

Linear Algebra and Several Variable Calculus

Willoughby Seago

September 17, 2019

These are my notes for the *linear algebra and several variable calculus (LASVC)* course from the University of Edinburgh as part of the second year of the theoretical physics degree. When I took this course in the 2019/20 academic year it was taught by Professor Philip Clark¹. These notes are based on the lectures delivered as part of this course, and the notes provided as part of this course. The content within is correct to the best of my knowledge but if you find a mistake or just disagree with something or think it could be improved please let me know.

These notes were produced using L^AT_EX². Graphs were plotted using Python³, Matplotlib⁴, NumPy⁵, and SciPy⁶. Diagrams were drawn with tikz⁷. Some images were taken from the lecture notes provided.

This is version 1.0 of these notes, which is up to date as of 04/01/2021.

Willoughby Seago

s1824487@ed.ac.uk

¹<https://www.ph.ed.ac.uk/people/philip-clark>

²<https://www.latex-project.org/>

³<https://www.python.org/>

⁴<https://matplotlib.org/>

⁵<https://numpy.org/>

⁶<https://scipy.org/scipylib/>

⁷<https://www.ctan.org/pkg/pgf>

Contents

1 Basic Vectors	4
2 More Basic Vectors	4
3 Orthonormal Basis Sets	5
3.1 Equation of a Line	5
3.2 Equation of a Plane	5
3.3 Orthonormal Basis Set	6
3.4 Linear Independence	6
4 Orthonormalisation	7
4.1 Creating orthonormal basis sets	7
4.2 Gram-Schmidt Orthonormalisation	8
4.3 Triple Products	8
4.4 Levi-Civita Symbol	8
5 Matrices	9
5.1 More on the Scalar Triple Product	9
5.2 Matrices	9
6 Orthogonal Matrices	10
6.1 Proof of Equivalence	10
7 Active Transforms	11
7.1 Linearity	11
8 Eigenvectors and Eigenvalues	11
8.1 Symmetric and Diagonal Matrices	11
8.2 Eigenvectors and Eigenvalues definition	12
9 Determinants	12
9.1 The Eigenvalue Equation	12
9.2 Determinants as Volume Scale Factor	13
9.3 Permutation Method	13
9.4 Recursive Method	13
10 The Eigenvalue Equation	14
10.1 Useful Determinant Properties	14
10.2 Trace of a Matrix	14
10.3 Solving the Eigenvalue Equation	15
10.4 Coupled Equations	15
11 Linear Systems	16
11.1 Linear System of Equations	16
12 Inverses	17
12.1 Rouchè-Capelli Theorem	17
12.2 Cramer's Rule	17
12.3 Gaussian Elimination	18
12.4 Complex Vectors and Matrices	18
12.5 Dirac Notation	18
13 Partial Derivative	19
13.1 Last Notes on Linear Algebra	19

13.2 Linear Vector Spaces	19
13.3 Single Variable Derivative	20
13.4 Derivatives of Multiple Variables	20
13.5 Product Rule	21
13.6 Gradient	21
14 Gradient Vector	22
15 Change of Variables and Chain Rules	23
15.1 Change of Variables	24
16 Higher Order Derivatives and Taylor Expansions	25
16.1 Taylor Expansions	26
16.2 Stationary Points	26
17 Integration	27
17.1 1D Integrals	27
17.2 2D Integrals	28
17.3 Evaluating 2D Integrals	28
18 Change of Basis	32
18.1 Polar Coordinates	32
18.2 The Jacobian	33
18.3 Cylindrical Coordinates	34
19 Spherical Coordinates and the Laplacian	35
19.1 Integrating in Spherical Coordinates	35
19.2 The Laplacian	35

List of Figures

3.1 A point on the line $\mathbf{r} = \mathbf{p} + \lambda \mathbf{a}$	5
3.2 Multiple orthonormal basis sets	6
6.1 Rotation of basis	10
10.1 Spring mass system	15
15.1 Small step δs in the direction of $\hat{\mathbf{u}}$	23
17.1 Integrating $f(x)$	27
17.2 Integrating $f(x, y)$	28
17.3 A simple 2D integral	29
17.4 Splitting the same shape multiple ways	30
17.5 Region bounded by $x = 0$, $y = x$ and $y = 2 - x$	31
18.1 Integrating over a circle	33

1 Basic Vectors

A vector \mathbf{a} has magnitude, $|\mathbf{a}| = a$, and direction.

A binary operator $*$ acting on elements a and b is commutative if $a * b = b * a$. The same operator is said to be anti-commutative if instead $a * b = -b * a$.

Vector addition is commutative as $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.

The vector $c\mathbf{a}$ for $c > 0$ points in the same direction as \mathbf{a} and has magnitude ca . For $c < 0$ the vector $c\mathbf{a}$ points in the opposite direction to \mathbf{a} and has magnitude $-ca$.

A unit vector is any vector with magnitude 1. It is denoted as $\hat{\mathbf{a}}$ where $\hat{\mathbf{a}}$ points in the same direction as \mathbf{a} and has magnitude 1. The unit direction in the direction of $\hat{\mathbf{a}}$ can be calculated as $\hat{\mathbf{a}} = \mathbf{a}/a$

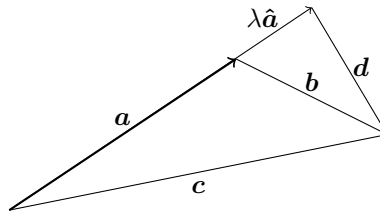
The scalar (dot) product between two vectors \mathbf{a} and \mathbf{b} is defined as $\mathbf{a} \cdot \mathbf{b} = ab \cos \vartheta$ where ϑ is the angle between the vectors.

2 More Basic Vectors

A point P on a line in direction \mathbf{a} can be given as $\overrightarrow{OP} = \lambda \mathbf{a}$. For this reason a line is a 1D vector space.

A point P on a plane containing linearly independent vectors \mathbf{a} and \mathbf{b} can be given as $\overrightarrow{OP} = \lambda \mathbf{a} + \mu \mathbf{b}$. For this reason a plane is a 2D vector space.

For 2D and up it is easier to work with orthogonal vectors as basis vectors as this keeps the scalar coefficients as low as possible. If we have two vectors \mathbf{a} and \mathbf{b} that are linearly independent and a third vector $\mathbf{c} = \mathbf{a} + \mathbf{b}$ then we can perform a process called orthonormalisation to find a set of orthonormal vectors.



\mathbf{d} is perpendicular to \mathbf{a}

$$\begin{aligned}
 \mathbf{c} &= \mathbf{a} + \mathbf{b} \\
 \mathbf{c} &= \lambda \hat{\mathbf{a}} + \mathbf{d} \\
 \mathbf{c} \cdot \hat{\mathbf{a}} &= \underbrace{\lambda \hat{\mathbf{a}} \cdot \hat{\mathbf{a}}}_{=1} + \underbrace{\mathbf{d} \cdot \hat{\mathbf{a}}}_{=0} \\
 \lambda &= \mathbf{c} \cdot \hat{\mathbf{a}} \\
 \mathbf{d} &= \mathbf{c} - \lambda \hat{\mathbf{a}} \\
 \mathbf{d} &= \mathbf{c} - (\mathbf{c} \cdot \hat{\mathbf{a}}) \hat{\mathbf{a}}
 \end{aligned}$$

Given a plane containing linearly independent vectors \mathbf{a} and \mathbf{b} any vector \mathbf{c} can be decomposed into an in-plane (transverse) component \mathbf{v}_t and a normal component \mathbf{v}_n . We assume here that the normal unit vector $\hat{\mathbf{n}}$ for the plane is known as this is the most common way to define a plane. Otherwise this can be calculated as $\mathbf{a} \times \mathbf{b} / |\mathbf{a} \times \mathbf{b}|$.

$$\begin{aligned}
 \mathbf{c} &= \mathbf{v}_t + \mathbf{v}_n \\
 \mathbf{c} &= \mathbf{v}_t + \lambda \hat{\mathbf{n}}
 \end{aligned}$$

$$\begin{aligned}
\mathbf{c} \cdot \hat{\mathbf{n}} &= \mathbf{v}_t \cdot \hat{\mathbf{n}} + \lambda \hat{\mathbf{n}} \cdot \hat{\mathbf{n}} \\
\mathbf{c} \cdot \hat{\mathbf{n}} &= 0 + \lambda \\
\mathbf{c} \cdot \hat{\mathbf{n}} &= \lambda \\
\mathbf{v}_t &= \mathbf{c} - (\mathbf{c} \cdot \hat{\mathbf{n}}) \hat{\mathbf{n}}
\end{aligned}$$

The cross product of \mathbf{a} and \mathbf{b} is defined as the vector normal to both \mathbf{a} and \mathbf{b} following the right hand rule and with magnitude $|\mathbf{a} \times \mathbf{b}| = ab \sin \vartheta$.

3 Orthonormal Basis Sets

3.1 Equation of a Line

A line is in direction \mathbf{a} and point \mathbf{p} lies on the line. The position of a general point \mathbf{r} on the line is given by

$$\begin{aligned}
\mathbf{r} &= \mathbf{p} + \lambda \mathbf{a} \\
\mathbf{r} \times \mathbf{a} &= \mathbf{a} \times \mathbf{p} + \lambda \mathbf{a} \times \mathbf{a} \\
(\mathbf{r} - \mathbf{p}) \times \mathbf{a} &= \mathbf{0}
\end{aligned}$$

This can be understood as \mathbf{p} taking you from the origin to the line and $\lambda \mathbf{a}$ taking you along the line to the desired point.

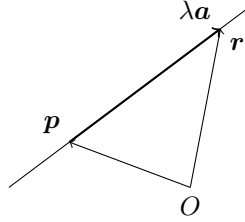


Figure 3.1: A point on the line $\mathbf{r} = \mathbf{p} + \lambda \mathbf{a}$

3.2 Equation of a Plane

For two vectors \mathbf{a} and \mathbf{b} in a plane normal to $\hat{\mathbf{n}}$ with non-zero angle between \mathbf{a} and \mathbf{b} and a point on the plane \mathbf{p} the equation of the plane can be given as:

$$\begin{aligned}
\mathbf{r} &= \mathbf{p} + \lambda \mathbf{a} + \mu \mathbf{b} \\
\mathbf{r} - \mathbf{p} &= \lambda \mathbf{a} + \mu \mathbf{b} \\
(\mathbf{r} - \mathbf{p}) \cdot \hat{\mathbf{n}} &= \lambda \mathbf{a} \cdot \hat{\mathbf{n}} + \mu \mathbf{b} \cdot \hat{\mathbf{n}} \\
(\mathbf{r} - \mathbf{p}) \cdot \hat{\mathbf{n}} &= 0
\end{aligned} \tag{3.1}$$

This can be generalised to any d -dimensional space with d -dimensional normal $\hat{\mathbf{n}}$, two d -dimensional points \mathbf{p} and \mathbf{r} will satisfy equation 3.1.

3.3 Orthonormal Basis Set

An orthonormal basis set is a set of mutually perpendicular unit vectors, usually denoted \mathbf{e}_i where $i \in \mathbb{N}$, $|\mathbf{e}_i| = 1$ and if $i \neq j$ then $\mathbf{e}_i \cdot \mathbf{e}_j = 0$. A d -dimensional vector \mathbf{a} can be decomposed into components of this basis set

$$\mathbf{a} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \cdots + a_d \mathbf{e}_d$$

The coefficient of \mathbf{e}_i can be found using the dot product

$$\mathbf{a} \cdot \mathbf{e}_i = a_i$$

since this goes to 0 for all terms not including \mathbf{e}_i and $\mathbf{e}_i \cdot \mathbf{e}_i = 1$.

There is more than one orthonormal basis set

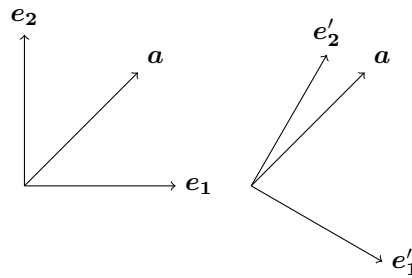


Figure 3.2: Multiple orthonormal basis sets

\mathbf{a} in figure 3.2 can be given as:

$$\mathbf{a} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 = a'_1 \mathbf{e}'_1 + a'_2 \mathbf{e}'_2$$

or in terms of column vectors

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}_B \quad \mathbf{a} = \begin{pmatrix} a'_1 \\ a'_2 \end{pmatrix}_{B'}$$

Where the subscript B or B' tells us the basis we are in. Note that in general

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}_B \neq \begin{pmatrix} a'_1 \\ a'_2 \end{pmatrix}_{B'}$$

as the basis vectors aren't included in this representation in the way that they are in $\mathbf{a} = a_i \mathbf{e}_i$.

