51

12.2.2 The standard single-variable integral as a path integral . . . 52

12.2.3 Rules for a scalar path integral . . . 52

12.2.4 Independence of path in scalar fields . . . 52

12.3 Integrating 1-forms . . . 52

12.3.1 1-Forms . . . 52

12.3.2 Gradient vector field, gradient 1-form field, and the derivative . . . 53

12.3.3 Integrating 1-forms . . . 53

12.4 Independence of Path . . . 53

12.5 Detecting gradients . . . 53

12.6 Work, Circulation and Flux . . . 53

12.6.1 Work . . . 53

12.6.2 Circulation . . . 54

12.6.3 Flux . . . 54

13 Week 13 **55**

13.1 Green's Theorem . . . 55

13.1.1 The Theorem . . . 55

13.1.2 Orientation and Boundaries . . . 55

13.1.3 Differentiability, orientation & Green's Theorem . . . 55

13.2 Gradient, Curl & Divergence . . . 56

13.2.1 Curl and divergence densities . . . 56

13.2.2 Green Theorem through the perspective of curl and divergence . . . 57

13.2.3 Divergence in 3-D . . . 57

13.2.4 Curl in 3-D . . . 57

13.3 Differential Forms in 3-D . . . 57

13.3.1 Euclidean Forms . . . 57

13.3.2 The Wedge . . . 58

13.3.3 Differentiation of form fields . . . 58

14 Week 14 **60**

14.1 Integrating 2-forms . . . 60

14.1.1 Computing the integral . . . 60

14.1.2 Computing flux . . . 60

14.2 Gauss' Theorem . . . 61

14.2.1 The Theorem . . . 61

14.3 Stoke's Theorem . . . 61

14.4 Which theorem to use? . . . 62

Chapter 1: Week 1

1.1 Lines & Planes

1.1.1 Lines in 2-D

Recall that Lines in 2-D can be defined in the following forms

- The Slope-Intercept form
$$y = mx + b.$$
where m is the gradient, and b is the y intercept
- The Point-Slope form
$$y - y_0 = m(x - x_0).$$
where (x_0, y_0) is a point on the line
- The Intercept form
$$\frac{x}{a} + \frac{y}{b} = 1.$$
where a is the x intercept and b is the y intercept

1.1.2 Planes in 3-D

For a plane in standard coordinates, the points (x, y, z) can be defined by

- Point-"Slope" form
$$n_x(x - x_0) + n_y(y - y_0) + n_z(z - z_0) = 0.$$
where $\left(\begin{smallmatrix} n_x \\ n_y \\ n_z \end{smallmatrix}\right)$ is a vector normal to the plane, and (x_0, y_0, z_0) is a point on the plane
- Intercept form
$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1.$$
where a, b, c are the respective intercepts on the x, y, z axis.

1.1.3 Lines in 3D

Lines in 3-D can be defined by a system of equations w.r.t parameter t, for example

$$\begin{aligned} x(r) &= 3r - 5 \\ y(r) &= r + 3 \\ z(r) &= -4r + 1 \end{aligned}$$

which is a line that contains the point $\begin{pmatrix} -5 \\ 3 \\ 1 \end{pmatrix}$ and parallel to the vector $\begin{pmatrix} 3 \\ 1 \\ -4 \end{pmatrix}$

1.2 Curves & surfaces in 3D

Planar curves and surfaces in n-d can be described *implicitly* or *parametrically*

implicit	parametric
$y = x^2 - 3$	$x(t) = t$ $y(t) = t^2 - 3$
$x^2 + y^2 = 4$	$x(t) = 2cost$ $y(t) = 2sint$
$z = 3x^2 + y^2 - 5$	$x(t) = s$ $y(t) = t$ $z(t) = 3s^2 + t^2 - 5$
$x^2 + y^2 + z^2 = 4$	$x(t) = 2cosssint$ $y(t) = 2sinssintt$ $z(t) = 2cost$

1.2.1 Classical Quadratic Surfaces

The follow classical quadratic surfaces are implicitly defined by

- Sphere
$$\frac{(x-x_0)^2}{r^2} + \frac{(y-y_0)^2}{r^2} + \frac{(z-z_0)^2}{r^2} = 1$$
- Ellipsoid
$$\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} + \frac{(z-z_0)^2}{c^2} = 1$$
- 1-sheeted hyperboloid
$$\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} - \frac{(z-z_0)^2}{c^2} = 1$$
- 2-sheeted hyperboloid
$$\frac{(x-x_0)^2}{a^2} - \frac{(y-y_0)^2}{b^2} - \frac{(z-z_0)^2}{c^2} = 1$$
- Elliptic Paraboloid
$$z - z_0 = \frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2}$$
- Hyperbolic Paraboloid
$$z - z_0 = \frac{(x-x_0)^2}{a^2} - \frac{(y-y_0)^2}{b^2}$$
- Cones
$$\frac{(z-z_0)^2}{c^2} = \frac{(x-x_0)^2}{a^2} - \frac{(y-y_0)^2}{b^2}$$

1.3 Coordinates & Points

The n-dimensional plane is coordinatized via

$$\mathbb{R}^n = (x_1, x_2, ..., x_{n-1}, x_n).$$

Distance between 2 points

The distance between two points in 2-d is

$$\sqrt{(Q_1 - P_1)^2 + (Q_2 - P_2)^2}$$

The distance between two points in 3-d is

$$\sqrt{(Q_1 - P_1)^2 + (Q_2 - P_2)^2 + (Q_3 - P_3)^2}$$

Generalizing for n-dimensions

$$D = \sqrt{\sum_i (Q_i - P_i)^2}$$

1.4 Vectors

A vector $\underline{v}$ in $\mathbb{R}^n$ is specified by

$$\underline{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

A vector $\vec{QP}$ is the vector between points Q and P. It can also be denoted as $\underline{\mathrm{v}}$

1.4.1 Vector Algebra

Vectors can be scaled and added (only if they have the same number of components.

- Addition: $(\underline{u} + \underline{v})_n = \underline{u_n} + \underline{v_n}$
- Scaling: $(c\underline{u})_n = c\underline{u}_n$

Vectors have the following properties

- Comutativity: $\underline{u} + \underline{v} = \underline{v} + \underline{u}$
- Addition of zero $\underline{u} + \underline{0} = \underline{u}$
- Subtraction $\underline{u} - \underline{v} = \underline{u} + (-\underline{v})$

1.4.2 Norm of a Vector

The norm or "length" of a vector is given by

$$\|\underline{v}\| = \sqrt{\sum_i v_i^2}$$

and has the following properties

- $\|\underline{u} + \underline{v}\| \leq \|\underline{u}\| + \|\underline{v}\|$
- $\|\underline{u}\| = 0 \leftrightarrow \|\underline{u}\| = \underline{0}$
- $c\|\underline{u}\| = |c|\|\underline{u}\|$

1.4.3 Application of vectors in parameterization

Vectors can be used to parameterize lines and planes, where the parameter is a scalar of the tangent vector. For example

- A line can be parameterized as
$$\underline{x}t = \underline{x}_0 + t\underline{v}$$
- A plane can be parameterized as
$$\underline{x}(s,t) = \underline{x}_0 + s\underline{u} + t\underline{v}$$

1.4.4 Basis vectors

A vector $\underline{v}$ in $\mathbb{R}^n$ has n basis vectors
$$\underline{e}_k = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$
where only the k-th term is 1
The vector $\underline{v}$ is given by
$$\underline{v} = \sum_{k=1} n v_k e_k$$

Chapter 2: Week 2

2.1 The Dot Product

The dot product for two vectors $\underline{u}$ and $\underline{v}$ in $\mathbb{R}^n$ separated by angle θ is defined as

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

For two vectors, $\underline{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots u_n \end{pmatrix}$ and $\underline{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \vdots \\ u_n \end{pmatrix}$ their dot product is computed as

$$\underline{u} \cdot \underline{v} = u_1 v_1 + u_2 v_2 + u_3 v_3 + \ldots + u_n v_n$$

Properties of the dot product

The dot product has the following properties

- Commutativity: $\underline{u} \cdot \underline{v} = \underline{v} \cdot \underline{u}$
- $\underline{u} \cdot \underline{0} = 0$
- $\underline{v} \cdot \underline{v} = \|\underline{v}\|^2$

Note that the dot product of two vectors is a *scalar*

2.1.1 Orthogonality and the dot product

For two vectors $\underline{u}$ and $\underline{v} \neq \underline{0}$

$$\underline{u} \perp \underline{v}$$
$$\Updownarrow$$
$$cos\theta = 0$$
$$\Updownarrow$$
$$\underline{u} \cdot \underline{v} = 0$$

2.1.2 Orthogonal Projection

For two vectors $\underline{v}$ and $\underline{u}$, the length of $\underline{v}$ along the $\underline{u}$ axis is

$$\frac{\underline{v} \cdot \underline{u}}{\|\underline{u}\|}$$

2.1.3 Implicit definition of hyperplanes using the dot product

The hyperplane passing through a point $\underline{x}_0$ and orthogonal to vector $\underline{n}$ is

$$(\underline{x} - \underline{x}_0) \cdot \underline{n} = 0$$

2.2 The Cross Product

The cross product of two vectors $\underline{u}$ and $\underline{v}$ is

$$\underline{u} \times \underline{v} = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}$$

2.2.1 Properties of the cross product

The cross product has the following properties

- Anti-commutativity: $\underline{u} \times \underline{v} = -\underline{v} \times \underline{u}$
- Mutual orthagonality: $\underline{u} \times \underline{v} \perp \underline{u}$
- $\underline{u} \times \underline{0} = \underline{0}$
- $\underline{v} \times \underline{v} = -\underline{v} \times \underline{v} = 0$

2.2.2 Geometric understanding of the cross product

For two vectors $\underline{u}$ and $\underline{v}$ separated by angle θ

$$\|\underline{u} \times \underline{v}\| = \|\underline{u}\|\|\underline{v}\| sin\theta$$

2.2.3 Distance to a line and the cross product

The distance from a point P to a line containing point Q and parallel to the vector $\underline{v} is \frac{\|\vec{QP} \times \underline{v}\|}{\|\underline{v}\|}$

2.2.4 The Scalar Triple product

The volume of the parallelopiped generated by vectors $\underline{u}$, $\underline{v}$ and $\underline{w}$ is

$$|\underline{u} \cdot (\underline{v} \times \underline{w})|$$

Note that

$$|\underline{u} \cdot (\underline{v} \times \underline{w})| = |\underline{v} \cdot (\underline{w} \times \underline{u})| = |\underline{w} \cdot (\underline{u} \times \underline{v})| =$$

2.3 Intro to vector calculus

For a curve in 3-D space $\gamma : \mathbb{R} - > \mathbb{R}^3$

Position	Velocity	Acceleration
$\gamma(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$	$\gamma' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$	$\gamma'' = \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix}$
$\underline{r}$	$\underline{v}$	$\underline{a}$

2.3.1 Derivative of a vector

$$\underline{v} = \underline{r}' = \lim_{h \to 0} \frac{\underline{r}(t+h) - \underline{r}(t)}{h}$$

2.3.2 Arclength of a parameterized curve

Given a parameterized curve $\gamma : \mathbb{R} \to \mathbb{R}^n$

$$l = \int dl$$
$$dl = \|\gamma'\| dt$$

Hence

$$l = \int \|\gamma'\| dt$$

2.3.3 Rules of vector differentiation

For vector functions $\underline{u}(t)$ and $\underline{v}(t)$

$$(\underline{u} \cdot \underline{v})' = \underline{u}' \cdot \underline{v} + \underline{u} \cdot \underline{v}'$$
$$(\underline{u} \times \underline{v})' = \underline{u}' \times \underline{v} + \underline{u} \times \underline{v}'$$

2.4 Vectors & Physical Motion

2.4.1 Finding the velocity and position vector with acceleration

For a particle with position vector $\underline{r}(0) = \begin{pmatrix} 2 \\ 4 \\ -1 \end{pmatrix}$ and velocity vector $\underline{v}(0) = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$ and acceleration vector $\underline{a}(t) = \begin{pmatrix} -cos(t) \\ 3e^{-t} \\ 6t \end{pmatrix}$

The velocity vector at time t is

$$\begin{aligned} \underline{v}(t) &= \int_0^t \underline{a}(t)dt + \underline{v}(0) \\ &= \int_0^t \begin{pmatrix} -cost \\ 3e^{-t} \\ 6t \end{pmatrix} + \underline{v}(0) \\ &= \begin{pmatrix} -sint \\ -3e^{-t} \\ 3t^2 \end{pmatrix} dt + \underline{v}(0) \\ &= \begin{pmatrix} -sint - 1 \\ -3e^{-t} - 3 \\ 3t^2 + 2 \end{pmatrix} \end{aligned}$$

The position vector at time t is

$$\begin{aligned} \underline{r}(t) &= \int_0^t \underline{v}(t)dt + \underline{r}(0) \\ &= \int_0^t \begin{pmatrix} -sint - 1 \\ -3e^{-t} - 3 \\ 3t^2 + 2 \end{pmatrix} dt + \underline{r}(0) \\ &= \begin{pmatrix} cost - t - 1 \\ -3e^{-t} + 3t - 3 \\ t^3 + 2t \end{pmatrix} + \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} \end{aligned}$$

2.4.2 Unit tangent and normal vectors

The velocity and acceleration vectors at time t can be expressed as a combinations of two orthogonal vectors.