3.4 Linear Independence

A set of d vectors spans a d -dimensional space if the vectors are linearly independent.

A set of d vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ is linearly independent if and only if $c_1 = c_2 = \cdots = c_d = 0$ is the only solution to

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_d \mathbf{v}_d = 0$$

For example two co-linear vectors \mathbf{a} and \mathbf{b} aren't linearly independent as $\mathbf{a} = \lambda \mathbf{b}$ so $\mathbf{a} - \lambda \mathbf{b} = 0 \forall \mathbf{a}$ and \mathbf{b} .

A 3-dimensional orthonormal basis set is right handed if

$$\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3 \quad \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1 \quad \mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2$$

4 Orthonormalisation

For two n -dimensional vectors \mathbf{a} and \mathbf{b} we can write them as sums

$$\sum_{i=1}^n a_i \mathbf{e}_i \quad \& \quad \sum_{j=1}^n b_j \mathbf{e}_j$$

We define the Kronecker delta δ as

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

This allows us to write the dot product for two unit vectors \mathbf{e}_i and \mathbf{e}_j as

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$$

We can use this to write the dot product for \mathbf{a} and \mathbf{b} as sums for which we will use the notation \sum_i to indicate that we are summing over all values of i .

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= \left(\sum_i a_i \mathbf{e}_i \right) \cdot \left(\sum_j b_j \mathbf{e}_j \right) \\ &= \sum_i \sum_j a_i b_j \mathbf{e}_i \cdot \mathbf{e}_j \\ &= \sum_i \sum_j \delta_{ij} a_i b_j \end{aligned}$$

Since all terms have a factor of δ_{ij} any term where $i \neq j$ goes to 0 so we can sum over only one index by setting $i = j$

$$= \sum_i a_i b_i$$

4.1 Creating orthonormal basis sets

In 1-dimension if we have a vector \mathbf{a} we can just use $\mathbf{e}_1 = \hat{\mathbf{a}} = \mathbf{a}/a$ as a basis vector.

In 2-dimensions if we have linearly independent vectors \mathbf{a} and \mathbf{b} we start by using $\mathbf{e}_1 = \hat{\mathbf{a}} = \mathbf{a}/a$. We then need to find a second perpendicular unit vector to this. We do this by taking \mathbf{b} and subtracting the component in the direction of \mathbf{a} to construct a vector \mathbf{c} perpendicular to \mathbf{e}_1 .

$$\mathbf{c} = \mathbf{b} - (\mathbf{b} \cdot \hat{\mathbf{a}}) \hat{\mathbf{a}}$$

We can then set our second basis vector to $\mathbf{e}_2 = \hat{\mathbf{c}} = \mathbf{c}/c$, written in terms of \mathbf{a} and \mathbf{b} this is:

$$\mathbf{e}_2 = \frac{\mathbf{b} - (\mathbf{b} \cdot \hat{\mathbf{a}}) \hat{\mathbf{a}}}{|\mathbf{b} - (\mathbf{b} \cdot \hat{\mathbf{a}}) \hat{\mathbf{a}}|}$$

Note that if \mathbf{a} and \mathbf{b} aren't linearly independent then $\vartheta = 0$ and

$$\begin{aligned} \mathbf{c} &= \mathbf{b} - (\mathbf{b} \cdot \hat{\mathbf{a}}) \hat{\mathbf{a}} \\ &= \mathbf{b} - b \cos(\vartheta) \hat{\mathbf{a}} \\ &= \mathbf{b} - b \hat{\mathbf{a}} \end{aligned}$$

Since \mathbf{a} and \mathbf{b} are in the same direction $\mathbf{b} = b \hat{\mathbf{a}}$

$$\begin{aligned} &= \mathbf{b} - \mathbf{b} \\ &= 0 \end{aligned}$$

This process is called Gram-Schmidt orthonormalisation

4.2 Gram-Schmidt Orthonormalisation

We can iterate over this process to find \mathbf{e}_3 if we have some vector \mathbf{d} which is linearly independent to \mathbf{a} and \mathbf{b} .

$$\mathbf{e}_3 = \frac{\mathbf{d} - (\mathbf{d} \cdot \mathbf{e}_1)\mathbf{e}_1 - (\mathbf{d} \cdot \mathbf{e}_2)\mathbf{e}_2}{|\mathbf{d} - (\mathbf{d} \cdot \mathbf{e}_1)\mathbf{e}_1 - (\mathbf{d} \cdot \mathbf{e}_2)\mathbf{e}_2|}$$

For the n^{th} step of this process:

- Have $n - 1$ unit vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}$
- Given an n^{th} vector \mathbf{v}_n , which is linearly independent to these unit vectors, we can find a vector \mathbf{v}'_n :

$$\mathbf{v}'_n = \mathbf{v}_n - (\mathbf{v}_n \cdot \mathbf{e}_1)\mathbf{e}_1 - (\mathbf{v}_n \cdot \mathbf{e}_2)\mathbf{e}_2 - \dots - (\mathbf{v}_n \cdot \mathbf{e}_{n-1})\mathbf{e}_{n-1}$$

- Normalise this vector to get

$$\mathbf{e}_n = \frac{\mathbf{v}'_n}{|\mathbf{v}'_n|}$$

You run out of vectors when \mathbf{v}_n can be written as $\mathbf{v}_n = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots a_{n-1}\mathbf{v}_{n-1}$

4.3 Triple Products

The area of a parallelogram with adjacent sides \mathbf{a} and \mathbf{b} is given by $|\mathbf{a} \times \mathbf{b}|$.

The volume of a parallelepiped constructed from the above parallelogram with a third side \mathbf{c} is $|(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}|$. This is called the scalar triple product.

The vector triple product is

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$$

Example 4.1

An associative operator \circ satisfies $A \circ (B \circ C) = (A \circ B) \circ C$. Is the cross product \times associative?

We can show that it isn't using the vector triple product. First we use the anti-commutativity of the cross product

$$\begin{aligned} (\mathbf{a} \times \mathbf{b}) \times \mathbf{c} &= -\mathbf{c} \times (\mathbf{a} \times \mathbf{b}) \\ &= \mathbf{a}(\mathbf{c} \cdot \mathbf{b}) - \mathbf{b}(\mathbf{c} \cdot \mathbf{a}) \end{aligned} \tag{4.1}$$

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}) \tag{4.2}$$

These are clearly different as equation 4.1 has components in the \mathbf{a} and \mathbf{b} directions where as equation 4.2 has components in the \mathbf{b} and \mathbf{c} directions.

4.4 Levi-Civita Symbol

The Levi-Civita symbol ε is defined as

$$\varepsilon_{ijk} = \begin{cases} +1 & ijk = 123, 312, 231 \\ -1 & ijk = 132, 321, 213 \\ 0 & i = j, j = k, i = k \end{cases}$$

For example $\varepsilon_{312} = 1$, $\varepsilon_{213} = -1$ and $\varepsilon_{121} = 0$.

In more dimensions $\varepsilon_{i_1 i_2 \dots i_n}$ is defined as 1 if $i_1 i_2 \dots i_n$ is an even permutation (requires an even number of two element swaps to return to $12 \dots n$), -1 if it is an odd permutation (requires an odd number of swaps to return to $12 \dots n$) or 0 if there are repeated indices.

We can use this to write the cross product in a more concise way

$$(\mathbf{a} \times \mathbf{b})_i = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{e}_i = \sum_{j=1}^3 \sum_{k=1}^3 \varepsilon_{ijk} a_j b_k$$

$$\mathbf{a} \times \mathbf{b} = \sum_i \sum_j \sum_k \varepsilon_{ijk} \mathbf{e}_i a_j b_k$$

5 Matrices

5.1 More on the Scalar Triple Product

If \mathbf{a} , \mathbf{b} and \mathbf{c} aren't linearly independent then their scalar triple product $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0$. This means that the scalar triple product can be used to test for linear independence.

Consider the parity transform $\mathbf{e}'_i = -\mathbf{e}_i$. This changes the parity of the coordinate system.

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = -\mathbf{a} \cdot (-\mathbf{b} \times -\mathbf{c}) = -\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$$

So the sign of the scalar triple product changes under a parity change. For this reason we call the scalar triple product a pseudo-scalar

5.2 Matrices

Consider a 2-dimensional vector \mathbf{a} . It can be written in multiple different ways:

$$\mathbf{a} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2$$

$$\mathbf{a} = a'_1 \mathbf{e}'_1 + a'_2 \mathbf{e}'_2$$

Where $\{\mathbf{e}_1, \mathbf{e}_2\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2\}$ are orthonormal basis sets. Our aim is to find the components a'_i . In general we can find the \mathbf{e}_i component by dotting with \mathbf{e}_i :

$$\mathbf{a} \cdot \mathbf{e}_i = a'_1 (\mathbf{e}_i \cdot \mathbf{e}'_1) + a'_2 (\mathbf{e}_i \cdot \mathbf{e}'_2)$$

For an n -dimensional vector \mathbf{a}

$$a'_i = \sum_{j=1}^n a_j (\mathbf{e}_i \cdot \mathbf{e}'_j)$$

a'_i is a weighted sum of the components in the original basis set. We can write this transform as

$$\begin{pmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_n \end{pmatrix} = P \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

P is an $n \times n$ array of numbers and specifies the weights. P_{ij} is the element of P at row i and column j . We can write P out as an array of dot products

$$\begin{pmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_n \end{pmatrix} = \begin{pmatrix} \mathbf{e}'_1 \cdot \mathbf{e}_1 & \mathbf{e}'_1 \cdot \mathbf{e}_2 & \cdots & \mathbf{e}'_1 \cdot \mathbf{e}_n \\ \mathbf{e}'_2 \cdot \mathbf{e}_1 & \mathbf{e}'_2 \cdot \mathbf{e}_2 & \cdots & \mathbf{e}'_2 \cdot \mathbf{e}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{e}'_n \cdot \mathbf{e}_1 & \mathbf{e}'_n \cdot \mathbf{e}_2 & \cdots & \mathbf{e}'_n \cdot \mathbf{e}_n \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

A matrix provides a way to define a linear transformation. It doesn't have a physical meaning itself.

For matrices A and B the matrix product AB is defined if and only if the number of columns in A is the same as the number of rows in B . The element AB_{ij} is the dot product of row i of A and column j of B . Doing transformation A and then B is equivalent to doing the transformation BA . In general $AB \neq BA$.

6 Orthogonal Matrices

The identity matrix $\mathbb{1}$ is the matrix $[\mathbb{1}]_{ij} = \delta_{ij}$

The inverse of matrix M is defined as the matrix M^{-1} such that

$$MM^{-1} = M^{-1}M = \mathbb{1}$$

The transpose of a matrix M is defined as the matrix M^T that results from swapping the rows and columns of M

$$[M^T]_{ij} = M_{ji}$$

A passive transformation is one where the vector that it acts on is unaffected but the basis changes.

An example of this is a rotation, anticlockwise, about the origin, by an angle of ϑ as shown in figure 6.1 By

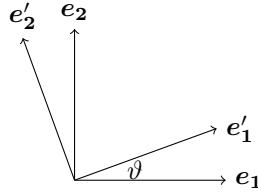


Figure 6.1: Rotation of basis

projecting the the new basis vectors onto the old ones we can calculate the transformation matrix $R(\vartheta)$:

$$P = \begin{pmatrix} \mathbf{e}'_1 \cdot \mathbf{e}_1 & \mathbf{e}'_1 \cdot \mathbf{e}_2 \\ \mathbf{e}'_2 \cdot \mathbf{e}_1 & \mathbf{e}'_2 \cdot \mathbf{e}_2 \end{pmatrix} = \begin{pmatrix} \cos \vartheta & \sin \vartheta \\ -\sin \vartheta & \cos \vartheta \end{pmatrix}$$

The inverse of a rotation is a rotation by the same amount in the opposite direction so in this case $R(-\vartheta)$

$$R^{-1}(\vartheta) = \begin{pmatrix} \cos -\vartheta & \sin -\vartheta \\ -\sin -\vartheta & \cos -\vartheta \end{pmatrix} = \begin{pmatrix} \cos \vartheta & -\sin \vartheta \\ \sin \vartheta & \cos \vartheta \end{pmatrix}$$

From this we can see that

$$R^{-1}(\vartheta) = R^T(\vartheta)$$

If $O^T O = \mathbb{1}$ then we say that the matrix O is orthogonal. Equivalent definitions are:

1. A matrix O is orthogonal if and only if $O^T = O^{-1}$
2. A matrix O is orthogonal if and only if its rows make up an orthonormal basis set
3. A matrix O is orthogonal if and only if its columns make up an orthonormal basis set
4. A matrix O is orthogonal if and only if it describes a transformation from one orthonormal basis set to another

6.1 Proof of Equivalence

Claim - A matrix O is orthogonal if and only if $O^T = O^{-1}$

Proof - Starting with the condition of a matrix being orthogonal if $O^T O = \mathbb{1}$

$$O^T O = \mathbb{1} \iff O^T O O^{-1} = \mathbb{1} O^{-1} \iff O^T = O^{-1}$$

Claim - A matrix O is orthogonal if and only if its rows or columns make up an orthonormal basis set

Proof - A matrix has a set of orthonormal basis vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ as its rows. Its transpose therefore has this set of vectors as its columns. This means that the product of these two matrices is $[\mathbf{v}_i \cdot \mathbf{v}_j]_{ij}$. Since these are orthonormal basis vectors this means $[\mathbf{v}_i \cdot \mathbf{v}_j]_{ij} = \delta_{ij} = \mathbb{1}$

7 Active Transforms

A passive (aka alias) transformation expresses how the coordinates must change to keep a vector \mathbf{a} unchanged in the new basis.

An active (aka alibi) transformation fixes the basis and changes all other vectors.

$$\mathbf{a} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2$$

$$\mathbf{a}' = \tilde{a}_1 \mathbf{e}_1 + \tilde{a}_2 \mathbf{e}_2$$

$\mathbf{a} \neq \mathbf{a}'$ however the set of basis vectors for both is still $\{\mathbf{e}_i\}$. How do we find \tilde{a}_i ? We start with the dot product with \mathbf{e}_i

$$\tilde{a}_i = \mathbf{a}' \cdot \mathbf{e}_i$$

$$\mathbf{a}' = \begin{pmatrix} \tilde{a}_1 \\ \tilde{a}_2 \end{pmatrix} = A \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 \cdot \mathbf{a}' \\ \mathbf{e}_2 \cdot \mathbf{a}' \end{pmatrix}$$

$$\mathbf{a}' = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 = \tilde{a}_1 \mathbf{e}_1 + \tilde{a}_2 \mathbf{e}_2$$

Where $\{\mathbf{v}_i\}$ is the set that $\{\mathbf{e}_i\}$ transforms to under A . We can use the dot product with \mathbf{a}' written in terms of $\{\mathbf{v}_i\}$ to calculate A

$$\begin{pmatrix} \tilde{a}_1 \\ \tilde{a}_2 \end{pmatrix} = \begin{pmatrix} a_1 \mathbf{v}_1 \cdot \mathbf{e}_1 + a_2 \mathbf{v}_2 \cdot \mathbf{e}_1 \\ a_1 \mathbf{v}_1 \cdot \mathbf{e}_2 + a_2 \mathbf{v}_2 \cdot \mathbf{e}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 \cdot \mathbf{v}_1 & \mathbf{e}_1 \cdot \mathbf{v}_2 \\ \mathbf{e}_2 \cdot \mathbf{v}_1 & \mathbf{e}_2 \cdot \mathbf{v}_2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

7.1 Linearity

\mathbf{a} and \mathbf{b} are vectors, $\mathbf{a} + \mathbf{b} = \mathbf{c}$. Under a certain transformation $\mathbf{a} \rightarrow \mathbf{a}'$ and $\mathbf{b} \rightarrow \mathbf{b}'$. If under that same transformation $(\mathbf{a} + \mathbf{b})' = \mathbf{c}' = \mathbf{a}' + \mathbf{b}'$ then this is a linear transformation. If this transformation is encoded in matrix A then $\mathbf{a}' = A_{ij}a_j$ and $\mathbf{b}' = A_{ij}b_j$.

$$\mathbf{c}' = (\mathbf{a} + \mathbf{b})' = A_{ij}(a_i + b_i) = A_{ij}a_j + A_{ij}b_j = \mathbf{a}' + \mathbf{b}'$$

so A is a linear transformation.

$$\mathbf{a}' = A\mathbf{a} = \sum_i a_i \mathbf{v}_i = \sum_i \tilde{a}_i \mathbf{e}_i$$

$$\mathbf{a}' = A\mathbf{a} \implies A^{-1}\mathbf{a}' = \mathbf{a}$$

A^{-1} is a passive transformation from an orthonormal basis set to a non-orthonormal basis set. This only works if the transformation is orthogonal (ie $P = A^T = A^{-1}$).

8 Eigenvectors and Eigenvalues

8.1 Symmetric and Diagonal Matrices

For this section we will need some new types of matrix:

- A symmetric matrix S is one where $S = S^T$ or equivalently $S_{ij} = S_{ji}$. This means that S is necessarily square.
- A diagonal matrix D is one where for $i \neq j$ $D_{ij} = 0$.

8.2 Eigenvectors and Eigenvalues definition

Let S be an $n \times n$ symmetric matrix. Then S has at most n eigenvalues λ_i where $i \in \{1, \dots, n\}$ and n eigenvectors \mathbf{v}_i such that

$$S\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

$\{\hat{\mathbf{v}}_i\}$ forms an orthonormal basis set.

$$\mathbf{a} = a_1 \hat{\mathbf{v}}_1 + a_2 \hat{\mathbf{v}}_2 + \dots + a_n \hat{\mathbf{v}}_n$$

$$\mathbf{a}' = S\mathbf{a} = a_1 S\hat{\mathbf{v}}_1 + a_2 S\hat{\mathbf{v}}_2 + \dots + a_n S\hat{\mathbf{v}}_n = a_1 \lambda_1 \hat{\mathbf{v}}_1 + a_2 \lambda_2 \hat{\mathbf{v}}_2 + \dots + a_n \lambda_n \hat{\mathbf{v}}_n$$

$$\mathbf{a}^{(m)} = S^m \mathbf{a} = a_1 \lambda_1^m \hat{\mathbf{v}}_1 + a_2 \lambda_2^m \hat{\mathbf{v}}_2 + \dots + a_n \lambda_n^m \hat{\mathbf{v}}_n$$

Every symmetric matrix S can be written as $S = O^T D O$ where O is orthogonal and has S 's normalised eigenvectors as rows and D is diagonal and has S 's eigenvalues along its leading diagonal.

To show that S has only n eigenvalues/vectors we consider the element S_{ij} which is $\mathbf{v}_i^T S \mathbf{v}_j$. By doing this calculation in two different orders we get two equivalent answers:

$$\begin{aligned} \mathbf{v}_i^T S \mathbf{v}_j &= \mathbf{v}_i^T \lambda_j \mathbf{v}_j = \lambda_j \mathbf{v}_i^T \mathbf{v}_j = \lambda_j \mathbf{v}_i \cdot \mathbf{v}_j \\ \mathbf{v}_i^T S \mathbf{v}_j &= (S^T \mathbf{v}_i)^T \mathbf{v}_j = (S \mathbf{v}_i)^T \mathbf{v}_j = (\lambda_i \mathbf{v}_i)^T \mathbf{v}_j = \lambda_i \mathbf{v}_i^T \mathbf{v}_j = \lambda_i \mathbf{v}_i \cdot \mathbf{v}_j \\ &\implies \lambda_i \mathbf{v}_i \cdot \mathbf{v}_j = \lambda_j \mathbf{v}_i \cdot \mathbf{v}_j \\ &\implies (\lambda_i - \lambda_j) \mathbf{v}_i \cdot \mathbf{v}_j = 0 \end{aligned}$$

Therefore $\lambda_i = \lambda_j$ or $\mathbf{v}_i \cdot \mathbf{v}_j = 0$. If the later is the case then \mathbf{v}_i and \mathbf{v}_j must be perpendicular. Therefore for two distinct eigenvalues their corresponding eigenvectors must be orthogonal. Since there is no linearly independent set larger than n that spans an n -dimensional space there can only be at most n eigenvalue/vector pairs.

9 Determinants

9.1 The Eigenvalue Equation

S is a symmetric matrix. \mathbf{v} is one of its eigenvectors and has components c_i

$$\begin{aligned} S\mathbf{v} &= \lambda \mathbf{v} \\ S\mathbf{v} - \lambda \mathbf{v} &= \mathbf{0} \\ S\mathbf{v} - \lambda \mathbf{1} \mathbf{v} &= \mathbf{0} \\ (S - \lambda \mathbf{1})\mathbf{v} &= \mathbf{0} \end{aligned} \tag{9.1}$$

Let $S - \lambda \mathbf{1} = S'$

$$\begin{aligned} \begin{pmatrix} S_{11} - \lambda & S_{12} & \dots & S_{1n} \\ S_{21} & S_{22} - \lambda & \dots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \dots & S_{nn} - \lambda \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0} \\ \implies \sum_i c_i u_i &= 0 \end{aligned}$$

where u_i is a row of S' . $\mathbf{v} = \mathbf{0}$ is the trivial solution. If we have a non-trivial solution the $\{u_i\}$ is a linearly dependent set. The procedure to find values of λ and the eigenvectors is

1. Find λ such that the rows of $S'(\lambda)$ are linearly dependent.
2. λ is a eigenvalue
3. Put λ in equation 9.1 and solve for \mathbf{v} .
4. normalise \mathbf{v} to get the eigenvectors

9.2 Determinants as Volume Scale Factor

If we have a unit volume before an active transformation A then after A is applied the volume is now $V(A)$. All volumes are scaled by a factor of $V(A)$.

$$V(A) = |\det(A)|$$

since $V(A)$ is just a scale factor

$$V(AB) = V(A)V(B) = V(B)V(A) = V(BA)$$

We want linearly independent values as solutions to the eigenvector equation so we look for transformations where $V(A) = 0$ as this means that two or more rows/columns must be linearly dependent. Since we want $V(A) = 0$ we need to find cases where $\det(A) = 0$.

To find λ we solve

$$\det(S - \lambda \mathbf{1}) = 0$$

9.3 Permutation Method

A permutation P on the set $N = \{1, \dots, n\}$ for some $n \in \mathbb{N}$ is defined here as an ordered set $\{x | x \in N\}$ with no repetition allowed. We define $P(i)$ as the i^{th} element of P . The sign of P , denoted $\text{sign}(P)$ is defined to be

$$\text{sign}(P) = \varepsilon_{P(1)P(2)\dots P(n)}$$

The determinant of an $n \times n$ matrix A is then

$$\det(A) = \sum_{P \in S_N} \text{sign}(P) A_{1P(1)} A_{2P(2)} \cdots A_{nP(n)}$$

where S_N is the set of all possible permutations, P , of the set N .

9.4 Recursive Method

The minor $M^{(i,j)}$ of matrix A is defined as the determinant of the matrix that results from removing row i and column j from A . For an $n \times n$ matrix the determinant can then be calculated as

$$\det(A) = \sum_{j=1}^n A_{ij} M^{(i,j)}$$

for any $i \in \{1, 2, \dots, n\}$. Or equivalently

$$\det(A) = \sum_{i=1}^n A_{ij} M^{(i,j)}$$

for any $j \in \{1, 2, \dots, n\}$. This does require use of the fact that the determinant of a 1×1 matrix A is A_{11} which can be shown using the permutation method. For an $n \times n$ matrix A_n for the first three values of n the determinants are:

$$\det(A_1) = A_{11}$$

$$\det(A_2) = A_{11}A_{22} - A_{12}A_{21}$$

$$\det(A_3) = A_{11}(A_{22}A_{33} - A_{23}A_{32}) - A_{12}(A_{21}A_{33} - A_{31}A_{23}) + A_{13}(A_{21}A_{32} - A_{22}A_{31})$$

10 The Eigenvalue Equation

10.1 Useful Determinant Properties

It can clearly be seen that the determinant of an $n \times n$ diagonal matrix D is the product of its diagonal elements

$$\det(D) = \prod_{i=1}^n D_{ii}$$

This can be seen from the fact that $D_{1i} = 0$ if $i \neq 1$ so only the first term of the determinant need be considered and once the first row and column are removed the same logic applies as it is still a diagonal matrix.

Likewise an $n \times n$ symmetric matrix S with eigenvalues λ_i has determinant

$$\det(S) = \prod_{i=1}^n \lambda_i$$

This follows from the fact that $\det(AB) = \det(BA)$:

$$\det(S) = \det(O^T D O) = \det(O O^T D) = \det(D) = \prod_{i=1}^n \lambda_i$$

$\det(\mathbb{1}) = 1$, this follows from the fact that $\mathbb{1}$ is a diagonal matrix with all non-zero elements equal to 1 so its determinant is

$$\det(\mathbb{1}) = \prod_{i=1}^n 1 = 1$$

For an $n \times n$ matrix A and scalar c

$$\det(cA) = c^n \det(A)$$

This is clear when looking at the permutation definition of a determinant as it has n factors of A_{1i} in each term of the sum so multiplying by c leads to a multiple of c^n in each term.

$$\det(A) = \frac{1}{\det(A^{-1})} \iff \det(A) \det(A^{-1}) = 1 \iff \det(AA^{-1}) = 1 \iff \det(\mathbb{1}) = 1$$

$\det(A) = 0$ if the rows or columns of A form a linearly dependent set. This follows from the fact that this means that their are collinear vectors and hence the volume is 0.