$\underline{v}(t)$ and $\underline{a}(t)$ can be expressed in terms of

- $\underline{T}$, *unit tangent* vector to $\underline{v}$ at time t
- $\underline{N}$, *unit normal* vector to $\underline{v}$ at time t

where

$$\begin{aligned} \underline{T}(t) &= \frac{\underline{v}(t)}{\|\underline{v}\|} \\ \underline{N}(t) &= \frac{\underline{T}'(t)}{\|\underline{T}'\|} \end{aligned}$$

The acceleration vector a T can be decomposed as

$$
\begin{aligned}
\underline{a} &= a_T \underline{T} + a_N \underline{N} \\
&= \left(\frac{d}{dt}\|\underline{v}\|\right) + \left(\kappa\|\underline{v}\|^2\right)\underline{N}
\end{aligned}
$$

where

$$
\begin{aligned}
\kappa(t) = \|\frac{d}{dl}\underline{T}dl\| &= \frac{\|\frac{d}{dt}\underline{T}(t)\|}{\frac{dl}{dt}} \\
&= \frac{\|\underline{v}\times\underline{a}\|}{\|\underline{v}\|^3} \text{ (in 3-D)}
\end{aligned}
$$

2.4.3 The osculating circle

The osculating circle, or *best fit circle* lies in the normal-tangent plane, and has the reciprocal of the curvature as its radius

The osculating center is given by

$$\underline{r}(t) + \frac{1}{\kappa(t)}\underline{N}(t)$$

where $\kappa(t)$ is the radius

2.4.4 The unit binormal vector

The degree of *twist* out of the osculating plane is given by the unit binormal vector

For a curve $\gamma : \mathbb{R} - > \mathbb{R}^n$, the binormal is

$$\underline{B} = \underline{T} \times \underline{N}$$

The strength of the twisting is given by

$$\tau(t) = -\underline{N}(t) \cdot \frac{d}{dl}\underline{B}(l)$$

Chapter 3: Week 3

3.1 Matrix Equations

3.1.1 Linear System of Equations

A linear system of equations, such as

$$x - 3y + 3z = 8$$
$$5x + y - 2z = 7$$
$$-2x + y - z = -6$$

Can be expressed as

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

Where A is a matrix $A = \begin{bmatrix} 1 & -3 & 3 \\ 5 & 1 & -2 \\ -2 & 1 & -1 \end{bmatrix}$, $\underline{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ and $\underline{b} = \begin{pmatrix} 8 \\ -7 \\ 6 \end{pmatrix}$

The system has a unique solution where

$$\underline{x} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$$

3.2 Gaussian elimination, or row reduction

For a system of linear equations

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 6 & 4 & -2 \\ 0 & 2 & 0 & -1 & 4 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \\ 4 \\ 3 \\ 10 \end{pmatrix}$$

We can convert it into the augmented matrix

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & -2 \\ 0 & 0 & 6 & 4 & -2 & 4 \\ 0 & 2 & 0 & -1 & 4 & 3 \\ 0 & 0 & 0 & 0 & 2 & 10 \end{array}\right]$$

The system is unchanged under the following operations

- Switching 2 rows

$$R_m \leftrightarrow R_n$$

- Recaling a row

$$m \times R_1, m \in \mathbb{R}$$

- Adding a multiple of one row to another

$$R_2 - \frac{1}{2}R_1$$

Given these 3 operations, for each column j in matrix A,

1. Swap R_i with any row R_n provided $n > i$ such that the value A_{ij} is non-zero
2. Scale R_i such that A_{ij} is 1 or -1
3. Clear values in column j below A_{ij} by adding a multiple of R_i
4. Repeat for the next j

3.3 Inverse matrices

The **inverse** of A is a matrix A^{-1} which satisfies

$$AA^{-1} = I = A^{-1}A$$

3.3.1 Inverse 2-by-2 matrices

For a 2-by-2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the inverse is

$$A^{-1} = \frac{1}{ad - cb} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \text{ provided } ab - dc \neq 0$$

3.3.2 Uniqueness of the inverse

Claim: The inverse of a matrix is unique if it exists

If A has multiple inverses B and C, then by definition

$$AB = I = BA, AC = I = CA$$

By associative property,

$$B = IB = (CA)B = C(AB) = CI = C$$

3.3.3 Solving a system of linear equations with the inverse

Given a system $A\underline{x} = \underline{b}$, applying the inverse to both sides gives

$$A^{-1}A\underline{x} = \underline{x} = A^{-1}\underline{b}$$

3.3.4 Finding the inverse via Gaussian Elimination or row reduction

Given $A = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 3 & -1 \\ -4 & 0 & 2 \end{bmatrix}$,the inverse A^{-1} can be found via row reduction,

$$\left[\begin{array}{ccc|ccc} 1 & 3 & 1 & 1 & 0 & 0 \\ 2 & 3 & -1 & 0 & 1 & 0 \\ 4 & 0 & 2 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & 1 & 0 & 0 & \frac{1}{3} & \frac{1}{6} \\ 0 & 0 & 1 & \frac{2}{3} & -\frac{2}{3} & -\frac{1}{6} \end{array}\right]$$

3.3.5 Finding the inverse of a block diagonal matrix

For a matrix

$$A = \begin{bmatrix} A_1 & \dots & \dots \\ \vdots & A_2 & \vdots \\ \vdots & \dots & A_3 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} A_1^{-1} & \dots & \dots \\ \vdots & A_2^{-1} & \vdots \\ \vdots & \dots & A_3^{-1} \end{bmatrix}$$

Chapter 4: Week 4

4.1 Linear Transformations

A matrix is a **linear transformation** sending vectors to vectors

For a $m \times n$ matrix A, $A : \mathbb{R}^n \to \mathbb{R}^m$

$$\underline{u} = A\underline{x}$$

A is a **linear** function, meaning

$$A(c\underline{x}) = cA\underline{x}$$

$$A(\underline{x} + \underline{y}) = A\underline{x} + A\underline{y}$$

4.1.1 Diagonal matrices as rescaling function

A matrix A of the form

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

Rescales each axis independently by factor a and b, if a or b is negative, the respective axis flipped
A special case of the rescaling function is a **projection**, such as

$$\begin{bmatrix} 0 & 0 \\ 0 & a \end{bmatrix}$$

Note: some matrices can project along non-axis aligned directions, such as

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

4.1.2 Shearing matrices

A vector A of the form

$$\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$$

shears the plane along the horizontal

4.1.3 Rotation matrices

A matrix of the form

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

rotates the plane counterclockwise by angle θ
Note that for **special angles**, such as $\frac{\pi}{2}$, the rotation matrix is

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

4.1.4 Matrix composition

Performing the transformation described by matrix A, followed by the transformation described by matrix B gives a composition C

$$C = BA$$

In general, any linear transformation

$$\mathbb{R}^n \to \mathbb{R}^n$$

Is a composition of scaling, shearing, rotation, and projection

4.1.5 Example in Euler Angles, 3D, and 4D matrices

Euler Angles & 3D rotation matrices
The *roll*, *pitch* and *yaw* describes rotations in the x, y, z planes respectively, by angles γ, β, α respectively, giving

$$Roll = R_{x:\gamma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\gamma & -sin\gamma \\ 1 & sin\gamma & cos\gamma \end{bmatrix}$$

$$Pitch = R_{y:\beta} = \begin{bmatrix} cos\beta & 0 & sin\beta \\ 0 & 1 & 0 \\ -sin\beta & 0 & cos\beta \end{bmatrix}$$

$$Yaw = R_{z\alpha} = \begin{bmatrix} cos\alpha & -sin\alpha & 0 \\ sin\alpha & cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For a 4D matrix

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

Rotates the $x_1 - x_2$ plane by α, and shears the $x_3 - x_4$ plane

4.2 Change in Basis

4.2.1 Basis vectors

A basis for $\mathbb{R}^n$ has exactly n independent vectors

Given a basis $B = (u_k)_{k=1...n}$, any vector can be uniquely expressed as

$$\underline{v} = c_1\underline{u}_1 + c_2\underline{u}_2 + \dots c_n\underline{u}_n$$

The scalars c_i are the **coordinates** of $\underline{v}$ in B

A basis is

- **Orthogonal**, if the basis vectors are mutually orthogonal
- **Orthonormal**, if the basis vectors are mutually orthogonal **and** basis vectors are all of **unit length**

For an orthonomal basis B, the coordinates c_i can be computed as

$$c_i = \underline{v} \cdot \underline{u}_i, i \in n$$

4.2.2 Coordinate changes

For a vector $\underline{v}$, expressed in a new coordinate system described by A, the coordinates $\underline{v}$ in the new coordinate system $\underline{w}$is

$$\underline{w} = A^{-1}\underline{v}$$

$$\underline{v} = A\underline{w}$$

4.3 The Determinant

4.3.1 Computing the determinant

There are a few strategies for computing the determinant

1. By formula (for cases in $2d$ or $3d$
2. By minor expansion
3. By row reduction (as an extension of minor expansion)
4. By Block Diagonal matrices

4.3.2 By formula

In 1-by-1,

$$|[a]| = a$$

In 2-by-2,

$$\left|\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right| = ad - bc$$

In 3-by-3, for matrix A given by

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$|A| = (a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}) - (a_{13}a_{22}a_{31} + a_{12}a_{21}a_{31} + a_{11}a_{23}a_{32})$$

4.3.3 By Minor Expansion

For a matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$|A| = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Following the "sign convention", where

- odd columns / rows: $+$
- even columns / rows: $-$

4.3.4 By Row Operations

Since triangle matrices can have their determinants easily computed by taking the product of the diagonal, we can perform row operations on the matrix to convert it to a diagonal matrix

Row operation	Matrix multiplication of A by R	Determinant of R	Effect on determinant
Swapping of rows 1 and 2	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$Det(R) = -1$	$Det(RA) = -Det(A)$
Scaling a row by factor of c	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$Det(R) = c$	$Det(RA) = cDet(A)$
Linear combination of two rows	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & c & 0 & 1 \end{bmatrix}$	$Det(R) = 1$	No change

4.3.5 By Block Diagonal Matrices

For a matrix

$$A = \begin{bmatrix} A_1 & \dots & \dots \\ \vdots & A_2 & \vdots \\ \vdots & \dots & A_3 \end{bmatrix}$$

$$|A| = |A_1|\,|A_2|\,|A_3|$$

4.4 Geometry of determinants

In 2D the determinant is related to the area of parallelogram, where

$$\text{Area of parallelogram formed by } \begin{pmatrix} c \\ d \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix} = \left| DET \begin{bmatrix} a & c \\ b & d \end{bmatrix} \right|$$

In 3D, the determinant is the *oriented volume* of the parallelopiped, where

$$\text{Oriented volume of parallelopiped formed by } \underline{u}, \underline{v}, \underline{w} = \left| DET \begin{bmatrix} \vert & \vert & \vert \\ \underline{u} & \underline{v} & \underline{w} \\ \vert & \vert & \vert \end{bmatrix} \right|$$

In the general, the determinant describes how n - dimensional volume changes under a linear transformation $A : \mathbb{R}^n - > \mathbb{R}^n$

In arbitrary dimensions, the linear transformation described by A sends the unit **hypercube** spanned by the unit basis vectors to a **parllelopiped** with **"n-volume"** $|DET(A)|$

4.4.1 Properties of the determinant

$$Det(A^T) = Det(A)$$

$$Det(AB) = DET(A)DET(B)$$

$$A \text{ is invertible if and only if } DetA \neq 0$$

$$\text{If } A \text{ is invertible then } Det(A)^{-1} = \frac{1}{DetA}$$

Chapter 5: Week 5

5.1 Multivariate functions

Some examples of multivariate functions covered so far include

- Parameterized curves:
$$f : \mathbb{R} \to \mathbb{R}^n$$
- Parameterized surfaces
$$f : \mathbb{R}^2 \to \mathbb{R}^n$$
- Implicit curves
$$f : \mathbb{R}^2 \to \mathbb{R}$$
- Implicit surfaces
$$f : \mathbb{R}^3 \to \mathbb{R}$$

5.1.1 Multivariate functions as non-linear transforms

Multivariate functions can be used to describe non-linear transformations, just as matrices describe linear transformations.

For example, translating between polar and euclidean coordinates

$$\begin{pmatrix} x \\ y \end{pmatrix} = f \begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} rcos\theta \\ rsin\theta \end{pmatrix}$$

$$\begin{pmatrix} r \\ \theta \end{pmatrix} = f^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sqrt{x^2 + y^2} \\ ARCTAN(\frac{y}{x}) \end{pmatrix}$$

5.1.2 Multivariate functions example - Market Equilibrium

Suppose a market has

- Supply:
$$S(P) = S_0 + aP$$
- Demand:
$$D(P) = D_0 - bP$$

Then the market has some equilibrium price and quantity P_m and Q_m respectively, where

$$Q_m = S_0 + aP_m = D_0 - bP_m$$

Equating demand and supply, and making P_m the subject of the equation, we get

$$P_m = \frac{D_0 - S_0}{a + b}$$

and

$$Q_m \frac{aD_0 + bS_0}{a + b}$$

Hence, we see that P_m and Q_m are two outputs of a function of four inputs D_0, S_0, a, b, or

$$\begin{pmatrix} P_m \\ Q_m \end{pmatrix} = f \begin{pmatrix} D_0 \\ S_0 \\ a \\ b \end{pmatrix}$$

5.2 Partial Derivatives

The **partial derivative** of a function $f(x)$ with respect to the jth input x_j is

$$\frac{\partial f}{\partial x_j} = \lim_{h \to 0} \frac{f(\underline{x} + h\underline{e}_j) - f(x)}{h}$$

5.2.1 Computing the partial derivative and implicit differentiation

For a function

$$f \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x^3 y - 5xyz^2$$

We can compute the following partial derivatives

$$\tfrac{\partial f}{\partial x} = 3x^2 y - 5y^2 \;\Big|\; \tfrac{\partial f}{\partial y} = x^3 - 5xz^2 \;\Big|\; \tfrac{\partial f}{\partial z} = -10xyz$$

Recall that for for implicit differentiation,

$$df = (3x^2 y - 5y^2)dx + (x^3 - 5xz^2)dy + (-10xyz)dz$$

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial y}{\partial y} dy + \frac{\partial f}{\partial z} dz$$

5.2.2 Partial derivatives and rates of change

For a function that takes 2 inputs x, y and returns 3 outputs u, v, w,

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x^4 \\ y^2 - xy \\ 5x - 2y \end{pmatrix}$$

We can compute the following partial derivatives

$$\begin{array}{c|c} \tfrac{\partial u}{\partial x} = -4x^3 & \tfrac{\partial u}{\partial y} = 0 \\ \tfrac{\partial v}{\partial x} = -y & \tfrac{\partial v}{\partial y} = 2y - x \\ \tfrac{\partial w}{\partial x} = 5 & \tfrac{\partial w}{\partial y} = -2 \end{array}$$

5.3 The Derivative

The derivative of a function

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

can be written as a $m \times n$ matrix where the partial derivative of the jth output to the $i - th$ input makes up the $j - th$ row and $i - th$ column element

$$[Df] = \left[\frac{\partial f_i}{\partial x_j} \right]_{ij}$$

The derivative transforms vectors of rates of change of inputs to vectors of rates of change of outputs

5.3.1 A note on notation

For a function f that maps input $\underline{x}$ to $\underline{y}$, we can denote

- rate of change of input: $\underline{h} = (h_j)$ or $\dot{\underline{x}} = (\dot{x}_j)$
- rate of change of output: $\underline{l} = (l_j)$ or $\dot{\underline{y}} = (\dot{y}_j)$
- derivative as $[Df]$ or $\frac{\partial \underline{y}}{\partial \underline{x}}$

Hence the relationship between the input and output rates of change can be expressed as

$$\underline{l} = [Df]\underline{h}$$

Or

$$\dot{\underline{y}} = \frac{\partial \underline{y}}{\partial \underline{x}} \dot{\underline{x}}$$

5.3.2 Computing inputs and outputs rates of change

For function

$$\begin{pmatrix} u \\ v \end{pmatrix} = f \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} xy^2 - x^2 z \\ 3xy + z^3 \end{pmatrix}$$

With inputs changing at

$$\underline{h} = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}$$

At a point

$$\underline{a} = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$$

We can find the derivative by

$$[Df] = \begin{bmatrix} y^2 - 2xz & 2xy & -x^2 \\ 3y & 3x & 3z^2 \end{bmatrix}$$

Evaluating the derivative at $\underline{a}$, we get

$$[Df]_{\underline{a}} = \begin{bmatrix} -11 & -4 & -4 \\ -3 & 6 & 27 \end{bmatrix}$$

We can find the rate of change of outputs by applying the linear transformation $[Df]_{\underline{a}}$ to $\underline{h}$

$$\begin{aligned} \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} &= [Df]_{\underline{a}} \underline{h} \\ &= \begin{bmatrix} -11 & -4 & -4 \\ -3 & 6 & 27 \end{bmatrix} \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 13 \\ 5 \end{pmatrix} \end{aligned}$$

5.3.3 Definition of the derivative

Definition: A function $f(\underline{x})$ is differentiable at $\underline{x} = \underline{a}$ if and only if there is a **linear transformation** $[Df]_{\underline{a}}$, the derivative of f at $\underline{a}$ such that

$$\lim_{\|\underline{h}\| \to 0} \frac{(f(\underline{a} + \underline{h}) - f(\underline{a})) - [Df]_{\underline{a}}\underline{h}}{|\underline{h}|} = \underline{0}$$

5.3.4 A Taylor Perspective

Recall from single variable calculus that for a function f, at very small values of h,

$$f(x+h) = f(x) + f'(x)h + \dots$$

In the multivariate case

$$f(\underline{x} + \underline{h}) = f(\underline{x}) + [Df]\underline{h} + \dots$$

Where

- $f(\underline{x})$ is the $0 - th$ term
- $[Df]\underline{h}$ is the linear term

This **is** the definition of the derivative!! $\rightarrow$ it is the coefficient of the linear term in the Taylor series

A weird example

For a function

$$S(A) = A^2$$

Where A is a 2 by 2 matrix, and the function S has 4 inputs and 4 outputs

$$S(A+H) = (A+H)^2 = A^2 + AH + HA + H^2$$

$$\begin{aligned} S(A+H) &= S(A) + [DS]_A H \quad + \dots \\ &= (A^2) + (AH + HA) \quad + H^2 \end{aligned}$$

Hence,

$$[DS]_A H = AH + HA$$

5.3.5 Existence of the derivative

The derivative does not exist, even if all the partial derivatives exist, if

$$\lim_{\|\underline{h}\| \to 0} \frac{(f(\underline{a} + \underline{h}) - f(\underline{a})) - [Df]_{\underline{a}}\underline{h}}{|\underline{h}|} \neq \underline{0}$$

Chapter 6: Week 6

6.1 The Chain Rule

6.1.1 The classical chain rule

The classical chain rule states that for functions f and $g : \mathbb{R} \to \mathbb{R}$, the composition of $f \circ g$ has derivative

$$(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$$

Alternatively, for $y = y(u)$ and $u = u(x)$,

$$\left.\frac{dy}{dx}\right|_a = \left.\frac{dy}{du}\right|_{u(a)} \cdot \left.\frac{du}{dx}\right|_a$$

Take note of where each derivative is evaluated at

- f' is evaluated at $g(a)$
- g' is evaluated at a

6.1.2 The multivariate chain rule

The multivariate chain rule states that for differentiable functions $f : \mathbb{R}^m \to \mathbb{R}^p$ and $g : \mathbb{R}^n \to \mathbb{R}^m$, the composition $f \cdot g : \mathbb{R}^n \to \mathbb{R}^p$ has derivative

$$[D(f \cdot g)]_{\underline{a}} = [Df]_{g(\underline{a})} \cdot [Dg]_{\underline{a}}$$

Or, for $\underline{y} = \underline{y}(u)$ and $\underline{u} = \underline{u}(x)$

$$\frac{\partial \underline{y}}{\partial \underline{x}} = \frac{\partial \underline{y}}{\partial \underline{u}} \cdot \frac{\partial \underline{u}}{\partial \underline{x}}$$

6.1.3 Note on notation

Most difficulties with chain rule problems arise due to difficulties with notation. Be careful with variables!

Problem: For example, given a system

$$z = uv^3$$

$$u = s^2 - t^2$$

$$v = st$$

Compute the partial derivatives of z with respect to s, t
Solution: We let

$$z = z(\underline{y})$$

$$\underline{y} = \underline{y}(\underline{x})$$

Where

$$\underline{y} = \begin{pmatrix} u \\ v \end{pmatrix}, \underline{x} = \begin{pmatrix} s \\ t \end{pmatrix}$$

We can compute

$$\begin{aligned} \frac{\partial z}{\partial \underline{y}} &= \begin{bmatrix} \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{bmatrix} \\ &= \begin{bmatrix} v^3 & 3uv^2 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \underline{y}}{\partial \underline{x}} &= \begin{bmatrix} \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial s} & \frac{\partial v}{\partial t} \end{bmatrix} \\ &= \begin{bmatrix} 2s & -2t \\ t & s \end{bmatrix} \end{aligned}$$

Applying the chain rule,

$$\begin{aligned} \frac{\partial z}{\partial \underline{x}} &= \frac{\partial z}{\partial \underline{y}} \cdot \frac{\partial \underline{y}}{\partial \underline{x}} \\ &= \begin{bmatrix} v^3 & 3uv^2 \end{bmatrix} \begin{bmatrix} 2s & -2t \\ t & s \end{bmatrix} \\ &= \begin{bmatrix} 2sv^3 + 3tuv^2 & -2tv^3 + 3suv^2 \end{bmatrix} \\ &= \begin{bmatrix} 5s^4t^3 - 3s^2t^5 & -5s^3t^4 + 3s^5t^2 \end{bmatrix} \end{aligned}$$

6.1.4 Classical chain rules as a variation on the ONE single chain rule

Classical calculus textbooks state that there are different chain rules, all of which need to be memorized. For example
1. For a function of the form $y = y(x_1, x_2, \ldots, x_n)$, with each $x_j = x(t)$,

$$\frac{dy}{dt} = \frac{\partial y}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial y}{\partial x_2}\frac{dx_2}{dt} + \ldots + \frac{\partial y}{\partial x_n}\frac{dx_n}{dt}$$

2. For a function $u(x, y)$ with $x = x(s, t)$ and $y = y(s, t)$,

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s}$$

However, recognize that all these *special* derivatives can be derived using the one chain rule.