10.2 Trace of a Matrix

The trace of an $n \times n$ matrix A is denoted as $\text{Tr}(A)$. It is defined as the sum of the values on the leading diagonal

$$\text{Tr}(A) = \sum_{i=1}^n A_{ii}$$

Since the trace only uses the diagonal values of the matrix it is clear that

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

The trace of a matrix is also cyclic, that is that for a matrix product $A = A_1 A_2 \cdots A_n$ the trace $\text{Tr}(A)$ is the same as for any other cyclic permutation of A_i . For example:

$$\text{Tr}(ABC) = \text{Tr}(BCA) = \text{Tr}(CAB) \neq \text{Tr}(ACB) = \text{Tr}(BAC) = \text{Tr}(CBA)$$

The trace of an $n \times n$ symmetric matrix S with eigenvalues λ_i is

$$\text{Tr}(S) = \text{Tr}(O^T D O) = \text{Tr}(O O^T D) = \text{Tr}(D) = \sum_{i=1}^n \lambda_i$$

10.3 Solving the Eigenvalue Equation

The characteristic polynomial of a symmetric matrix S is $\det(S - \lambda \mathbf{1}) = 0$. The roots of this polynomial are the eigenvalues λ_i of S . To find the eigenvectors \mathbf{v}_i we solve

$$S\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

for a generic

$$\mathbf{v}_i = v_i^1 \mathbf{e}_1 + v_i^2 \mathbf{e}_2 + \cdots + v_i^n \mathbf{e}_n = \sum_{j=1}^n v_i^j \mathbf{e}_j$$

Since there are linearly dependent vectors in $S - \lambda \mathbf{1}$ (as the determinant is 0) this gives a system of equations with infinitely many solutions. The best that we can do from this is to find the ratio of the components. To find a specific eigenvector we impose the condition that the eigenvector has magnitude 1. This allows us to calculate a specific eigenvector since only the direction is important not the length.

10.4 Coupled Equations

When values in one equation depend on another equation which in turn depends on the first we say that the equations are coupled. It is possible to de-couple equations using matrices.

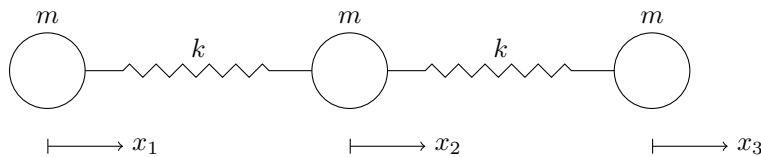


Figure 10.1: Spring mass system

The spring mass system in figure 10.1 is an example of one such coupled system. x_i is the displacement of particle i from its equilibrium position. The extension of the first spring is $x_1 - x_2$. From this we can see that the force f_1 on the first particle is

$$f_1 = -k(x_1 - x_2)$$

The same logic works to calculate the force f_3 on the third particle

$$f_3 = -k(x_3 - x_2)$$

Using Newton's third law we can then calculate the force f_2 on the middle particle

$$f_2 = -(f_1 + f_3) = k(x_1 - 2x_2 + x_3)$$

We can then apply Newton's second law to any of the particles to get

$$m \frac{d^2 x_i}{dt^2} = f_i$$

However since f_i and x_i are dependent on the other two particles this is a coupled system. We rewrite this in terms of matrices and vectors

$$m \frac{d^2 \mathbf{x}}{dt^2} = k S \mathbf{x}(t)$$

$$m \frac{d^2 \mathbf{x}}{dt^2} = k \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}$$

11 Linear Systems

Continuing from the last section we can write $\mathbf{x}(t)$ as

$$\mathbf{x}(t) = \sum_{i=1}^3 a_i(t) \mathbf{v}_i$$

where \mathbf{v}_i is an eigenvector of S . This gives us

$$m \frac{d^2}{dt^2} \mathbf{x}(t) = m \sum_{i=1}^3 \frac{d^2 a_i(t)}{dt^2} \mathbf{v}_i$$

$$kS\mathbf{x}(t) = k \sum_{i=1}^3 \lambda_i a_i(t) \mathbf{v}_i$$

By dotting both of these with \mathbf{v}_i we get

$$\ddot{a}_i(t) = \gamma_i a_i(t)$$

where $\gamma_i \neq 0$. We know that for SHM

$$\frac{d^2 x}{dt^2} = -\omega^2 x$$

$$x(t) = A \cos \omega t + B \sin \omega t$$

Starting from rest means that

$$\dot{a}_i(0) = 0$$

We can calculate $\dot{x}(t)$ as

$$\dot{x}(0) = -\omega A \sin 0 + \omega B \cos 0 = 0 \implies B = 0$$

From this we get

$$\dot{a}_i(t) = a_i(0) \cos \left(\sqrt{-\frac{k\lambda_i}{m}} t \right)$$

We now need to find the eigenvalues of S . To do this we solve the characteristic polynomial

$$\det(S - \lambda \mathbf{1}) = 0$$

This gives us $\lambda = -3, -1, 0$ Finally we get

$$\mathbf{x}(t) = \sum_{i=1}^3 a_i(0) \cos \left(\sqrt{-\frac{k\lambda_i}{m}} t \right)$$

11.1 Linear System of Equations

We can write a set of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots & \quad \quad \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

as one matrix equation

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

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \quad (11.1)$$

This means that the solution to the equations is given by

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

so we just need to be able to find A^{-1}

12 Inverses

Every linear equation corresponds to a plane. Solutions to the system lie where the planes intersect. This means that adding a new equation (and hence plane) can have one of three effects:

1. Same plane as one already in the system. This means there are infinite solutions.
2. Plane is parallel to one already in the system. This means there are no solutions.
3. The plane intersects all of the others at a line reducing the dimensions of the problem.

One of the first two options occurs if and only if the equations aren't linearly independent. We need a method to find the number of linearly independent equations. If all rows of a matrix A are linearly independent then $\det(A) \neq 0$. We need to find the dimension of the largest square sub-matrix A' such that $\det(A') \neq 0$. The dimension of A' is the known as the rank of A .

12.1 Rouchè-Capelli Theorem

We need to define the augmented matrix. For equation 11.1 the augmented equation is denoted $(A|b)$ and is

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

The Rouchè-Capelli theorem (also known as the Kronecker-Capelli theorem) is:

A system of linear equations with n unknowns has at least one solution if and only if the rank of its coefficient matrix A is equal to the rank of its augmented matrix $(A|b)$. If the rank of A is n then there is a unique solution, otherwise there are infinitely many solutions.

12.2 Cramer's Rule

Cramer's rule is a method for finding the inverse of a matrix where it exists. For the inverse of matrix A to exist we need $\det(A) \neq 0$. We can then use Cramer's rule to find each element of the inverse

$$[A^{-1}]_{ij} = \frac{(-1)^{i+j} M^{(j,i)}}{\det(A)}$$

where $M^{(a,b)}$ is the minor of A , that is the determinant of the matrix that results from removing row a and column b .

12.3 Gaussian Elimination

Calculating a determinant of an $n \times n$ matrix has time complexity $O(n!)$. This is not good. We need a faster method to find the inverses without needing to find the inverse. One such method is Gaussian elimination. In this we do elementary row operations (linear combinations of rows) to an augmented matrix such that we get every value below the diagonal to be 0. This means that if we expand the determinant by the first column only the first term needs to be considered as all others have a factor of A_{i1} which is 0 when $i \neq 1$.

12.4 Complex Vectors and Matrices

Vectors and matrices can have complex elements. For these to be useful we need to define some new properties. We define the complex conjugate of a matrix M as

$$[M^*]_{ij} = [M]_{ij}^*$$

that is every element is replaced with its complex conjugate. This means that for a real matrix R $R^* = R$. We denote the Hermitian conjugate (also known as the hermitian transpose) of matrix M as M^\dagger and define it as

$$[M^\dagger]_{ij} = [M^*]_{ji} = [M^T]_{ij}^*$$

that is the complex conjugate of the transpose (or transpose of the complex conjugate). This means that for a real matrix R $R^\dagger = R^T$. If a matrix H has the property that

$$H = H^\dagger$$

then we say H is Hermitian, this is analogous to a real matrix being symmetric. All symmetric matrices are Hermitian. The scalar product of two, n dimensional complex vectors \mathbf{u} and \mathbf{v} is

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\dagger \mathbf{v} = \sum_{i=0}^n u_i^* v_i$$

this is analogous to the dot product for real, n dimensional vectors \mathbf{a} and \mathbf{b} being

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b} = \sum_{i=0}^n a_i b_i$$

If a matrix U has the property that

$$U^\dagger U = \mathbb{1}$$

then we say U is a unitary matrix, this is analogous to a real matrix being orthogonal. All orthogonal matrices are also unitary.

12.5 Dirac Notation

This notation is also known as bra-ket notation. A vector \mathbf{v} in this notation is written as a ket: $|v\rangle$ and its hermitian transpose \mathbf{v}^\dagger is written as a bra: $\langle v|$. Using this notation the scalar product of \mathbf{u} and \mathbf{v} can be written as:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\dagger \mathbf{v} = \langle u | v \rangle$$

The matrix multiplication $\mathbf{u}^\dagger M \mathbf{v}$ can be written as

$$\mathbf{u}^\dagger M \mathbf{v} = \langle u | M | v \rangle$$

Quantum mechanical state can often be represented by a complex wave function ψ , this can be represented by a complex vector $|\psi\rangle$ as both functions and vectors are members of a vector space. If this space has an infinite number of dimensions it is known as a Hilbert space.

13 Partial Derivative

13.1 Last Notes on Linear Algebra

Any Hermitian matrix H can be written as

$$H = U^\dagger D U$$

where U is a unitary matrix and D is a real diagonal matrix with $D_{ii} = \lambda_i$ where λ_i is an eigenvalue of H .

For complex vectors \mathbf{a} and \mathbf{b} the dot product is no longer commutative

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^\dagger \mathbf{b} = (\mathbf{b}^\dagger \mathbf{a})^\dagger = (\mathbf{b}^\dagger \mathbf{a})^* = (\mathbf{b} \cdot \mathbf{a})^*$$

Some consequences of this are, for complex vectors \mathbf{a} , \mathbf{b} and \mathbf{c} and scalars $z_i \in \mathbb{C}$:

$$(\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c} \neq \mathbf{c} \cdot \mathbf{a} + \mathbf{c} \cdot \mathbf{b} = \mathbf{c} \cdot (\mathbf{a} + \mathbf{b})$$

$$(z_1 \mathbf{a}) \cdot (z_2 \mathbf{b}) = z_1^* z_2 \mathbf{a} \cdot \mathbf{b}$$

13.2 Linear Vector Spaces

A vector space V is a set of objects that are known as vectors (note that non-vector objects can still be members of a vector space and in this case are still thought of as vectors) and have the following properties:

- *Closure under addition* - for $\mathbf{a}, \mathbf{b} \in V$ $\mathbf{a} + \mathbf{b} \in V$
- *Commutativity of addition* - for $\mathbf{a}, \mathbf{b} \in V$ $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$
- *Associativity of addition* - for $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$ $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$
- *Closure under scalar multiplication* - for a scalar λ and $\mathbf{a} \in V$ $\lambda \mathbf{a} \in V$
- *Distributivity of scalar multiplication over vector addition* - for a scalar λ and $\mathbf{a}, \mathbf{b} \in V$ $\lambda(\mathbf{a} + \mathbf{b}) = \lambda \mathbf{a} + \lambda \mathbf{b}$
- *Distributivity of scalar multiplication over scalar addition* - for scalars λ and μ , and $\mathbf{a} \in V$ $(\lambda + \mu)\mathbf{a} = \lambda \mathbf{a} + \mu \mathbf{a}$
- *Associativity of scalar multiplication* - for scalars λ and μ , and $\mathbf{a} \in V$ $\lambda(\mu \mathbf{a}) = (\lambda \mu)\mathbf{a}$
- *Existence of additive identity* - $\exists \mathbf{0}$ such that $\mathbf{a} + \mathbf{0} = \mathbf{a} \quad \forall \mathbf{a} \in V$
- *Existence of additive inverse* - $\exists (-\mathbf{a})$ such that $\mathbf{a} + (-\mathbf{a}) = \mathbf{0} \quad \forall \mathbf{a} \in V$
- *Existence of scalar multiplication identity* - there exists a scalar 1 such that $1\mathbf{a} = \mathbf{a} \quad \forall \mathbf{a} \in V$