For example
Given: a function $u(x, y)$ with $x = x(s, t)$ and $y = y(s, t)$,
Let: function $g : g\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$ and $f : f\begin{pmatrix} x \\ y \end{pmatrix} = u$
Compute:

$$\begin{aligned} u = f(x, y) \quad & [Df] = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{bmatrix} \\ \begin{pmatrix} x \\ y \end{pmatrix} = g(s, t) \quad & [Dg] = \begin{bmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{bmatrix} \end{aligned}$$

Applying the chain rule

$$\begin{aligned} [D(f \cdot g)] &= \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} & \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t} \end{bmatrix} \end{aligned}$$

6.2 Differentiation rules: Linearity of the derivative

Differentiation is a **linear operator** on functions

- Addition: $[D(f+g)] = [Df] + [Dg]$
- Scalar multiplication: $[D(cf)] = c\,[Df]$

For a linear or **affine** function of the form

$$\underline{y} = A\underline{x} + \underline{b}$$

$$\frac{\partial \underline{y}}{\partial \underline{x}} = A$$

Conclusion: Affine functions have constant derivatives

For the **identity function**

$$\underline{y} = \underline{x}$$

$$\frac{\partial \underline{y}}{\partial \underline{x}} = I = \frac{\partial \underline{x}}{\partial \underline{x}}$$

Conclusion: the identity function has the identity matrix as its derivative

6.3 Differentiation rules: Product Rule

For $\underline{u}, \underline{v}$ dependent on a variable t,

$$(\underline{u} \cdot \underline{v})' = \underline{u}' \cdot \underline{v} + \underline{u} \cdot \underline{v}'$$

To prove the rule mentioned above,

Claim:

$$(\underline{u} \cdot \underline{v})' = \underline{u}' \cdot \underline{v} + \underline{u} \cdot \underline{v}'$$

Let:

$$g(t) = \begin{pmatrix} \underline{u}(t) \\ \underline{v}(t) \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ \underline{u}_n \\ v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix}$$

and

$$f(\underline{u}, \underline{v}) = \underline{u} \cdot \underline{v} = \underline{u}^T \underline{v} = \sum_{j=1\dots n} \underline{u}_j \underline{v}_j$$

Note: the various notations that represent the same dot product
Compute

$$[Dg] = \begin{bmatrix} \underline{u}' \\ \underline{v}' \end{bmatrix} = \begin{bmatrix} u_1' \\ \vdots \\ u_n' v_1' \\ \vdots \\ v_n' \end{bmatrix}$$

$$[Df] = \begin{bmatrix} \underline{v}^T & \underline{u}^T \end{bmatrix} = \begin{bmatrix} v_1 & \dots & v_n & u_1 & \dots & u_n \end{bmatrix}$$

Applying the chain rule

$$\begin{aligned}(\underline{u} \cdot \underline{v})' &= [D(f \cdot g)] \\ &= [Df]\,[Dg] \\ &= \begin{bmatrix} \underline{v}^T & \underline{u}^T \end{bmatrix} \begin{bmatrix} \underline{u}' \\ \underline{v}' \end{bmatrix} \\ &= \underline{v}^T \underline{u}' + \underline{u}^T \underline{v}' \\ &= \underline{u}' \cdot \underline{v} + \underline{v}' \cdot \underline{u}\end{aligned}$$

Another example with quadratic functions
For a square matrix A with variable $\underline{x}$, consider the quadratic function

$$f(\underline{x}) = \underline{x}^T A \underline{x} = \underline{x} \cdot (A\underline{x})$$

Claim:

$$[Df] = \underline{x}^T (A + A^T)$$

Proof

$$\begin{aligned}[Df] &= \left[D(\underline{x}^T A \underline{x})\right], \quad \text{recall that } \left[D(\underline{u}^T \underline{v})\right] = \underline{u}^T\,[Dv] + \underline{v}^T\,[D\underline{u}] \\ &= \underline{x}^T\,[D(A\underline{x})] + (A\underline{x})^T\,[D\underline{x}], \quad \text{recall that } [DA\underline{x}] = A, [D\underline{x}] = I \\ &= \underline{x}^T A + (A\underline{x})^T I \quad \text{recall that} (AB)^T = B^T A^T \\ &= \underline{x}^T A + \underline{x}^T A^T \\ &= \underline{x}^T (A + A^T)\end{aligned}$$

6.3.1 The Material Derivative

For a function $f(t, \underline{x})$ with $\underline{x} = \underline{x}(t)$, the material derivative is defined as

$$\frac{Df}{Dt} = \frac{d}{dt} f(t, \underline{x}(t)) = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \underline{x}} \frac{dx}{dt}$$

Claim: the material derivative can be proved using the chain rule
Proof:

$$\frac{Df}{Dt} = \frac{d}{dt} f(t, \underline{x}(t)) = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \underline{x}} \frac{dx}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \underline{x}} \underline{v}$$

Let:

$$g(t) = \begin{pmatrix} t \\ \underline{x}(t) \end{pmatrix}$$

$$f = f(t, \underline{x})$$

We can compute the derivatives

$$[Dg] = \begin{bmatrix} 1 \\ \frac{dx}{dt} \end{bmatrix}$$

$$[Df] = \begin{bmatrix} \frac{\partial f}{\partial t} & \frac{\partial f}{\partial \underline{x}} \end{bmatrix} \begin{bmatrix} 1 \\ \frac{dx}{dt} \end{bmatrix}$$

6.4 The Inverse Function Theorem

Recall from single variable calculus that if a function $f : \mathbb{R} \to \mathbb{R}$ has an inverse f^{-1} then

$$\frac{d}{dx}(f^{-1}) = \frac{1}{df/dx}$$

6.4.1 Inverse multivariate functions

A function $f : \mathbb{R}^n \to \mathbb{R}^n$ has an inverse f^{-1} if and only if

$$f^{-1}(f(\underline{x})) = \underline{x}$$

for all $\underline{x}$

The existence of the inverse should not be taken for granted

6.4.2 Linearization of nonlinear functions

Given a function $f(\underline{x}) = \underline{u}$, we can linearize it locally via Taylor Exapnsion

For example **Given**

$$\underline{u} = f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e^{x+y} - cos(3x+y) \\ xy + sin(x-2y) \end{pmatrix}$$

For x, y close to zero, we can take the Taylor expansion

$$\begin{aligned} f\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 1 + (x+y) + O(\|\underline{x}\|^2) - (1 - O\left(\|\underline{x}\|^2\right)) \\ xy + ((x-2y) - O\left(\|\underline{x}\|^3\right)) \end{pmatrix} \\ &= \begin{pmatrix} x + y + O\left(\|\underline{x}\|^2\right) \\ x - 2y + O\left(\|\underline{x}\|^2\right) \end{pmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + O\left(\|\underline{x}\|^2\right) \end{aligned}$$

6.4.3 The Inverse Rule

For a function $f : \mathbb{R}^n \to \mathbb{R}^n$ that is invertible

$$\left[Df^{-1}\right] = [Df]^{-1}$$

Proof: Since f is invertible,

$$(f^{-1} \cdot f)\underline{x} = \underline{x}$$

Since the identity function has the identity matrix as its derivative

$$\left[D\left(f^{-1} \cdot f\right)\right] = I$$

By the chain rule

$$\left[Df^{-1}\right][Df] = I$$

Hence

$$\left[Df^{-1}\right] = [Df]^{-1}$$

6.4.4 The Inverse Function Theorem

A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is locally invertible near $f(\underline{a})$ if the derivative of f at $\underline{a}$ is invertible, i.e.

f is invertible near $f(\underline{a})$ if $Det\,[Df]_{\underline{a}} \neq 0$
Linear data controls the existence of **nonlinear** inverse functions locally

6.5 The Implicit Function theorem

6.5.1 Implicit differentiation and implicit functions

Recall in the classical case, for an implicit curve given by

$$F(x, y) = 0$$

By implicit differentiation,

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy = 0$$

We can solve for the slope

$$\frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}$$

Hence, we can solve for $y = y(x)$ (expressing y in terms of x, as long as $\frac{\partial F}{\partial y}$ is well defined

For example, for a circle in the plane,

$$F(x, y) = x^2 + y^2 - r^2 = 0$$

$$\frac{dy}{dx} = -\frac{2x}{2y} = -\frac{x}{y}$$

We can express y in terms of x except for points where $y = 0$

$$y = \sqrt{r^2 - x^2}, y > 0$$

$$y = -\sqrt{r^2 - x^2}, y < 0$$

For a sphere in 3d

$$F(x, y, z) = x^2 + y^2 + z^2 - r^2 = 0$$

z can be expressed in x, y as

$$z = \sqrt{r^2 - x^2 - y^2}, z > 0$$

$$z = -\sqrt{r^2 - x^2 - y^2}, z > 0$$

We can **locally solve** for $z = z(x, y)$ as long as $z \neq 0$, which coincides with $\frac{\partial}{\partial z}(x^2 + y^2 + z^2 - r^2) = 0$

6.5.2 The Implicit Function Theorem

Given a collection $F(\underline{x}, \underline{y}) = \underline{0}$ of m equations defined in terms of $\underline{x}$ (n variables) and $\underline{y}$ (m variables)

Solutions to $F(\underline{x}, \underline{y}) = \underline{0}$ near a solution point $\underline{a}$ can be realised as an implicit function

$$\underline{y} = \underline{y}(\underline{x})$$

if and only if

$$Det \left[\frac{\partial F}{\partial \underline{y}}\right]_{\underline{a}} \neq 0$$

This local solution is unique and differentiable, with

$$\left[\frac{\partial \underline{y}}{\partial \underline{x}}\right]_{\underline{a}} = -\left[\frac{\partial F}{\partial \underline{y}}\right]_{\underline{a}}^{-1} \left[\frac{\partial F}{\partial \underline{x}}\right]_{\underline{a}}$$

Chapter 7: Week 7

7.1 Gradient

7.1.1 The level set

For functions of the form

$$f : \mathbb{R}^m \to R$$

We can consider its level sets, which are subsets of the domain given by the implicit equations

$$f(\underline{x}) = c$$

for some constant c
We denote level sets with

$$f^{-1}(c) = \underline{x} : f(\underline{x}) = c$$

Note this does not imply f is invertible

7.1.2 Definition of the gradient

For scalar-value functions, there is a more primitive approach to the derivative

For $f : \mathbb{R}^n \to \mathbb{R}$, the **gradient** is the **vectorized derivative**

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Note that:

- ∇f is a **vector** at $\underline{a}$
- $[Df]_{\underline{a}}$ is a **linear transformation** acting on vectors at $\underline{a}$

7.1.3 Gradients and level sets

Lemma: At each point, the gradient is **orthogonal** to that point's level set and is oriented in the direction of **maximally increasing value**

Proof
For a point $\underline{a}$ where $f(\underline{a}) = c$, let $\underline{u}$ be a unit vector at $\underline{a}$

Let the rate of change of f along $\underline{u}$ be denoted by $D_{\underline{u}}f$.