It is trivial to show that all of these properties hold for functions, this means that we can apply other vector ideas to functions. A particularly powerful one is that a function f on variable x can be written as a sum of basis functions f_i :

$$f(x) = \sum_i a_i f_i(x)$$

for some scalars a_i . This idea is the basis of Fourier analysis. The coefficients a_i can be obtained the same way as for vectors, through a scalar product, which we define as:

$$\langle f, g \rangle = \int f^*(x) g(x) dx$$

where f and g are functions of x and $*$ denotes the complex conjugate. As with vectors we can have an orthonormal basis set which satisfies $\langle f_i, f_j \rangle = \delta_{ij}$ this allows us to find a_i :

$$a_i = \langle f_i, f \rangle$$

An analogue for a transformation can also be found for functions. This analogue is an operator which takes a function and returns a new function. An operator \mathcal{L} acts on a function f to give

$$\mathcal{L}f(x) = g(x)$$

Like transformations we have an idea of a eigenvalue λ where

$$\mathcal{L}f(x) = \lambda f(x)$$

here f is called an eigenfunction of \mathcal{L} . One example is $\mathcal{L} = d/dx$:

$$\mathcal{L}f(x) = \frac{d}{dx}f(x)$$

$$f(x) = e^{\lambda x}$$

$$\mathcal{L}f(x) = \frac{d}{dx}e^{\lambda x} = \lambda e^{\lambda x} = \lambda f(x)$$

13.3 Single Variable Derivative

A function f of a single variable x has a gradient at point x given by

$$\lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon) - f(x)}{\varepsilon}$$

this is its right derivative. The left derivative is

$$\lim_{\varepsilon \rightarrow 0} \frac{f(x) - f(x - \varepsilon)}{\varepsilon}$$

If both the right and left derivative exist and are equal then the total derivative exists and f is called differentiable at that point x . The total derivative is

$$\frac{df}{dx} = \lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon) - f(x)}{\varepsilon}$$

13.4 Derivatives of Multiple Variables

We can consider what happens to a function f of two variables at a point (x, y) and what the difference is from a nearby point $(x + \varepsilon, y + \delta)$. We can calculate the difference in the function at these two points and divide by the distance between the points. We get

$$\frac{f(x + \varepsilon, y + \delta) - f(x, y)}{\sqrt{\varepsilon^2 + \delta^2}} \quad (13.1)$$

The problem is that the value of this depends on which direction we take the small displacement (ε, δ) . If we take it along a contour line then this value will be 0 as contour lines are by definition lines, where $f(a, b)$ is constant for a point (a, b) on the line. If we take the displacement in any other direction then equation 13.1 will be non-zero. This is a problem. We can define the derivative along any path that approaches (x, y) in terms of the derivatives along two non-parallel paths that go through (x, y) . The most obvious choice for these paths is parallel to the x and y axes. If f was a function of n variables we would need n derivatives along linearly independent paths to define the total derivative. The derivative along one of these paths parallel to an axis is called a partial derivative.

A function f of variables x and y has partial derivatives with respect to x and y , these are respectively:

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

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

These are calculated in practice by considering all variable that we aren't differentiating with respect to as constants and then differentiating as normal. For this reason if we were to try to find the partial derivative of f with respect to x then this is the same as defining a function g of variable x such that by setting $y = y^*$ for some constant y^* we get $g(x) = f(x, y^*)$. The partial derivative is then

$$\frac{\partial f(x, y)}{\partial x} = \frac{\partial f(x, y^*)}{\partial x} = \frac{dg(x)}{dx}$$

It is common to use subscripts to denote the variable being held constant. All of the below are equivalent notations:

$$\frac{\partial f}{\partial x} = \frac{\partial f(x, y)}{\partial x} = \left(\frac{\partial f}{\partial x} \right)_y = \left(\frac{\partial f(x, y)}{\partial x} \right)_y$$

We can extend the idea of a partial derivative to a function f of n variables x_i . We define the partial derivative as

$$\frac{\partial f(x_1, \dots, x_i, \dots, x_n)}{\partial x_i} = \lim_{\varepsilon \rightarrow 0} \frac{f(x_1, \dots, x_i + \varepsilon, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{\varepsilon}$$

13.5 Product Rule

The product rule for one variable functions f_i is

$$\frac{d}{dx}[f_1(x)f_2(x)] = f_1(x)\frac{df_2(x)}{dx} + \frac{df_1(x)}{dx}f_2(x)$$

and it can be extended to multiple variables by considering again the single variable function g that is the result of setting all but one variable to a constant. For the functions f_1 and f_2 of n variables x_i :

$$\begin{aligned} \frac{\partial}{\partial x_i}[f_1(x_1, \dots, x_n)f_2(x_1, \dots, x_n)] &= \frac{d}{dx_i}[g_1(x_i)g_2(x_i)] \\ &= g_1 \frac{dg_2(x_i)}{dx_i} + \frac{dg_1(x_i)}{dx_i}g_2(x) \\ &= f_1(x_1, \dots, x_n) \frac{\partial f_2(x_1, \dots, x_n)}{\partial x_i} + \frac{df_1(x_1, \dots, x_n)}{dx_i} f_2(x_1, \dots, x_n) \end{aligned}$$

where $i \in \{1, \dots, n\}$.

13.6 Gradient

Now in a horrible abuse of notation and with complete disregard for what a derivative is we will motivate the gradient vector:

We will consider the 2 variable function f with variables x and y . Consider some small displacement ε in the x direction and some small displacement δ in the y direction.

$$\begin{aligned} \frac{\partial f(x, y)}{\partial x} &\underbrace{=}_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon, y) - f(x, y)}{\varepsilon} \\ \implies f(x + \varepsilon, y) &\underbrace{=}_{\varepsilon \rightarrow 0} f(x, y) + \varepsilon \frac{\partial f(x, y)}{\partial x} + o(\varepsilon) \end{aligned}$$

where o is a correction function that approaches 0 faster than ε does, that is $\lim_{\varepsilon \rightarrow 0} o(\varepsilon)/\varepsilon = 0$, this is known as little o notation. Likewise

$$f(x, y + \delta) \underbrace{=}_{\delta \rightarrow 0} f(x, y) + \delta \frac{\partial f(x, y)}{\partial y} + o(\delta)$$

14 Gradient Vector

The derivative of a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ along any path through (x, y) can be written as in terms of two derivatives along any two, linearly independent, paths through (x, y) . It is often most useful to use the partial derivatives as they are along lines that are orthogonal and they are simple to calculate. From page 21 section 13.6 we have:

$$f(x + \varepsilon, y) = f(x, y) + \varepsilon \frac{\partial f(x, y)}{\partial x} + o(\varepsilon) \quad (14.1)$$

$$f(x, y + \delta) = f(x, y) + \delta \frac{\partial f(x, y)}{\partial y} + o(\delta) \quad (14.2)$$

For a function g we can also show that

$$\lim_{\varepsilon \rightarrow 0} g(\varepsilon) o(\varepsilon) = 0$$

if $g(\varepsilon)$ converges as $\varepsilon \rightarrow 0$. If we take equation 14.2 we can replace x with $x + \varepsilon$ since equation 14.2 is true(ish) $\forall x$. This gives us

$$f(x + \varepsilon, y + \delta) = f(x + \varepsilon, y) + \delta \frac{\partial f(x + \varepsilon, y)}{\partial y} + o(\delta)$$

next we substitute in $f(x + \varepsilon, y)$ from equation 14.1. This gives us

$$f(x + \varepsilon, y + \delta) = f(x, y) + \varepsilon \frac{\partial f(x, y)}{\partial x} + o(\varepsilon) + \delta \frac{\partial}{\partial y} \left[f(x, y) + \varepsilon \frac{\partial f(x, y)}{\partial x} + o(\varepsilon) \right] + o(\delta)$$

$$f(x + \varepsilon, y + \delta) = f(x, y) + \varepsilon \frac{\partial f(x, y)}{\partial x} + o(\varepsilon) + \delta \frac{\partial f(x, y)}{\partial y} + \delta \varepsilon \frac{\partial}{\partial y} \left[\frac{\partial f(x, y)}{\partial x} \right] + \delta \frac{\partial o(\varepsilon)}{\partial y} + o(\delta)$$

$\frac{\partial o(\varepsilon)}{\partial y} = 0$ and the term with a factor of $\delta \varepsilon \approx 0$ also goes to 0 so this gives us

$$f(x + \varepsilon, y + \delta) = f(x, y) + \varepsilon \frac{\partial f(x, y)}{\partial x} + \delta \frac{\partial f(x, y)}{\partial y} + o(\varepsilon) + o(\delta)$$

Rearranging and taking limits as $\varepsilon \rightarrow 0$ and $\delta \rightarrow 0$ we get

$$\lim_{\varepsilon, \delta \rightarrow 0} [f(x + \varepsilon, y + \delta) - f(x, y)] = \lim_{\varepsilon, \delta \rightarrow 0} \left[\varepsilon \frac{\partial f(x, y)}{\partial x} + \delta \frac{\partial f(x, y)}{\partial y} \right]$$

We now divide by the magnitude of the small displacement (ε, δ) to get rate of change of the function in the direction

$$\hat{\mathbf{u}} = \frac{1}{\sqrt{\varepsilon^2 + \delta^2}} \begin{pmatrix} \varepsilon \\ \delta \end{pmatrix}$$

this gives us

$$\lim_{\varepsilon, \delta \rightarrow 0} \frac{f(x + \varepsilon, y + \delta) - f(x, y)}{\sqrt{\varepsilon^2 + \delta^2}} = \lim_{\varepsilon, \delta \rightarrow 0} \frac{\varepsilon \frac{\partial f(x, y)}{\partial x} + \delta \frac{\partial f(x, y)}{\partial y}}{\sqrt{\varepsilon^2 + \delta^2}}$$

We define the directional derivative, denoted D (often with subscript $\hat{\mathbf{u}}$ and/or $f(x, y)$), as

$$D = \hat{\mathbf{u}} \cdot \begin{pmatrix} \frac{\partial f(x, y)}{\partial x} \\ \frac{\partial f(x, y)}{\partial y} \end{pmatrix} = \hat{\mathbf{u}} \cdot \nabla f(x, y)$$

Where $\nabla f(x, y)$ is the gradient vector of f .

The gradient vector ∇f also denoted **grad** f can be generalised, along with the directional derivative to n dimensions as follows:

For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ the gradient vector is a vector valued function $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ whose value at point $\mathbf{p} = (x_1, \dots, x_n)$ is the vector whose components are the partial derivatives of f at each basis in p :

$$\nabla f(p) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

$$[\nabla f]_i = \frac{\partial f}{\partial x_i}$$

One way to think of ∇ is as an n dimensional vector of operators acting on f

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

The directional derivative along the tangent to a contour line is 0 since a contour line is defined as a line where $f(x, y)$ is constant. Therefore if $\hat{\mathbf{u}}$ is a vector along a contour line then

$$D = \hat{\mathbf{u}} \cdot \nabla f = 0$$

this means that ∇f must be perpendicular to $\hat{\mathbf{u}}$ and the contour line. If instead we pick $\hat{\mathbf{u}} = \nabla f / |\nabla f|$ then

$$\hat{\mathbf{u}} \cdot \nabla f = \frac{\nabla f}{|\nabla f|} \cdot \nabla f = \frac{\nabla f \cdot \nabla f}{|\nabla f|} = \frac{|\nabla f|^2}{|\nabla f|} = |\nabla f|$$

This is the maximum value this dot product can take since the vectors are collinear. This means that the largest directional derivative is in the direction of ∇f . This gives us two possible values of ∇f , one in the direction of increase and one in the direction of decrease. We take the value of ∇f as the one pointing in the direction of increase.

It is important to note that ∇f doesn't point "uphill" as it is sometimes said to. ∇f is in the space \mathbb{R}^n which maps to the space \mathbb{R} . For the case $n = 2$ we can think of this as ∇f being in the x, y plane as a projection of a vector in \mathbb{R}^3 pointing "uphill".

15 Change of Variables and Chain Rules

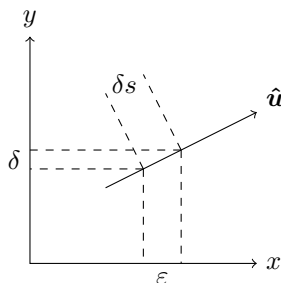


Figure 15.1: Small step δs in the direction of $\hat{\mathbf{u}}$

If we take a small step δs in the direction of $\hat{\mathbf{u}}$ as shown in figure 15.1 then the derivative in that direction can be taken as

$$\lim_{\epsilon, \delta \rightarrow 0} \frac{f(x + \epsilon, y + \delta) - f(x, y)}{\sqrt{\epsilon^2 + \delta^2}} = \frac{df}{ds} = \hat{\mathbf{u}} \cdot \nabla f(x, y)$$

Similarly for a small step δs along a curve we can just consider the derivative in the direction of the tangent at that point.

15.1 Change of Variables

Say we want to change variables from (x, y) to $s = s(x, y)$. We can rewrite our old variables as functions of s : $x = x(s)$ and $y = y(s)$. Then our derivative in the direction of the curve $s = s(x, y)$ is

$$\begin{aligned}\frac{df}{ds} &= \lim_{\delta s \rightarrow 0} \frac{f(s + \delta s) - f(s)}{\delta s} \\ \frac{df}{ds} &= \lim_{\delta s \rightarrow 0} \frac{1}{\delta s} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} \cdot \nabla f\end{aligned}$$

where $f(x, y) \equiv f(x(s), y(s)) \equiv f(s)$. We can expand the dot product to get

$$\frac{df}{ds} = \begin{pmatrix} \frac{dx}{ds} \\ \frac{dy}{ds} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \frac{dx}{ds} \frac{\partial f}{\partial x} + \frac{dy}{ds} \frac{\partial f}{\partial y} \quad (15.1)$$

This is a chain rule. If we let s vary with time such that $s = s(t)$ then we can use the normal chain rule to see how f changes

$$\begin{aligned}\frac{ds}{dt} \frac{df}{ds} &= \frac{ds}{dt} \frac{dx}{ds} \frac{\partial f}{\partial x} + \frac{ds}{dt} \frac{dy}{ds} \frac{\partial f}{\partial y} \\ \frac{df}{dt} &= \frac{dx}{dt} \frac{\partial f}{\partial x} + \frac{dy}{dt} \frac{\partial f}{\partial y}\end{aligned}$$

This is the total derivative along a line with some arbitrary time dependence. If we have instead $f(x, t)$ then

$$\frac{df}{dt} = \frac{dx}{dt} \frac{\partial f}{\partial x} + \frac{dt}{dt} \frac{\partial f}{\partial t} = \frac{dx}{dt} \frac{\partial f}{\partial x} + \frac{\partial f}{\partial t}$$

Example 15.1

A wave on a string has wave function $\psi(x, t) = A \cos(kx - \omega t)$. An observer follows the trajectory $x(t = \alpha t)$ for some constant α . What is the rate of change of displacement of the string as seen by the observer?

$$\begin{aligned}\frac{d\psi}{dt} &= \frac{dx}{dt} \frac{\partial \psi}{\partial x} + \frac{\partial \psi}{\partial t} \\ &= -\alpha k A \sin(kx - \omega t) + \omega A \sin(kx - \omega t) \\ &= (\omega - \alpha k) A \sin(kx - \omega t)\end{aligned}$$

If instead we want to change from variables (x, y) to (u, v) where $x = x(u, v)$ and $y = y(u, v)$ how does $f(x, y)$ change if we increase u and hold v constant?

$$\frac{\partial f(x, y)}{\partial u} = \frac{\partial}{\partial u} f(x(u, v), y(u, v)) \quad (15.2)$$

This has contributions from $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$. We set $v = v^*$ which is a constant. This gives us a statement equivalent to equation 15.2

$$\frac{d}{du} f(x(u, v^*), y(u, v^*))$$

which is just like the situation where we had $x = x(s)$ and $y = y(s)$ which gave equation 15.1 so

$$\frac{\partial f(x, y)}{\partial u} = \frac{dx}{du} \frac{\partial f}{\partial x} + \frac{dy}{du} \frac{\partial f}{\partial y}$$

however when we set v back to a variable instead of a constant the derivatives with respect to u become partial derivatives

$$\frac{\partial f(x, y)}{\partial u} = \frac{\partial x}{\partial u} \frac{\partial f}{\partial x} + \frac{\partial y}{\partial u} \frac{\partial f}{\partial y}$$

Writing this and explicitly noting what we are holding constant we get

$$\begin{aligned}\frac{\partial}{\partial u}f(x(u, v), y(u, v)) &= \left(\frac{\partial x}{\partial u}\right)_v \left(\frac{\partial f}{\partial x}\right)_y + \left(\frac{\partial y}{\partial u}\right)_v \left(\frac{\partial f}{\partial y}\right)_x \\ \frac{\partial}{\partial v}f(x(u, v), y(u, v)) &= \left(\frac{\partial x}{\partial v}\right)_u \left(\frac{\partial f}{\partial x}\right)_y + \left(\frac{\partial y}{\partial v}\right)_u \left(\frac{\partial f}{\partial y}\right)_x\end{aligned}$$

This is a chain rule.

Two lines are at right angles. The directions of these lines are described by $\mathbf{u} = (\Delta x_1 \ \Delta y_1)^\top$ and $\mathbf{v} = (\Delta x_2 \ \Delta y_2)^\top$. What is the product of their gradients (the slopes not the gradient vector)? They are perpendicular

$$\implies \mathbf{u} \cdot \mathbf{v} = 0$$

$$\Delta x_1 \Delta x_2 + \Delta y_1 \Delta y_2 = 0$$

$$-\Delta x_1 \Delta x_2 = \Delta y_1 \Delta y_2$$

$$\frac{\Delta y_1 \Delta y_2}{\Delta x_1 \Delta x_2} = -1$$

$$\frac{\Delta y_1}{\Delta x_1} \frac{\Delta y_2}{\Delta x_2} = -1$$

$$\text{slope of } \mathbf{u} \cdot \text{slope of } \mathbf{v} = -1$$

so the product of their gradients is -1 . For $\hat{\mathbf{u}}$ pointing along a contour $\hat{\mathbf{u}} \cdot \nabla f = 0$. Therefore if we multiply the gradients of $\hat{\mathbf{u}}$ and ∇f we get -1 :

$$-1 = \left(\frac{dy}{dx} \Big|_{\text{contour}} \right) \frac{\left(\frac{\partial f}{\partial y} \right)_x}{\left(\frac{\partial f}{\partial x} \right)_y}$$

We can consider the first term as the partial derivative of y with respect to x while holding f constant since that is the definition of a contour. This gives us

$$\left(\frac{\partial y}{\partial x} \right)_f = - \left(\frac{\partial y}{\partial x} \right)_f \left(\frac{\partial f}{\partial y} \right)_x$$

This is a chain rule. Sometimes called the cyclic chain rule or triple product rule. The reason for these names is clear if we rearrange to get

$$\left(\frac{\partial x}{\partial y} \right)_f \left(\frac{\partial y}{\partial f} \right)_x \left(\frac{\partial f}{\partial x} \right)_y = -1$$

16 Higher Order Derivatives and Taylor Expansions

Both $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are functions of x and y . This gives four options for second order partial derivatives:

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

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

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

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

For a continuous function it can be shown that $f_{xy} = f_{yx}$.

16.1 Taylor Expansions

For a Taylor series we expect each possible derivative to play some part in the expansion. Since Taylor series only exist if the function is infinitely differentiable at the point of interest this means that the function must be continuous at all derivatives at that point so $f_{xy} = f_{yx}$ and the same is true for higher order derivatives. Taking this into account we expect the Taylor series of function f to look like:

$$f(x, y) \approx a + b_x x + b_y y + c_{xx} x^2 + c_{xy} xy + c_{yy} y^2 + \dots$$

By taking derivatives we can find the values of the constants.

$$f(0, 0) = a \implies f \Big|_{\substack{x=0 \\ y=0}} = a$$

$$\frac{\partial f}{\partial x} = b_x + 2c_{xx}x + c_{xy}y + \dots \implies \frac{\partial f}{\partial x} \Big|_{\substack{x=0 \\ y=0}} = b_x$$

$$\frac{\partial f}{\partial y} = b_y + c_{xy}y + 2c_{yy}y + \dots \implies \frac{\partial f}{\partial y} \Big|_{\substack{x=0 \\ y=0}} = b_y$$

$$\frac{\partial^2 f}{\partial x^2} = 2c_{xx} + \dots \implies \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \Big|_{\substack{x=0 \\ y=0}} = c_{xx}$$

$$\frac{\partial^2 f}{\partial y^2} = 2c_{yy} + \dots \implies \frac{1}{2} \frac{\partial^2 f}{\partial y^2} \Big|_{\substack{x=0 \\ y=0}} = c_{yy}$$

$$\frac{\partial^2 f}{\partial y \partial x} = c_{xy} + \dots \implies \frac{\partial^2 f}{\partial y \partial x} \Big|_{\substack{x=0 \\ y=0}} = c_{xy}$$

This gives us the Taylor series for f at the point $(0, 0)$:

$$f(x, y) = f(0, 0) + f_x(0, 0)x + f_y(0, 0)y + \frac{1}{2}f_{xx}(0, 0)x^2 + f_{xy}(0, 0)xy + \frac{1}{2}f_{yy}(0, 0)y^2 + \dots$$

or around a point (x_0, y_0) :

$$\begin{aligned} f(x, y) &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ &\quad + \frac{1}{2}f_{xx}(x_0, y_0)(x - x_0)^2 + f_{xy}(x_0, y_0)(x - x_0)(y - y_0) + \frac{1}{2}f_{yy}(y - y_0)^2 + \dots \end{aligned}$$

16.2 Stationary Points

At a stationary point the gradient is 0 in all directions. Therefore $\hat{\mathbf{u}} \cdot \nabla f = 0 \forall \hat{\mathbf{u}} \implies \nabla f = \mathbf{0}$. Using the Taylor series we can see that at a stationary point

$$f(x, y) - f(0, 0) \approx \frac{1}{2}f_{xx}(0, 0)x^2 + f_{xy}(0, 0)xy + \frac{1}{2}f_{yy}(0, 0)y^2$$

This can be written as a matrix/vector product

$$f(x, y) - f(0, 0) \approx \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} H \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2} \mathbf{r}^T H \mathbf{r}$$

where \mathbf{r} is the position vector from the origin to the point (x, y) and H is the Hessian matrix defined as

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

since

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

we can see that $H_{ij} = H_{ji}$ so H is symmetric. This means that H has orthogonal eigenvectors \mathbf{v}_i with eigenvalues λ_i . We can rewrite \mathbf{r} in terms of these eigenvectors as

$$\mathbf{r} = q_1 \mathbf{v}_1 + q_2 \mathbf{v}_2$$

$$\begin{aligned} f(x, y) - f(0, 0) &\approx \frac{1}{2} \mathbf{r}^T H \mathbf{r} \\ &= \frac{1}{2} (q_1 \hat{\mathbf{v}}_1 + q_2 \hat{\mathbf{v}}_2)^T H (q_1 \hat{\mathbf{v}}_1 + q_2 \hat{\mathbf{v}}_2) \\ &= \frac{1}{2} (q_1 \hat{\mathbf{v}}_1 + q_2 \hat{\mathbf{v}}_2)^T (q_1 H \hat{\mathbf{v}}_1 + q_2 H \hat{\mathbf{v}}_2) \\ &= \frac{1}{2} (q_1 \hat{\mathbf{v}}_1 + q_2 \hat{\mathbf{v}}_2)^T (q_1 \lambda_1 \hat{\mathbf{v}}_1 + q_2 \lambda_2 \hat{\mathbf{v}}_2) \\ &= \frac{1}{2} (q_1 \hat{\mathbf{v}}_1 + q_2 \hat{\mathbf{v}}_2) \cdot (q_1 \lambda_1 \hat{\mathbf{v}}_1 + q_2 \lambda_2 \hat{\mathbf{v}}_2) \\ &= \frac{1}{2} (q_1^2 \lambda_1 + q_2^2 \lambda_2) \end{aligned}$$

This lets us write

$$\begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = O \begin{pmatrix} x \\ y \end{pmatrix}$$

where O is the orthogonal matrix such that $H = O^T D O$ where $D_{ii} = \lambda_i$ and O has eigenvectors $\hat{\mathbf{v}}_i$ as rows.

The nature of a stationary point can be determined by the eigenvalues of the Hessian matrix for that point. If $\lambda_i > 0 \forall i$ then the stationary point is a local minimum. If $\lambda_i < 0 \forall i$ then the stationary point is a local maximum. In other cases the stationary point is a saddle point.