Recall that the rate of change of outputs can be given by

$$D_{\underline{u}}f = [Df]_{\underline{a}}\,\underline{u} = \nabla f \cdot \underline{u}$$

Since $[Df]\,\underline{u}$ is equivalent to $\nabla f \cdot \underline{u}$ and $\nabla f \cdot u$ is the length of projection of ∇f onto $\underline{u}$, we can conclude that

- when $\underline{u}$ is tangent to $f^{-1}(c)$, f is not changing, hence
$$\nabla f \perp f^{-1}(c)$$
- when $\underline{u}$ points along ∇f, the rate of change is maximized, hence gradient points in the direction of maximal increasing value

7.1.4 Directional Derivatives

There is only one derivative, but another set of terminology exists

For a real value function, to compute the derivative along a direction given by the unit vector $\underline{u}$, the **directional derivative** is defined as

$$D_{\underline{u}}f = \nabla f \cdot \underline{u} = [Df]\,\underline{u}$$

7.2 Tangent Spaces

7.2.1 The tangent space of an implicit surface

For an implicit surface in $3d$ given by $F(x, y, z) = 0$, we can find the tangent plane to a point $\underline{x}_0$ by using the gradient

$$\nabla F \cdot (\underline{x} - \underline{x}_0) = 0$$

7.2.2 The tangent space of an parameterized curve

For a parameterized curve

$$f(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$$

The velocity is

$$[Df] = \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \end{pmatrix}$$

The tangent line at t_0 is parameterized by another variable, s. A point on the tangent line can be expressed as

$$\underline{u}(s) = f(t_0) + [Df]_{t_0}\, s$$

7.2.3 The tangent space of a parameterized surface

Generalizing to a case in $3d$, for a parameterized surface of the form

$$f : \mathbb{R}^2 \to \mathbb{R}^3$$

Given a parameterization

$$\underline{x} = f(\underline{t})$$

where

$$\underline{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \underline{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

The tangent plane at $f(t_0)$ is given by the same formula

$$\underline{x} = f(\underline{t}_0) + [Df]\,\underline{s}$$

where there are two parameters s_1 and s_2 in $\underline{s}$

$$s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

The two columns in $[Df]$ form the basis vectors of the tangent plane, a point in which is specified by two parameters s_1 and s_2
Parametric to implicit!!: to find the implicit equation of the plane, recall that a plane can be implicitly defined using the normal to the plane, which can be found by taking the cross product of the two columns in $[Df]$

7.2.4 Images, Kernels, and tangent spaces in higher dimensions

Bonus content, will fill in later

7.3 Linearization

Finding tangent spaces is essentially a linearization of a function about a certain point, which provides a locally valid approximation

7.3.1 Differentials vs derivatives

Differentials and derivatives provide two different notations for the same idea

For a function

$$f(\underline{x}) = f(x_1, x_2, \ldots x_n)$$

Differentials
df is a linear combination of differentials

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$$

Derivatives

$$[Df] = \frac{\partial f}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

$[Df]$ is a linear transformation that acts on vectors of rates of change of inputs

if $d\underline{x}$ is a vector of differentials x_i, then the rate of change of output is

$$\frac{\partial f}{\partial \underline{x}} d\underline{x} = df$$

7.3.2 Relative rates of change and percentage errors

For a quantity u, the relative rate of change of u is given by

$$d(lnu) = \frac{du}{u}$$

The rate of change of output is a linear combination of rate of change of inputs

Given:

$$u(\underline{x}) = u(x_1, x_2, \ldots, x_n)$$

Assume that u is of the form

$$u = K\Pi_{i=1}^{n} x_i^{c_i}$$

Then

$$\begin{aligned}\frac{du}{u} &= d(lnu)\\ &= d(lnk + \sum_{i=1}^{n} ln(x_i^{c_i}))\\ &= \sum_{i=1}^{n} c_i d(ln(x_i))\\ &= \sum_{i=1}^{n} c_i \frac{dx_i}{x_i}\end{aligned}$$

7.4 Taylor Series

7.4.1 Single variable case

Recall that for a function

$$f : \mathbb{R} \to \mathbb{R}$$

The Taylor series is

$$\begin{aligned}f(x) &= \sum_{i=0}^{\infty} \frac{1}{i!} \left.\frac{d^i f}{dx^i}\right|_0 x^i\\ f(x) &= \sum_{i=0}^{\infty} \frac{1}{i!} \left. D^i f\right|_0 x^i\\ f(a+h) &= \sum_{i=0}^{\infty} \frac{1}{i!} \left. D^i f\right|_a h^i\\ f(x) &= \sum_{i=0}^{\infty} \frac{1}{i!} \left. D^i f\right|_a (x-a)^i\end{aligned}$$

7.4.2 Multivariable case

For a function

$$f : \mathbb{R}^n \to \mathbb{R}$$

The Taylor Series is given by

$$f(\underline{x}) = \sum_{I} \frac{1}{I!} \left. D^I f\right|_{\underline{0}} \underline{x}^I$$

$$f(\underline{x}) = \sum_{I} \frac{1}{I!} \left. D^I f\right|_{\underline{a}} (\underline{x} - \underline{a})^I$$

$$f(\underline{a} + \underline{h}) = \sum_{I} \frac{1}{I!} \left. D^I f\right|_{\underline{a}} (\underline{h})^I$$

The multi-index
Given n variables
$$\underline{x} = (x_1, x_2, x_3, \dots x_n)$$
The multi-index is n ordered indices
$$I = (i_1, i_2, i_3, \dots, i_n)$$
The degree of a multi-index Where the degree is given by
$$|I| = i_1 + i_2 + \dots + i_n$$
Taking a vector to the power of the multi-index: given $\underline{x}$ and I,
$$\underline{x}^I = x_1^{i_1} x_2^{i_2} \dots x_n^{i_n}$$
Taking the factorial of a multi-index For a multi-index I,
$$I = (i_1, i_2, \dots, i_n)$$
The factorial is defined as
$$I! = i_1! i_2! \dots i_n!$$
Multi-index and the derivative
Given
$$f = f(\underline{x}) = f(x_1, x_2, \dots, x_n)$$
We define
$$D^I f = \frac{\partial^{i_1}}{\partial x_1^{i_1}} \frac{\partial^{i_2}}{\partial x_2^{i_2}} \dots \frac{\partial^{i_n}}{\partial x_n^{i_n}} f$$

7.5 Computing Taylor Series

7.5.1 Commonly used Taylor Series

For all x
$$\begin{aligned}
e^x &= 1 + x + \tfrac{x^2}{2!} + \tfrac{x^2}{3!} + \dots &&= \sum_{n=0}^{\infty} \tfrac{x^n}{n!} \\
sin\ x &= x - \tfrac{x^3}{3!} + \tfrac{x^5}{5!} - \tfrac{x^7}{7!} + \dots &&= \sum_{n=0}^{\infty} \tfrac{(-1)^n x^{2n+1}}{(2n+1)!} \\
cos\ x &= 1 - \tfrac{x^2}{2!} + \tfrac{x^4}{4!} - \tfrac{x^6}{6!} + \dots &&= \sum_{n=0}^{\infty} \tfrac{(-1)^n x^{2n}}{(2n)!} \\
sinh\ x &= x + \tfrac{x^3}{3!} + \tfrac{x^5}{5!} + \tfrac{x^7}{7!} + \dots &&= \sum_{n=0}^{\infty} \tfrac{x^{2n+1}}{(2n+1)!} \\
cosh\ x &= 1 + \tfrac{x^2}{2!} + \tfrac{x^4}{4!} + \tfrac{x^6}{6!} + \dots &&= \sum_{n=0}^{\infty} \tfrac{x^{2n}}{(2n)!}
\end{aligned}$$
For x in restricted interval
$$\begin{aligned}
\tfrac{1}{1-x} &= 1 + x + x^2 + x^3 + \dots &&= \sum_{n=0}^{\infty} x^n &&\text{for } x \in (-1,1) \\
\tfrac{1}{(1-x)^2} &= 1 + 2x + 3x^2 + 4x^3 + \dots &&= \sum_{n=1}^{\infty} n x^{n-1} &&\text{for } x \in (-1,1) \\
ln(1+x) &= x - \tfrac{x^2}{2} + \tfrac{x^3}{3} - \tfrac{x^4}{4} + \dots &&= \sum_{n=0}^{\infty} \tfrac{(-1)^n x^{n+1}}{n+1} &&\text{for } x \in (-1,1] \\
arctan\ x &= x - \tfrac{x^3}{3} + \tfrac{x^5}{5} - \tfrac{x^7}{7} + \dots &&= \sum_{n=0}^{\infty} \tfrac{(-1)^n x^{2n+1}}{2n+1} &&\text{for } x \in [-1,1] \\
(1+x)^k &= 1 + kx + \tfrac{k(k-1)}{2!} x^2 + \tfrac{k(k-1)(k-2)(k-3)}{3!} + \dots && &&\text{for } x \in (-1,1)
\end{aligned}$$

7.5.2 Computing the Taylor Expansion of a function with 2 inputs and 1 output

For a planar function about the origin,

$$\begin{aligned} f(x,y) = f(0,0) \\ &+ \left.\frac{\partial f}{\partial x}\right|_{\underline{0}} x + \left.\frac{\partial f}{\partial y}\right|_{\underline{0}} y \\ &+ \frac{1}{2}\left.\frac{\partial^2 f}{\partial x^2}\right|_{\underline{0}} x^2 + \left.\frac{\partial^2 f}{\partial x \partial y}\right|_{\underline{0}} xy + \frac{1}{2}\left.\frac{\partial^2 f}{\partial y^2}\right|_{\underline{0}} y^2 \\ &+ \frac{1}{6}\left.\frac{\partial^3 f}{\partial x^3}\right|_{\underline{0}} x^3 + \frac{1}{2}\left.\frac{\partial^3 f}{\partial x^2 \partial y}\right|_{\underline{0}} x^2 y + \frac{1}{2}\left.\frac{\partial^3 f}{\partial x \partial y^2}\right|_{\underline{0}} xy^2 + \frac{1}{6}\left.\frac{\partial^3 f}{\partial y^3}\right|_{\underline{0}} y^3 \end{aligned}$$

7.5.3 Taylor Series and the Chain Rule

The composition of Taylor Series is the Taylor Series of the composition

7.5.4 Using Taylor Series for local analysis

Taylor Series can be used to analyze functions locally

For example, given a system

$$sin(xy) = e^{x^3} - cos(x^2 y)$$

We can find solutions locally near (0.0) by taking the Taylor series of both sides

$$xy - O((xy)^3) = \left(1 + x^3 + O(X^6) - (1 - O(x^2 y)^2)\right)$$

Simplifying, we get

$$xy - x^3 = 0$$

7.5.5 Taylor Series up to the second order

$$f(\underline{a} + \underline{h}) = f(\underline{a}) + \frac{1}{1!}\left[Df\right]_{\underline{a}} \underline{h} + \frac{1}{2!} h^T \left[D^2 f\right]_{\underline{a}} \underline{h} + O(\|h\|^3)$$

7.5.6 The Hessian

Definition:

$$\left[D^2 f\right]_{\underline{a} i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Hence,

$$\left[D^2 f\right] = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \frac{\partial^2 f}{\partial x_j^2} & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Chapter 8: Week 8

8.1 Critical points and extrema

8.1.1 Critical points

For a function $f : \mathbb{R}^n \to \mathbb{R}$, a critical point is an input whose derivative is zero or undefined (i.e. all partials are zero or undefined)

f attains a local maximum or minimum at $\underline{a}$ only if the input $\underline{a}$ is a critical point

Said contrapositively, non-critical points cannot be local maxima or minima

8.1.2 Trace-Determinant Method

Recall that

$$f(\underline{a} + \underline{h}) = f(\underline{a}) + [Df]_{\underline{a}}\, \underline{h} + \frac{1}{2!}\underline{h}^T \left[D^2 f\right]_{\underline{a}} \underline{h} + O(\|h\|^3)$$

Since at a critical point, $[Df]_{\underline{a}} = 0$, the second derivative dominates the local behavior

For all $\underline{h} \neq 0$,

- if $\underline{h}^T \left[D^2 f\right]_{\underline{a}} \underline{h} > 0$ then critical point is a local minimum
- if $\underline{h}^T \left[D^2 f\right]_{\underline{a}} \underline{h} > 0$ then critical point is a local maximum

However, to compute the result requires eigenvalues beyond the scope of this course.
Hence, we focus on the case

$$f : \mathbb{R}^2 \to \mathbb{R}$$

The second order derivative is

$$\left[D^2 f\right] = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Recall that the determinant is

$$Det = ad - bc$$

The trace is given by

$$TR = a + d$$

We can follow the following steps

- if $Det < 0$
  - saddle point
- if $Det > 0$
  - if $trace > 0$

    * local minima
  - if $trace < 0$
    * local maxima

8.2 Optimization & Linear regression

Linear regression refers to finding the best fit line for a dataset.

8.2.1 Least-squares method

The least squares method is one of the most common forms of computing a best fit line.

Given data presented as n pairs of (x_i, y_i), find a best fit line given by

$$y = mx + b$$

By minimizing the squared sum of vertical differences between points and the line. The sum is given by

$$f(m, b) = \sum_i (y_i - (mx_i + b))^2$$

To minimize

$$f(m, b) = \sum_i (y_i - (mx_i + b))^2$$

$$\begin{aligned}
\frac{\partial f}{\partial m} &= \sum_i \frac{\partial}{\partial m} (y_i - (mx_i + b))^2 \\
&= \sum_i 2 (y_i - (mx_i + b)) \frac{\partial}{\partial m} (y_i - (mx_i + b)) \\
&= \sum_i 2 (y_i - (mx_i + b)) (-x_i) \\
&= \sum_i -2x_i (y_i - (mx_i + b))
\end{aligned}$$

$$\begin{aligned}
\frac{\partial f}{\partial b} &= \sum_i 2 (y_i - (mx_i + b)) \frac{\partial}{\partial b} (y_i - (mx_i + b)) \\
&= \sum_i -2 (y_i - (mx_i + b))
\end{aligned}$$

To find the critical points wrt m and b, we solve for when the partial derivatives are 0
Solving for $\frac{\partial f}{\partial m}$

$$\begin{aligned}
&\sum_i x_i (y_i - mx_i - b) = 0 \\
&\sum_i x_i y_i - \sum_i m x_i^2 - \sum_i b x_i = 0 \\
&\sum_i x_i y_i - m \sum_i x_i^2 - b \sum_i x_i = 0 \\
&m \sum_i x_i^2 + b \sum_i x_i = \sum_i x_i y_i
\end{aligned}$$

Solving for $\frac{\partial f}{\partial b}$

$$\begin{aligned}
\sum_i (y_i - mx_i - b) = 0 \\
\sum_i y_i - \sum_i mx_i - \sum_i b = 0 \\
\sum_i y_i - m\sum_i x_i - b\sum_1 b = 0 \\
m\sum_i x_i + b\sum_1 b = \sum_i y_i
\end{aligned}$$

Notice that we end up with a linear system of the form

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

Written as

$$\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & \sum 1 \end{bmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

To solve the system of linear equations, recall that we can find the inverse of linear transformation A, and find $A^{-1}\underline{b}$

Notice that

$$Det\ A = \sum 1 \sum x_i^2 - \left(\sum x_i\right)^2 = n\sum x_i^2 - \left(\sum x_i\right)^2$$

Hence the inverse A^{-1} is

$$\frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} n & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{bmatrix}$$

Finally, we can solve for

$$\begin{pmatrix} m \\ b \end{pmatrix} = A^{-1}\underline{b}$$

8.2.2 Optimization & Nash Equilibria

In some games, choosing strategies at random with a probability distribution could lead to predictable expected outcome, and can be optimized.
Given a payoff matrix for player A

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

Suppose players choose their strategy with the following distributions

$$\underline{a} = \begin{pmatrix} a \\ 1-a \end{pmatrix}, \underline{b} = \begin{pmatrix} b \\ 1-b \end{pmatrix}$$

The expected payoff is

$$f(a,b) = \underline{a}^T P \underline{b} = -4 + 7a + 7b - 12ab$$

We can optimize f by finding the critical points.

$$\frac{\partial f}{\partial a} = 7 - 12b = 0$$

$$\frac{\partial f}{\partial b} = 7 = 12a = 0$$

$$b = a = \frac{7}{12}$$

By taking the Hessian, we observe that this point is a saddle point. The significance of saddle points is that they are the Nash Equilibrium, where neither players can do any better.

8.2.3 The Minimax Theorem

Theorem: Given any payoff matrix P, there exists a Nash Equilibrium $(\underline{a}, \underline{b})$ in mixed strategy such that

$$MAX\underline{x}^T(P\underline{b}) = MIN(\underline{a}^T P)\underline{y}$$

8.2.4 Constrained Optimization

Recall that in 1-D, for functions of the form

$$f : \mathbb{R} \to \mathbb{R}$$

Where the input is bounded within an upper and lower bound, to find the local minimum or maximum, we

- check within the boundary (differentiate and find the critical points)
- check at the boundary (calculate $f(a)$ where a is the boundary value

In higher dimensions such as 2D or 3D, we can use the constraints to convert the problem into a 2D case of the form

$$f : \mathbb{R}^2 \to \mathbb{R}$$

And solve for $[Df]_{\underline{a}} = 0$
In general, the steps for solving constrained optimization is

- Solve for $[Df]_{\underline{a}} = 0$
- Check that the points fall within the constraints
- Classify critical points
- Check the boundaries & end points

8.2.5 The Lagrange Multiplier

Consider a function of the form $f : \mathbb{R}^2 \to \mathbb{R}$ and a constraint G, one way to approach such a problem would be to parameterize the constraint, and solve as a single variable optimization problem.

However, if the constraint cannot be easily parameterized, we will need to solve using level sets.

More specifically, at the critical point, the level sets of the function to be optimized and the level sets of the constraints are tangent. Hence, their gradients are parallel.

Given a differentiable function of the form

$$f : \mathbb{R}^n \to \mathbb{R}$$

and a constraint

$$G : \mathbb{R}^n \to \mathbb{R}$$

Any extrenmum $\underline{a}$ of $f(\underline{x})$ restricted to $\underline{G(x) = 0}$ must satisfy

$$[Df]_{\underline{a}} = \lambda [DG]_{\underline{a}}$$

for some λ, provided $[DG]_{\underline{a}} \neq 0$

This is equivalent to finding the unconstrained optima of a new function

$$L(\underline{x}, \lambda) = f - \lambda G : \mathbb{R}^{n+1} \to \mathbb{R}$$

Chapter 9: Integration

Recall from single variable that there were two different kinds of integrals

	Definite	Indefinite
Notation	$\int_{x=a}^{b} f(x)dx$	$\int f(x)dx)$
Result	Scalar value	Class of functions
Definition	A limit of Riemann Sums	An antiderivative
Application	Used to compute area, work, force, etc	Used to compute definite integrals

For this course, we will focus on definite integrals.

Recall that the single variable definite integral is defined as

$$\int_{x=a}^{b} f(x)dx = \lim_{\triangle x \to 0^+} \sum_i f(x_i)\triangle x$$

The two objects are joined by the Fundamental Theorem of Integral Calculus

- From the indefinite integral, find the antiderivative, evaluate at the end points and take the difference
- You could think of the indefinite integral as a **means to an end**

In multivariable calculus, there is **no longer the indefinite integral**

- The indefinite integral has no multivariate analogue

There are certain 1-D interpretations of the derivative and integral that do not persist into higher dimensions. For example

- The derivative as *the slope*
- The integral as *the area*

Instead, we work with the intuition

- The derivative as a **linear transformation**
- The integral as a **mass**

In the multivariate case, we use the following notation

For $f : \mathbb{R}^n \to R$ the integral is defined as

$$\int_R d\underline{x}$$

OR

$$\int \ldots \iint_R f dx_1 \; dx_2 \; \ldots dx_n$$

9.1 Computing integrals and the Fubini theorem

Theorem: Some complicated stuff
Idea: For some function which depends on a bunch of variables

$$f(\underline{x})d\underline{x}$$

where

$$\underline{x} = (x_1, x_2, \dots x_n)$$

$$d\underline{x} = dx_1, dx_2, \dots dx_n$$

The integral over $\mathbb{R}^n$ with respect to the volume element $d\underline{x}$ is

$$\int_{\mathbb{R}^n} f(\underline{x})d\underline{x} = \int \left(\dots \left(\int \left(\int f dx_1 \right) dx_2 \right) \right) dx_n$$

Most importantly: the order in which the integral is computed does not matter

9.2 Double & Triple Integrals

A double integral is given by

$$\iint_R f\ dA = \iint_R f(x,y)\ dx\ dy$$

A triple integral is given by

$$\iiint_R f\ dV = \int_R f(x,y,z)\ dx\ dy\ dz$$

9.2.1 Finding the area in 2 dimensions

To find the area, we integrate the area element dA,

$$A = \int dA$$

Where dA, or the area of an infinitesimal rectangle, is given by

$$dA = dx\ dy = dy\ dx$$

Therefore,

$$A = \int dA = \iint 1dx\ dy = \iint 1dy\ dx$$

9.2.2 Changing the limits

In general, to figure out the limits of integration, we

1. choose a variable to integrate with respect to, such as dy
2. **fix** x, i.e. for a given value of x, read the top and bottom boundaries, which are some functions of $y = f(x)$
3. determine if integral needs to be split into the sum of multiple integrals
4. move on to the next variable

9.3 Averages

For the classical case, in 0-dimensions

$$f : 1, 2, \dots, n \to \mathbb{R}$$

f maps discrete values to a point each, the average is

$$\overline{f} = \frac{1}{n} \sum_{i=1}^{n} f_i$$

In 1-dimension

$$f : [a, b] -> \mathbb{R}$$

f maps values between the boundaries a, b to a value, the average is

$$\overline{f} = \frac{1}{b-a} \int_a^b f$$

In the multivariate case

$$\overline{f} = \frac{\int_R f d\underline{x}}{\int_R d\underline{x}} = \frac{1}{VOL_n(R)} \int_R f d\underline{x}$$

Chapter 10: Week 10

10.1 Centroids & Centers of mass

10.1.1 Centroids in 2D

The centroids of a region between graphs can be found using double integrals and averages

$$\begin{aligned}
\overline{x} &= \frac{1}{A}\iint_R x\ dA \\
&= \frac{1}{A}\int_{x=a}^{b}\int_{y=g(x)}^{f(x)} x\ dy\ dx
\end{aligned}$$

$$\begin{aligned}
\overline{y} &= \frac{1}{A}\iint_R y\ dA \\
&= \frac{1}{A}\int_{x=a}^{b}\int_{y=g(x)}^{f(x)} y\ dy\ dx
\end{aligned}$$

10.1.2 Centroids in 3D

The centroid of a region R in $\mathbb{R}^n$ is the point $\overline{\underline{x}}$ with coordinates The $i-th$ component of the centroid coordinates in $x_1, x_2 \dots x_n$ is

$$\begin{aligned}
\overline{x_i} &= \frac{1}{V}\int_R x_i d\underline{x} \\
&= \frac{\int_R x_i d\underline{x}}{\int_R 1 d\underline{x}}
\end{aligned}$$

10.1.3 Center of Mass

The **center of mass** of a region R in $\mathbb{R}^n$ with mass density $\rho(\underline{x})$ is the point $\overline{\underline{x}}$ with coordinates

$$\begin{aligned}
\overline{x_i} &= \frac{\int_R x_i \rho(\underline{x}) d\underline{x}}{\int_R \rho(\underline{x}) d\underline{x}} \\
&= \frac{\int_R x_i dM}{\int_R dM} \\
&= \frac{1}{M}\int_R x_i dM
\end{aligned}$$

10.2 Moments of inertia

Let $R \in \mathbb{R}^n$ be a solid body to be rotated about an axis
The moment of inertia measures the resistance to rotation, where

$$I = \int_R dI$$

and

$$dI = r^2 dM$$

10.2.1 Parallel Axis Theorem

If I_0 is the moment of inertia of an object about an axis through its center of mass, then the moment of inertia about a parallel axis D away equals

$$I_D = I_0 + MD^2$$

10.2.2 Radius of Gyration

Let $\mathbb{R} \in \mathbb{R}^n$ be a solid body to be rotated about an axis. If all the mass were concentrated at a single point, the radius of gyration is the radius at which the concentrated mass would have the same moment of inertia

$$I = Mr_g^2$$

$$r_g = \sqrt{\frac{I}{M}}$$

10.3 Inertia Matrix

10.4 Solid Body Mechanics

10.5 Probability & Integration

For a random $\underline{X}$ taking on values in R^n

The probability density function is

$$\rho : \mathbb{R}^n \to \mathbb{R}$$

such that

$$\rho \geq 0$$

and

$$\int_{\mathbb{R}^n} \rho(\underline{x}) d\underline{x} = 1$$

The probability element is given by

$$dP = \rho(\underline{x}) d\underline{x}$$

Hence the probability of $P(\underline{X} \in A)$ is

$$P(\underline{X} \in A) = \int_A dP = \int_A \rho(\underline{x}) d\underline{x}$$

The expectation $E(\underline{X})$ is the ρ weighted centroid

$$E(\underline{X}) = \int_{\mathbb{R}^n} \underline{x}\rho(\underline{x})d\underline{x}$$