17 Integration

17.1 1D Integrals

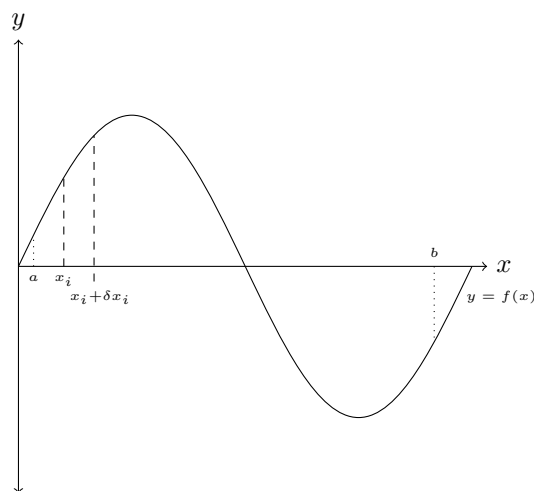


Figure 17.1: Integrating $f(x)$

The (Riemann) integral of $f(x)$ over the range $[a, b]$ is defined as splitting the area between the curve and the x axis into n vertical slices at positions x_i where $x_0 = a$ and $x_n = b$ and $x_i < x_{i+1}$. The i^{th} slice has

thickness δx_i . The integral is then found by summing up f evaluated at x_i times the width of the slice as $n \rightarrow \infty$ and $\delta x_i \rightarrow 0$. This gives

$$I = \lim_{n \rightarrow \infty} \sum_{i=0}^n f(x_i) \delta x_i = \int_a^b f(x) dx$$

This integral can be interpreted as an area if and only if $f(x)$ can be interpreted as the height above $y = 0$ at x . Even then it is still a signed area (area below the x axis is negative).

17.2 2D Integrals

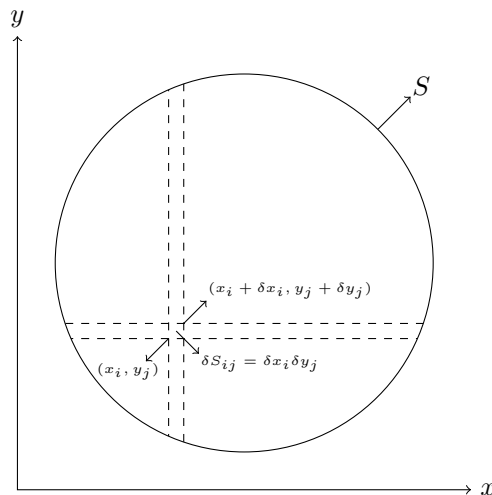


Figure 17.2: Integrating $f(x, y)$

The (Riemann) integral of $f(x, y)$ for $x, y \in S$ is defined as splitting area S into n vertical slices at positions x_i where $x_i < x_{i+1}$ and n horizontal slices at positions y_j where $y_j < y_{j+1}$. The width of the i^{th} vertical slice is δx_i and the width of the j^{th} horizontal slice is δy_j . The area of the segment where $x \in [x_i, x_{i+1}]$ and $y \in [y_j, y_{j+1}]$ is then $\delta S_{ij} = \delta x_i \delta y_j$. The integral is then found by summing up f evaluated at (x_i, y_j) times the area of the segment as $n \rightarrow \infty$ and $\delta S_{ij} \rightarrow 0$. This gives

$$I = \lim_{n \rightarrow \infty} \sum_{i=0}^n \sum_{j=0}^n f(x_i, y_j) \delta S_{ij} = \int_S f(x, y) dS$$

To avoid confusion later we also introduce the notation that the order of the $f(x, y)$ and dS does not matter so the following are equivalent:

$$\int_S f(x, y) dS = \int_S dS f(x, y)$$

This integral can be interpreted as the volume between the surface and the x, y plane if and only if $f(x, y)$ can be interpreted as the height above the x, y plane. Even then it is still a signed area (area below the x, y plane axis is negative).

17.3 Evaluating 2D Integrals

How do we evaluate a two dimensional integral? As an example we will integrate the function f over the area shown in figure 17.3 We can think of this being a 4 step process:

1. Increase x from a to b in steps of dx so $a \rightarrow a + dx \rightarrow a + 2dx \rightarrow \dots \rightarrow b - dx \rightarrow b$

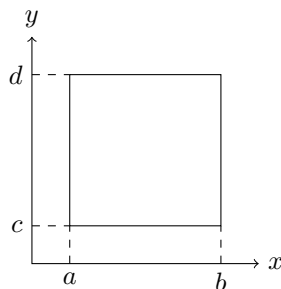


Figure 17.3: A simple 2D integral

2. For each x value from step one increase y from c to d in steps of dy so $c \rightarrow c + dy \rightarrow c + 2dy \rightarrow \dots \rightarrow d - dy \rightarrow d$
3. At each (x, y) from steps one and two evaluate $f(x, y)dx dy$
4. The integral is the sum of every result from step 3

If we do the above for $dx, dy \rightarrow 0$ then this is the integral. We can approximate the integral using non-zero dx and dy with the following C++ code:

```
double f(double x, double y) {
    ...
}

double dx = 0.001;
double dy = 0.001;
double a = 1;
double b = 2;
double c = 1;
double d = 2;

double integral = 0;
for (double x = 0; x < b; x += dx) {
    for (double y = 0; y < d; y += dy) {
        integral += f(x, y) * dx * dy;
    }
}
```

We can write this process with integrals as:

$$\int_a^b dx \int_c^d dy f(x, y) \equiv \int_S dS f(x, y)$$

The first integral written represents step one, the second integral written represents step two and the evaluation and summing in steps three and four are implicit in integrals. This is also written sometimes as the following:

$$\int_a^b \int_c^d f(x, y) dy dx = \int_a^b dx \int_c^d f(x, y) dy \equiv \int_S dS f(x, y)$$

but the order of integration can be tricky to deduce from this especially since not everyone is consistent. This means that the first way of writing it is probably best. It is also possible to be explicit in the limits about the variable for example replacing the a in the integral with respect to x with $x = a$ this gives the most explicit (and maybe slightly pedantic) way of writing this as

$$\int_{x=a}^{x=b} dx \int_{y=c}^{y=d} dy f(x, y) \equiv \int_S dS f(x, y)$$

This is especially important when the limits aren't constants.

Example 17.1

$$\begin{aligned}
 I &= \int_{x=0}^3 dx \int_{y=0}^4 dy xy \\
 &= \int_{x=0}^3 dx x \left[\frac{y^2}{2} \right]_0^4 \\
 &= \int_{x=0}^3 dx 8x \\
 &= 36
 \end{aligned}$$

We can take any variable that we aren't integrating with respect to as a constant. For constant limits a , b , c and d and functions g and h it is possible to do a separation of variables:

$$\int_a^b dx \int_c^d dy g(x)h(y) = \left[\int_a^b g(x) dx \right] \left[\int_c^d h(y) dy \right]$$

An example where this isn't possible is integrating over the triangle between the x axis, the line $x = a$ and the line $y = x$ for some function f . If we split the triangle in the x direction first and then in the y direction we end up with slices of height $y_i = x_i$ split into sections in the y direction. This results in the integral

$$I = \int_0^a dx \int_0^x dy f(x, y)$$

If instead we split the triangle in the y direction first and then in the x direction we get slices of length $x_i = y_i$ split into sections in the x direction. This results in the integral

$$I = \int_0^a dy \int_y^a dx f(x, y)$$

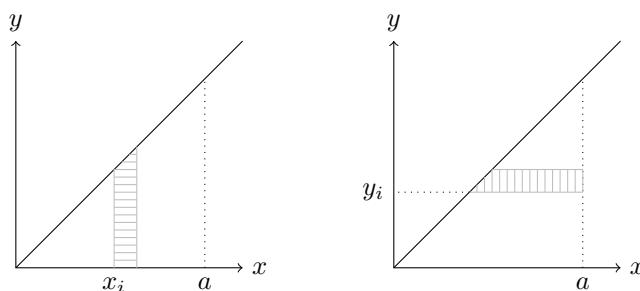


Figure 17.4: Splitting the same shape multiple ways

Example 17.2

$$\begin{aligned}
 I &= \int_1^2 dx \int_x^2 dy \frac{1}{y} \\
 &= \int_1^2 dx [\ln y]_x^2 \\
 &= \int_1^2 dx [\ln 2 - \ln x]
 \end{aligned}$$

$$\begin{aligned}
&= [x \ln 2 - x \ln x + x]_1^2 \\
&= 2 \ln 2 - 2 \ln 2 + 2 - (\ln 2 - \ln 1 + 1) \\
&= 2 - \ln 2 - 1 \\
&= 1 - \ln 2
\end{aligned}$$

This integral can be calculated much more easily by swapping the order of integration

$$\begin{aligned}
I &= \int_1^2 dx \int_x^2 dy \frac{1}{y} \\
&= \int_1^2 dy \int_1^y dx \frac{1}{y}
\end{aligned}$$

The integral with respect to x must be calculated first as its limits depend on y

$$\begin{aligned}
&= \left[\int_1^2 dy \frac{1}{y} \right] \left[\int_1^y dx \right] \\
&= \int_1^2 dy [x]_1^y \\
&= \int_1^2 dy \frac{1}{y} (y - 1) \\
&= \int_1^2 dy \left[1 - \frac{1}{y} \right] \\
&= [y - \ln y]_1^2 \\
&= 2 - \ln 2 - (1 - \ln 1) \\
&= 1 - \ln 2
\end{aligned}$$

The second way round avoids having to integrate $\ln x$. When swapping the order of integration one must take care to ensure that the limits of the integrals are correct.

Example 17.3

$$I = \int_0^1 dx \int_0^{2-x} dy \frac{x}{y}$$

(Note that this integral is actually undefined as x/y can't be evaluated at $y = 0$) This is asking us to integrate over the region R in figure 17.5. The way the integral is written the integral with respect to y is calculated first. If we want to swap the order of the integrals we need to consider that depending on the value of y the equation of the line bounding the space changes. This means that we have to split the integral at the intersection. The limits go from $y \rightarrow x$ and $y = 2 - x \rightarrow x = 2 - y$.

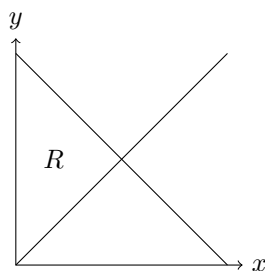


Figure 17.5: Region bounded by $x = 0$, $y = x$ and $y = 2 - x$

$$I = \left[\int_0^1 dy \int_0^y dx \frac{x}{y} \right] + \left[\int_1^2 dy \int_0^{2-y} dx \frac{x}{y} \right]$$

This is now separable as

$$I = \left[\int_0^1 dy \frac{1}{y} \int_0^y dx x \right] + \left[\int_1^2 dy \frac{1}{y} \int_0^{2-y} dx x \right]$$

This technique can make undoable (in elementary functions) integrals possible, for example the integral

$$I = \int_0^1 dy \int_y^1 dx \frac{\sin x}{x}$$

is impossible to do in elementary functions but by swapping the order of integration we get

$$\begin{aligned} I &= \int_0^1 dx \int_0^x dy \frac{\sin x}{x} \\ &= \int_0^1 \left[\frac{\sin x}{x} y \right]_0^x dx \\ &= \int_0^1 \sin x dx \\ &= [-\cos x]_0^1 \\ &= 1 - \cos 1 \end{aligned}$$

18 Change of Basis

18.1 Polar Coordinates

If we want to integrate over the surface S that is a circle radius R centre $(0,0)$ then we *can* do this in Cartesian coordinates. If we first split the surface in to strips in the x direction then this gives us x limits of $-R$ and R . We next split the each strip in the y direction. This gives y limits of $y = \pm\sqrt{R^2 - x^2}$ since the equation of the circle is $x^2 + y^2 = R^2$. Therefore the integral of the function f over S is

$$I = \int_S dS f(x, y) = \int_{-R}^R dx \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy f(x, y)$$

however the limits make this horrible even before we consider what $f(x, y)$ is.