10.6 Independence & Covariance

10.7 Covariance Matrices

Chapter 11: Week 11

11.1 Polar & Cylindrical coordinates

11.1.1 Polar Coordinates

The area element in polar coordinates is given by

$$dA = r\ dr\ d\theta$$

The infinitesimal area element is given by the product of

- dr in the radial direction
- $rd\theta$ in the angular direction

The transformation is

$$\begin{array}{ll} x = rcos\theta & r = \sqrt{x^2 + y^2} \\ y = rsin\theta & \theta = arctan(\frac{y}{x}) \end{array}$$

11.1.2 Cylindrical Coordinates

The volume element in cylindrical coordinates is given by

$$dV = r\ dr\ d\theta dz$$

The infinitesimal volume element is given by the product of

- dr in the radial direction
- $rd\theta$ in the angular direction
- z in the z axis

The transformation is

$$\begin{array}{ll} x = rcos\theta & r = \sqrt{x^2 + y^2} \\ y = rsin\theta & \theta = arctan(\frac{y}{x}) \end{array}$$

11.2 Spherical coordinates

The volume element is given by

$$dV = \rho^2 sin\ \phi d\rho\ d\theta\ d\phi$$

Spherical coordinates are defined by

- radius ρ

- inclination ϕ
- azimuth θ

Standard coordinates and spherical coordinates can be translated by

$$x = \rho \; cos \; \theta \; sin \; \phi$$
$$y = \rho \; sin \; \theta \; sin \; \phi$$
$$z = \rho \; cos \; \phi$$

Where

$$\rho \geq 0$$
$$0 \leq \theta \leq 2\pi$$
$$0 \leq \phi \leq \pi$$

To convert standard coordinates to polar coordinates

$$\rho = \sqrt{x^2 + y^2 + z^2} = \sqrt{r^2 + z^2}$$
$$\theta = arctan(\frac{y}{x})$$
$$\phi = arctan\left(\frac{r}{z}\right) = arccos\left(\frac{z}{\rho}\right)$$

11.3 Change in Variables Theorem

11.3.1 Linear transformations and the volume element

Recall from earlier in the semester that

- The determinant of a square matrix computes a volume
- Square matrices encode linear transformation

For a linear transformation A that takes $\underline{x}$ coordinates to $\underline{u}$ coordinates,

$$\underline{u} = A\underline{x}$$

The volume element is transformed as

$$d\underline{u} = |Det\,(A)|\, d\underline{x}$$

11.3.2 Non-linear transformations and the volume element

Given a change of coordinates $F : \mathbb{R}^n \to \mathbb{R}^n$ of the form

$$\underline{u} = F\,(\underline{x})$$

The volume element is transformed as

$$d\underline{u} = |Det\,[Df]|\, d\underline{x}$$

11.3.3 Change of variables theorem

Recall that in single variable integral calculus, u-sub is given by

$$\int_{u_0}^{u_1} h(u)du = \int_{x_0}^{x_1} h(u(x))\frac{du}{dx}dx$$

Theorem: Given a change of coordinates $F : \mathbb{R}^n \to \mathbb{R}^n$ of the form

$$\underline{u} = F\left(\underline{x}\right)$$

The integral of h over a region $F(R)$ is converted as

$$\int_{F(R)} h\left(\underline{u}\right) d\underline{u} = \int_{R} h\left(F\left(\underline{x}\right)\right) \left|Det\left[DF\right]\right| d\underline{x}$$

11.4 Surface Integrals

11.4.1 Parameterized surfaces

In 3-D, a parameterized surface has two inputs and three outputs, i.e.

$$G\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x(s,t) \\ y(s,t) \\ z(s,t) \end{pmatrix}$$

The derivative is a matrix with two columns, each of which gives a tangent vector.

$$[DG] = \begin{bmatrix} \frac{\partial G}{\partial t} & \frac{\partial G}{\partial s} \end{bmatrix}$$

The infinitesimal surface area element is given by the area of a parallelogram, bounded by $\frac{\partial G}{\partial t}$ and $\frac{\partial G}{\partial s}$

$$\text{area of parallelogram} = \left|\frac{\partial G}{\partial t} \times \frac{\partial G}{\partial s}\right|$$

Hence,

For a surface $S \in R^3$ parameterized by $G : \mathbb{R}^2 \to \mathbb{R}^3$

$$G\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

The surface area σ is given by

$$\int_{S=G(R)} d\sigma = \iint_{R} \left|\frac{\partial G}{\partial s} \times \frac{\partial G}{\partial t}\right| ds\ dt$$

In spherical coordinates

$$d\sigma = \mathbb{R}^2 sin\phi d\theta\ d\phi$$

Chapter 12: Week 12

12.1 Fields

In general, a Smiley field in $\mathbb{R}^n$ means that there is a smiley for every point in $\mathbb{R}^n$

12.1.1 Vector Fields

Recall that the gradient of a scalar field is a vector field.
In 2D

$$\vec{F} = \vec{F_x}\hat{\imath} + \vec{F_x}\hat{\jmath}$$

Some easy-to-remember planar vector fields include

- the rotational vector field: $\vec{F} = -y\hat{\imath} + x\hat{\jmath}$
- the saddle vector field: $\vec{F} = x\hat{\imath} - y\hat{\jmath}$
- radial vector fields: $\vec{F} = \sum_{i=1}^{n} x_i \underline{e_i}$

12.1.2 Other fields

More generally, there are other fields that could be useful, such as

- Taylor polynomial field: every point is associated with a Taylor Expansion
- Derivative: the derivative $[Df]$ of $f : \mathbb{R}^n \to \mathbb{R}^m$ is a $m \times n$ matrix field over $\mathbb{R}^n$

12.2 Path Integrals

12.2.1 Integrating a scalar field

Given a vector field $f : \mathbb{R}^n \to \mathbb{R}$ and a path $\gamma : [a, b] \to \mathbb{R}^n$

The path integral can be computed by integrating the arc length element, weighted by the scalar field

$$\begin{aligned} &\int_\gamma f dl \\ =& \int_\gamma f\,|\gamma'|\,dt \\ =& \int_{t=a}^{b} f\left(\gamma\left(t\right)\right)\left|\gamma'\left(t\right)\right| dt \end{aligned}$$

12.2.2 The standard single-variable integral as a path integral

The standard single variable integral

$$\int_{x=1}^{b} f(x)dx$$

Is really the path integral from a to b of a path

$$\gamma(t) = t : a \leq t \leq b$$

Based on the parameterization

$$|\gamma'(t)| = 1$$

Hence

$$\int_{\gamma} f \; dl = \int_{t=1}^{b} f(t)dt$$

12.2.3 Rules for a scalar path integral

Since the scalar path integral is what lines behind standard single variable integral, many of the rules are to be as expected

- Linearity: $\int_{\gamma} cf + gdl = c\int_{\gamma} fdl + \int_{x}$
- additivity $\int_{\gamma+\delta} = \int_{\gamma} fdl + \int_{\delta} fdl$
- reversal $\int_{-\gamma} = \int_{\gamma} fdl$

12.2.4 Independence of path in scalar fields

Path integrals are independent of the parameterization of the path

12.3 Integrating 1-forms

Given a vector field $\vec{F}$ on $\mathbb{R}^n$ and a parameterized path $\gamma : [a, b] \to \mathbb{R}^n$

$$\int_{\gamma} \vec{F} \cdot d\underline{x} = \int_{\gamma} \alpha_{\vec{F}} = \int_{\gamma} F_1 dx_1 + \ldots + F_n dx_n$$

12.3.1 1-Forms

Definition: a differential 1-form on $\mathbb{R}^n$ is a linear function α that takes in a vector $\underline{v}$ in $\mathbb{R}^n$ and returns a scalar $\alpha(\underline{v})$ linearly

For example, in $\mathbb{R}^3$, for a vector $\underline{v}$

$$\underline{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}$$

$$dx\,(\underline{v}) = v_x$$
$$dy\,(\underline{v}) = v_y$$
$$dz\,(\underline{v}) = v_z$$

dx can be thought of as an operator

$$dx \sim \hat{i}\cdot \text{ OR } \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot$$

12.3.2 Gradient vector field, gradient 1-form field, and the derivative

For a function $f : \mathbb{R}^n \to \mathbb{R}$, we have
The gradient field

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

The 1-form field

$$df = \frac{\partial f}{\partial x_1} dx_1 + \dots \frac{\partial f}{\partial x_n} dx_n$$

The derivative

$$[Df] = \begin{bmatrix} \frac{\partial f}{\partial x_1} \dots \frac{\partial f}{\partial x_n} \end{bmatrix}$$

For a rate of change of input denoted by $\underline{u}$

$$df(\underline{u}) = \nabla f \cdot \underline{u} = [Df]\, \underline{u}$$

12.3.3 Integrating 1-forms

Given a 1-form field α and a parameterized path $\gamma : [a, b] \to \mathbb{R}^n$

$$\int_\gamma \alpha = \int_{t=a}^{b} \alpha_{\gamma(t)} \left(\gamma'(t)\right) dt$$

12.4 Independence of Path

Theorem: let $\gamma : [a, b] \to \mathbb{R}^n$ be a path and given a scalar field $f : \mathbb{R}^n \to \mathbb{R}$
The integral of the gradient 1-form df equals

$$\int_\gamma df = f|_{\gamma(a)}^{\gamma(b)} = f(\gamma(b)) - f(\gamma(a))$$

12.5 Detecting gradients

Lemma: for a 1-form α where

$$\alpha = \sum_i \alpha_i dx_i$$

on $\mathbb{R}^n$ is a gradient if and only if

$$\frac{\partial \alpha_i}{\partial x_j} = \frac{\partial \alpha_j}{\partial x_i}$$

for all $i \neq j$

12.6 Work, Circulation and Flux

12.6.1 Work

For a planar vector field

$$\vec{F} = F_x i + F_y j$$

The work 1-form is

$$\alpha_{\vec{F}} F_x dx + F_y dy$$

The work done is

$$W = \int_\gamma \alpha_{\vec{F}}$$

12.6.2 Circulation

Circulation denotes work done around a closed loop

Gradient 1-forms have vanishing integrals around all closed loops

12.6.3 Flux

For a planar vector field

$$\vec{F} = F_x i + F_y j$$

The flux 1-form is

$$\phi \vec{F} F_x dy - F_y dx$$

We can compute the flux by

$$\phi_\gamma = \int_\gamma \phi_{\vec{F}} = \int_\gamma F_x dy - F_y dx$$

Chapter 13: Week 13

13.1 Green's Theorem

13.1.1 The Theorem

Let $D \in \mathbb{R}^2$ be a bounded domain with oriented boundary ∂D. Then, for P, Q differentiable on all of D,

$$\int_{\partial D} Pdx + Qdy = \iint_D \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} dA$$

Green's Theorem relates the integral of a 1-form over the **boundary of a domain** in the plane to the integral of a certain derivative over the **domain interior**

For some special 1-forms..., the resulting integral is equal to the area of the bounded domain

- $\alpha = xdy$

$$\int_\gamma xdy = \iint_D \frac{\partial}{\partial x} xdA = A$$

13.1.2 Orientation and Boundaries

Orientation on a planar domain can either be

- positive: counterclockwise (by default)
- negative: clockwise

The default is to assume a positive orientation on a planar domain. Such an orientation induces a counterclockwise rotation on the boundary.

A simply connected domain has one boundary component.

A multiply connected domain has multiple boundary components, consistent with the notion of a infinitesimal counterclockwise rotation.

13.1.3 Differentiability, orientation & Green's Theorem

Recall that in order to apply Green's Theorem over D, the 1-form must be differentiable on **all of** D
Consider a 1-form

$$\alpha = \frac{-y}{x^2+y^2} dx + \frac{x}{x^2+y^2} dy$$

Suppose we want to integrate the 1-form on $\mathbb{R}^2$ over the counterclockwise unit circle u, we can compute the integral using u-sub

$$\begin{aligned} \int_u \frac{-y}{x^2+y^2} dx + \frac{x}{x^2+y^2} dy &= \int_{t=0}^{2\pi} \frac{-sint(-sint)}{sin^2t + cos^2t} + \frac{cost(cost)}{sin^2t + cos^2t} dt \\ &= 2\pi \end{aligned}$$

However, if we try to use Green's Theorem, the same integral becomes

$$\iint_D \frac{\partial}{\partial x}\frac{x}{x^2+y^2} - \frac{\partial}{\partial y}\frac{-y}{x^2+y^2}dA = \iint_D \frac{(x^2+y^2)-2x^2+(x^2+y^2)-2y^2}{(x^2+y^2)^2} = 0 \neq 2\pi$$

This is because the 1-form is not differential at the origin

Now, if we want to integrate the 1-form over some arbitrarily complicated, singly connected domain D_0 that includes the origin, we can compute the integral by instead computing the integral over a **multiply connected** domain D_1 with the unit circle cut out

Note that D_1 has 2 boundaries, the outer boundary identical to D_0, and an inner **clockwise boundary** u, which is the unit circle.

Green's Theorem now applies to D_1, and the integral, as computed previously, is

$$\iint_{D_1} \frac{\partial}{\partial x}\frac{x}{x^2+y^2} - \frac{\partial}{\partial y}\frac{-y}{x^2+y^2}dA = 0$$

Let the outer boundary be γ, inner boundary be u

$$\int_{\gamma-u}\alpha = \int_\gamma \alpha - \int_u \gamma = 0$$

13.2 Gradient, Curl & Divergence

This section combines differentiation, 1-forms and what we learned previously about the gradient

13.2.1 Curl and divergence densities

For a vector field

$$\vec{F} = F_x\underline{i} + F_y\underline{j}$$

We can write down two new versions of Green's Theorem, for circulation and flux respectively

Circulation

$$\int_{\partial D} F_x dx + F_y dy = \iint_D \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}dA$$

Flux

$$\int_{\partial D} F_x dy - F_y dx = \iint_D \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y}dA$$

We can think of the integrands of the double integrals as some form of **densities**. In fact, they are two different types of **derivatives** of the vector fields

Curl

Curl indicates, at each point in a vector field, the infinitesimal counterclockwise spin exhibited by the vector field. It is a limiting intensity of circulation

Density is given by

$$\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}$$

It is denoted as

$$\nabla \times \vec{F}$$

Divergence

Divergence indicates, at each point in a vector field, the infinitesimal expansion or contraction exhibited by the vector field. It is a limiting intensity of flux

Density is given by

$$\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y}$$

It is denoted as

$$\nabla \cdot \vec{F}$$

13.2.2 Green Theorem through the perspective of curl and divergence

Given a vector field

$$\vec{F} = F_x \underline{i} + F_y \underline{j}$$

For $D \in \mathbb{R}^2$ a domain with oriented boundary ∂D

13.2.3 Divergence in 3-D

The divergence of a 3-D vector field is a scalar field.

The divergence at a point indicates the local expansion or contraction of the **volume element**

$$\nabla \cdot \vec{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

13.2.4 Curl in 3-D

The curl of a 3-D vector field is another vector field.

The formula for curl has as its components the three circulation densities in each plane

$$\nabla \times \vec{F} = Det \begin{bmatrix} i & \frac{\partial}{\partial x} & F_x \\ j & \frac{\partial}{\partial y} & F_y \\ k & \frac{\partial}{\partial z} & F_z \end{bmatrix}$$

$$\nabla \times \vec{F} = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) \underline{i} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) \underline{j} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \underline{k}$$

13.3 Differential Forms in 3-D

13.3.1 Euclidean Forms

1-forms

In 3-D, we have the basis 1-forms which give **oriented projected length** along the basis axes

$$dx, dy, dz$$

From the basis 1-forms, a **linear** 1-form is of the following form, and returns **oriented projected length** along some direction

$$\alpha = a\ dx + b\ dy + c\ dz$$

A 1-form field is one of the form

$$\alpha = f_1 dx + f_2 dy + f_3 dz$$

2-forms

In general, a k-form takes k ordered vectors and returns a scalar
In 3-D, we have basis 2-forms which give **oriented projected area**

$$dx \wedge dy, dy \wedge dz, dz \wedge dx$$

Each basis 2-form takes 2 vectors and returns some scalar, where

- dx extracts the x component
- dy extracts the y component
- dz extracts the z component

The 2-forms are thus

$$\begin{aligned}(dx \wedge dy)(\underline{u}, \underline{v}) &= Det \begin{bmatrix} u_x & v_x \\ u_y & v_y \end{bmatrix} \\ (dy \wedge dz)(\underline{u}, \underline{v}) &= Det \begin{bmatrix} u_y & v_y \\ u_z & v_z \end{bmatrix} \\ (dz \wedge dx)(\underline{u}, \underline{v}) &= Det \begin{bmatrix} u_z & v_z \\ u_x & v_x \end{bmatrix}\end{aligned}$$

A linear 2-form is one of the form

$$\beta = a\ dx \wedge dy + b\ dy \wedge dx + c\ dz \wedge dx$$

A 2-form field is one of the form

$$\beta = f_1 dx \wedge dy + f_2 dy \wedge dz + f_3 dz \wedge dx$$

3-form
The basis 3-form is

$$dx \wedge dy \wedge dz$$

The basis 3-form gives the oriented volume when given an ordered triplet of vectors $(\underline{u}, \underline{v}, \underline{w})$ in $\mathbb{R}^3$

$$(dx \wedge dy \wedge dz)(\underline{u}, \underline{v}, \underline{w}) = Det \begin{bmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{bmatrix}$$

13.3.2 The Wedge

The wedge product is a symbolic representation of stacking vectors into a determinant.
The rules are

- $dx_i \wedge dx_j = -dx_j \wedge dx_i$,
- $dx_i \wedge dx_i = 0$

13.3.3 Differentiation of form fields

Recall that the derivative of a scalar field f (0-form field) is a gradient 1-form

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz$$

Differentiating a 1-form field, we have

$$d(f dx_i) = df \wedge dx_i$$

For example,

$$\begin{aligned}d(f\ dx) &= df \wedge dx \\ &= \frac{\partial f}{\partial y}dy \wedge dx + \frac{\partial f}{\partial z}dz \wedge dx\end{aligned}$$

In general, taking the derivative of a k-form field gives a $k+1$ form field

Flux

$$\phi_{\vec{F}}(\underline{u}, \underline{v}) = Det\left[\vec{F}|\underline{u}|\underline{v}\right]$$

$$\phi_{\vec{F}} = F_x dy \wedge dz + F_y dz \wedge dx + F_z dx \wedge dy$$

Curl

Computing the derivative of

$$\alpha = f_1 dx + f_2 dy + f_3 dz$$

$$d\alpha = df_1 \wedge dx + df_2 \wedge dy + df_3 \wedge dz$$

Where

$$df_i = \frac{\partial f_i}{\partial x} dx + \frac{\partial f_i}{\partial y} dy + \frac{\partial f_i}{\partial z} dz$$

$$\begin{aligned} d\alpha &= \left(\frac{\partial f_1}{\partial y} dy + \frac{\partial f_1}{\partial z} dz\right) \wedge dx + \left(\frac{\partial f_2}{\partial x} dx + \frac{\partial f_2}{\partial z} dz\right) \wedge dy + \left(\frac{\partial f_3}{\partial x} dx + \frac{\partial f_3}{\partial y} dy\right) \wedge dz \\ &= \left(\frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z}\right) dy \wedge dz \left(\frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x}\right) dz \wedge dx \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\right) dx \wedge dy \end{aligned}$$

Hence

$$d\alpha_{\vec{F}} = \phi_{\nabla \times \vec{F}}$$

Divergence

Finding the derivative of

$$\beta = f_1 dy \wedge dx + f_2 dz \wedge dx + f_3 dx \wedge dy$$

$$d\beta = \left(\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z}\right) dx \wedge dy \wedge dz$$

Chapter 14: Week 14

14.1 Integrating 2-forms

14.1.1 Computing the integral

A 2-form is a weighted area form.

$$\int_D f(x,y)dx \wedge dy = \iint_D f(x,y)dA$$

Recall Green's Theorem

$$\int_{\partial D} Pdx + Qdy = \iint \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}dxdy$$

If we define

$$\alpha = Pdx + Qdy$$

then

$$d\alpha = \frac{\partial P}{\partial y}dy \wedge dx + \frac{\partial Q}{\partial x}dx \wedge dy$$

More generally,

$$\int_{\partial D} \alpha = \int_D d\alpha$$

To integrate a 2-form field over a parameterized surface defined by $G(s,t)$, a pair of basis vectors in the input is transformed by $[DG]_{(s,t)}$ to a pair of output vectors

At each point on the surface, there is a pair of tangent vectors $\frac{\partial G}{\partial s}$ and $\frac{\partial G}{\partial t}$, the columns of $[DG]_{(s,t)}$

Given a 2-form field β, we define the integral

$$\int_S \beta = \iint_R \beta_{G(s,t)} \left[DG\right]_{(s,t)} dsdt$$

14.1.2 Computing flux

2-forms fields can be thought of as flux forms in 3-D

For a 3-D vector field

$$\vec{F} = F_x\underline{i} + F_y\underline{j} + F_z\underline{k}$$

There is an associated flux 2-form

$$\Phi_{\vec{F}} = F_x dy \wedge dz + F_y dz \wedge dx + F_z dx \wedge dy$$

The flux 2-form tells you how much "stuff" is getting pushed through an infinitesimal oriented area

The integral of a flux 2-form gives the net flux, or how much "stuff" is getting pushed through an oriented surface

The flux of $\vec{F}$ across S is

$$\int_S \vec{F} \cdot \underline{n} d\sigma = \int_S \Phi_{\vec{F}}$$

14.2 Gauss' Theorem

Recall Green's Theorem

Flux form
Integrating the flux 1-form over some domain $D \in \mathbb{R}^2$ with oriented boundary ∂D is the same as integrating the **2-D** divergence with respect to the area element

$$\int_{\partial D} F_x dy - F_y dx = \iint_D \nabla \cdot \vec{F} dA$$

OR

$$\int_{\partial D} \Phi_{\vec{F}} = \int_D d\Phi_{\vec{F}}$$

In 3-D, this is Gauss' Theorem

14.2.1 The Theorem

Theorem: Let $D \in \mathbb{R}^3$ be a bounded domain with boundary ∂D, oriented by outward-pointing normal $\underline{n}$. Then, for a smooth vector field $\vec{F}$ on D,

$$\iint_{\partial D} \vec{F} \cdot \underline{n} d\sigma = \iiint_D \nabla \cdot \vec{F} dV$$

OR

$$\int_{\partial D} \Phi_{\vec{F}} = \int D d\Phi_{\vec{F}}$$

14.3 Stoke's Theorem

Gauss' Theorem is a 3-D generalization of the **flux form of Green's Theorem**

Now we consider the circulation form of Green's Theorem
For a vector field

$$\vec{F} = F_x \underline{i} + F_y \underline{j}$$

and $D \in \mathbb{R}^2$ a domain with oriented boundary ∂D

$$\int_{\partial D} F_x dx + F_y dy = \iint_D \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) dA = \iint_D \left(\nabla \times \vec{F} \right) \cdot \underline{k} dA$$

Or

$$\int_{\partial D} \alpha_{\vec{F}} = \int_D d\alpha_{\vec{F}}$$

Theorem: Let $D \in \mathbb{R}^3$ be a bounded surface with boundary ∂D oriented by a field of unit vectors $\underline{n}$ normal to D, then, for a smooth vector field $\vec{F}$ on D,

$$\int_{\partial D} \vec{F} \cdot dl = \iint_D \nabla \times \vec{F} \cdot \underline{n} d\sigma$$

Or
The circulation of a field around the boundary is equal to the net flux of the curl through the surface
Or

$$\int_{\partial D} \alpha_{\vec{F}} = \int_D d\alpha_{\vec{F}}$$

14.3.1 Independence of surface

For some fixed $\gamma = \partial D$, the choice of D does not matter as long as orientation remains the same

14.4 Which theorem to use?