It is easier to calculate this integral in polar coordinates which are related to Cartesian coordinates by:

$$x = \rho \cos \varphi \quad y = \rho \sin \varphi$$

To do this we split the circle as shown in figure 18.1. The area dS tends to a rectangle as $d\rho \rightarrow 0$. This rectangle has side lengths $d\rho$ and $\rho d\varphi$ so the area is given by $dS = dx dy = \rho d\rho d\varphi$. This circle has ρ limits 0 to R and φ limits 0 to 2π . This means our integral I is

$$I = \int_0^R d\rho \int_0^{2\pi} d\varphi \rho f(x, y)$$

Example 18.1

$$\begin{aligned} I &= \int dS xy \\ &= \int_0^R d\rho \int_0^{2\pi} d\varphi \rho \cos \varphi \sin \varphi \\ &= \int_0^R d\rho \int_0^{2\pi} d\varphi \rho^3 \cos \varphi \sin \varphi \end{aligned}$$

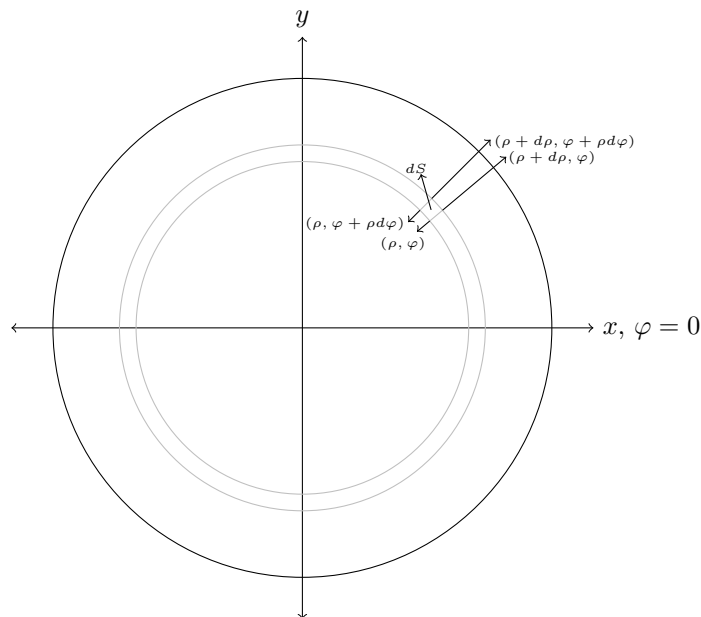


Figure 18.1: Integrating over a circle

$$\begin{aligned}
 &= \left[\int_0^R \rho^3 d\rho \right] \left[\int_0^{2\pi} \cos \varphi \sin \varphi d\varphi \right] \\
 &= \left[\frac{\rho^4}{4} \right]_0^R \left[\frac{1}{2} \sin^2 \varphi \right]_0^{2\pi} \\
 &= 0
 \end{aligned}$$

The fact that this is 0 makes sense since the surface is symmetric about both axis and the function will be negative in top left and bottom right quadrants and positive elsewhere so much like an odd function integrated over even limits the positive and negative parts cancel to 0.

18.2 The Jacobian

It was (relatively) easy to work out that $dx dy = \rho d\rho d\varphi$ geometrically but this will not always be the case, especially in higher dimensions. We want to come up with a method for working out the element we integrate in the general case.

Consider two coordinate systems the Cartesian (x, y) system and some irregular (u, v) system. Let \mathbf{a} be a vector along a line of constant v and \mathbf{b} be a vector along a line of constant u . We can work out what \mathbf{a} and \mathbf{b} are in terms of \mathbf{x} and \mathbf{y} which are vectors in the x and y directions respectively and du and dv which are small displacements in the u and v directions respectively

$$\begin{aligned}
 \mathbf{a} &= \frac{\partial \mathbf{x}}{\partial u} du + \frac{\partial \mathbf{y}}{\partial u} du \\
 \mathbf{b} &= \frac{\partial \mathbf{x}}{\partial v} dv + \frac{\partial \mathbf{y}}{\partial v} dv
 \end{aligned}$$

We can put \mathbf{a} and \mathbf{b} as the columns of a matrix and make this a matrix equation using the fact that a determinant is a scale factor for the transformation we can get dS :

$$dS = dx dy = \left| \det \begin{pmatrix} \frac{\partial x}{\partial u} du & \frac{\partial x}{\partial v} dv \\ \frac{\partial y}{\partial u} du & \frac{\partial y}{\partial v} dv \end{pmatrix} \right|$$

$$\begin{aligned}
&= \left| \frac{\partial x}{\partial u} du \frac{\partial y}{\partial v} dv - \frac{\partial x}{\partial v} dv \frac{\partial y}{\partial u} du \right| \\
&= \left| \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \right| du dv \\
&= \left| \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \right| du dv
\end{aligned}$$

The matrix in the last step is the Jacobian matrix (often just the Jacobian) change of basis. The (absolute value of the) determinant is then known as the Jacobian derivative (confusingly also often just the Jacobian). In general the Jacobian takes in two n and m dimensional sets of coordinates x_i and x'_j and gives the relationship between the unit volumes. There are various notations for the Jacobian, some of these are

$$J = J(x'_1, x'_2, \dots, x'_n) = \frac{\partial(x_1, x_2, \dots, x_m)}{\partial(x'_1, x'_2, \dots, x'_n)}$$

The $n \times m$ Jacobian matrix is then defined by

$$J_{ij} = \frac{\partial x_i}{\partial x'_j}$$

A few shortcuts with determinants are useful here, for an $n \times n$ matrix M :

- $\det(M) = \det(M^T)$ so if you accidentally calculate $\frac{\partial x'_i}{\partial x_j}$ instead for J_{ij} it doesn't matter
- A transformation $M \rightarrow M'$ that results in one row or column being multiplied by k has determinant $\det(M') = k \det(M)$
- Multiplication of M by a scalar k results in a determinant $\det(kM) = k^n \det(M)$. This can be thought of as each row individually being multiplied by k .

18.3 Cylindrical Coordinates

Cylindrical coordinates are related to Cartesian coordinates by

$$x = \rho \cos \varphi \quad y = \rho \sin \varphi \quad z = z$$

If we are going to integrate over a cylinder radius R height h placed such that $z = 0$ is halfway up the cylinder then the ρ limits are 0 and R , the φ limits are 0 and 2π and the z limits are $-h/2$ and $h/2$. The Jacobian is

$$\begin{aligned}
\frac{\partial(x, y, z)}{\partial(\rho, \varphi, z)} &= \left| \det \begin{pmatrix} \cos \varphi & \sin \varphi & 0 \\ -\rho \sin \varphi & \rho \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \right| \\
&= |\cos \varphi \rho \cos \varphi + \rho \sin \varphi \sin \varphi| \\
&= \rho
\end{aligned}$$

Hence a volume element dV is

$$dV = dx dy dz = \rho d\rho d\varphi dz$$

Hence an integral over a cylinder can be calculated as

$$\int_V dV f(x, y, z) = \int_0^R d\rho \int_0^{2\pi} d\varphi \int_{-h/2}^{h/2} dz \rho f(\rho \cos \varphi, \rho \sin \varphi, z)$$

Note that the limits for z depend on the location of the origin.

19 Spherical Coordinates and the Laplacian

19.1 Integrating in Spherical Coordinates

When trying to integrate over a spherically symmetric volume V it is best to do it in spherical coordinates.

$$x = r \sin \vartheta \cos \varphi, \quad y = r \sin \vartheta \sin \varphi, \quad z = r \cos \vartheta$$

where $r \in [0, \infty)$ is the distance from the origin to the point, $\varphi \in [0, 2\pi]$ is the angle from the x axis to the projection of the point in the x, y plane and $\vartheta \in [0, \pi]$ is the angle from the z axis to the line connecting the origin and the point. We need to calculate the value of the Jacobian for spherical coordinates:

$$\begin{aligned} J &= \frac{\partial(x, y, z)}{\partial(r, \vartheta, \varphi)} \\ &= \left| \det \begin{pmatrix} \sin \vartheta \cos \varphi & \sin \vartheta \sin \varphi & \cos \vartheta \\ r \cos \vartheta \cos \varphi & r \cos \vartheta \sin \varphi & -r \sin \vartheta \\ -r \sin \vartheta \sin \varphi & r \sin \vartheta \cos \varphi & 0 \end{pmatrix} \right| \\ &= |\cos \vartheta (r^2 \cos \vartheta \sin \vartheta \cos^2 \varphi + r^2 \cos \vartheta \sin \vartheta \sin^2 \varphi) \\ &\quad + r \sin \vartheta (r \sin^2 \vartheta \cos^2 \varphi + r \sin^2 \vartheta \sin^2 \varphi)| \\ &= |r^2 \cos^2 \vartheta \sin \vartheta + r^2 \sin^2 \vartheta \sin \vartheta| \\ &= r^2 \sin \vartheta \end{aligned}$$

since $\vartheta \in [0, \pi]$ $\sin \vartheta \geq 0 \forall \vartheta$. This means that a volume element dV is given by

$$dV = dx dy dz = r^2 \sin \vartheta, dr d\vartheta d\varphi$$

Hence the integral of a function f over all space is given by

$$\int_V dV = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz f(x, y, z) = \int_0^{\infty} dr \int_0^{\pi} d\vartheta \int_0^{2\pi} d\varphi f(r \sin \vartheta \cos \varphi, r \sin \vartheta \sin \varphi, r \cos \vartheta)$$

19.2 The Laplacian

On page 22 section 14 we defined the gradient operator ∇ as a vector of partial derivatives

$$\nabla_i = \frac{\partial}{\partial x_i}$$

another useful operator is it turns out this dotted with itself, in n dimensions this is:

$$\nabla \cdot \nabla = \nabla^2 = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$$

For example in 3 dimensions:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

This is very common in physics for example in the Schrödinger equation:

$$-\frac{\hbar^2}{2m} \nabla^2 \psi(\mathbf{r}, t) + V(\mathbf{r}, t) = i\hbar \frac{\partial}{\partial t} \psi(\mathbf{r}, t)$$

or the in the wave equation:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u$$

It is also possible to write these operators in polar coordinates of some kind. To do this we find the derivative of some arbitrary function f with respect to our new basis coordinates $(\rho, r, \varphi, \vartheta, \text{whatever else we are using})$

and set up and solve simultaneous equations for the differential operators with respect to our initial base (x , y , z etc.) and then we apply these to themselves to find the second derivatives. For example in plane polar coordinates:

$$x = r \cos \varphi, \quad y = r \sin \varphi$$

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

$$\begin{aligned} \frac{\partial f}{\partial \rho} &= \frac{\partial x}{\partial \rho} \frac{\partial f}{\partial x} + \frac{\partial y}{\partial \rho} \frac{\partial f}{\partial y} \\ &= \cos \varphi \frac{\partial f}{\partial x} + \sin \varphi \frac{\partial f}{\partial y} \\ \frac{\partial}{\partial \rho} &= \cos \varphi \frac{\partial}{\partial x} + \sin \varphi \frac{\partial}{\partial y} \end{aligned} \tag{19.1}$$

$$\begin{aligned} \frac{\partial f}{\partial \varphi} &= \frac{\partial x}{\partial \varphi} \frac{\partial f}{\partial x} + \frac{\partial y}{\partial \varphi} \frac{\partial f}{\partial y} \\ &= -\rho \sin \varphi \frac{\partial f}{\partial x} + \rho \cos \varphi \frac{\partial f}{\partial y} \\ \frac{\partial}{\partial \varphi} &= -\rho \sin \varphi \frac{\partial}{\partial x} + \rho \cos \varphi \frac{\partial}{\partial y} \end{aligned} \tag{19.2}$$

We can solve the simultaneous equations 19.1 and 19.2 to get our differential for x and y . Doing this gives us

$$\begin{aligned} \frac{\partial}{\partial x} &= \cos \varphi \frac{\partial}{\partial \rho} - \frac{\sin \varphi}{\rho} \frac{\partial}{\partial \varphi} \\ \frac{\partial}{\partial y} &= \sin \varphi \frac{\partial}{\partial \rho} + \frac{\cos \varphi}{\rho} \frac{\partial}{\partial \varphi} \end{aligned}$$

Finally we can calculate the Laplacian as

$$\begin{aligned} \nabla^2 &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \\ &= \left[\cos \varphi \frac{\partial}{\partial \rho} - \frac{\sin \varphi}{\rho} \frac{\partial}{\partial \varphi} \right] \left[\cos \varphi \frac{\partial}{\partial \rho} - \frac{\sin \varphi}{\rho} \frac{\partial}{\partial \varphi} \right] \\ &\quad + \left[\sin \varphi \frac{\partial}{\partial \rho} + \frac{\cos \varphi}{\rho} \frac{\partial}{\partial \varphi} \right] \left[\sin \varphi \frac{\partial}{\partial \rho} + \frac{\cos \varphi}{\rho} \frac{\partial}{\partial \varphi} \right] \\ &= \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial}{\partial \rho} \right) + \frac{1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \varphi^2} \\ &= \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \varphi^2} \end{aligned}$$

Following the same method it is possible to derive the Laplacian in cylindrical and spherical coordinates as well. Cylindrical:

$$\nabla^2 = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2}{\partial \varphi^2} + \frac{\partial^2}{\partial z^2}$$

Spherical:

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{r^2 \sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2}$$