IA Mathematics B Note

Ruize Li

Jesus College, Cambridge

January $\sim$ June 2024

Contents

Part I

Michaelmas Term

This note was originally not in my plan, and was created for Tripos exams. Therefore it contains mostly equations, derivation and proofs. I didn't go to mathematics lectures so there will be nothing from my own note nor the handout.

Chapter 1

Vectors

1.1 Definitions

Scalar is a value in $\mathbb{R}$, which has only magnitude but no direction, such as temperature, density, speed, etc.
Vector is a value in vector space $\mathcal{J}$[1], which has magnitude and direction, such as velocity, acceleration, force, etc.
Euclidean space is a space that satisfies:

- Two points uniquely define a straight line.
- Parallel lines exists.

In **Cartesian coordinates**, a vector can be written[2] as:

$$\vec{A} = A_x\hat{i} + A_y\hat{j} + A_z\hat{k} = (A_x, A_y, A_z). \tag{1.1}$$

The magnitude of that vector is defined to be

$$|\vec{A}| = A, \tag{1.2}$$

and the direction is described by the unit vector along $\vec{A}$

$$\hat{A} = \frac{\vec{A}}{|\vec{A}|}. \tag{1.3}$$

Cartesian coordinates is mutually perpendicular, emanate for fixed origin O, have fixed direction. In 3-D, by conversion, we always use right-handed coordinate system.

1.2 Algorithm of vectors

Summing two vectors is just summing their components, while multiplying a number to a vector is a kind of scaling to its magnitude.

$$\begin{aligned} \vec{A} + \vec{B} &= \vec{B} + \vec{A} \\ \vec{A} + (\vec{B} + \vec{C}) &= (\vec{A} + \vec{B}) + \vec{C} \\ |\alpha\vec{A}| &= |\alpha||\vec{A}|,\ \alpha \in \mathbb{R} \\ \vec{A} - \vec{B} &= \vec{A} + (-\vec{B}) \\ \alpha(\vec{A} + \vec{B}) &= \alpha\vec{A} + \alpha\vec{B} \end{aligned} \tag{1.4}$$

[1] This is not a good notation, but we won't be using it, so whatever.
[2] Using vector components, which are in $\mathbb{R}^3$

Two vectors can form two kinds of products, scalar product (dot product or inner product) and vector product (cross product or outer product).
Scalar product produces a scalar:

$$\vec{A}\cdot\vec{B}=|\vec{A}||\vec{B}|\cos\langle\vec{A},\vec{B}\rangle, \tag{1.5}$$

following these rules:

$$\begin{aligned}\vec{A}\cdot\vec{B}&=\vec{B}\cdot\vec{A}\\ \vec{A}\cdot(\vec{B}+\vec{C})&=\vec{A}\cdot\vec{B}+\vec{A}\cdot\vec{C}\\ (\vec{a}+\vec{b})^2&=\vec{a}^2+\vec{b}^2+2\vec{a}\cdot\vec{b}\end{aligned} \tag{1.6}$$

Vector product produces a vector:

$$\vec{A}\times\vec{B}=\begin{vmatrix}\hat{i}&\hat{j}&\hat{k}\\ A_x&A_y&A_z\\ B_x&B_y&B_z\end{vmatrix}. \tag{1.7}$$

One can easily find that $|\vec{A}\times\vec{B}|=|\vec{A}||\vec{B}|\sin\langle\vec{A},\vec{B}\rangle$, and the fact that the resulting vector is perpendicular to both $\vec{A}$ and $\vec{B}$. By definition its direction follows the right-handed conversion.
Vector product has the following rules:

$$\begin{aligned}\vec{A}\times\vec{B}&=-\vec{B}\times\vec{A}\\ \vec{A}\times(\vec{B}+\vec{C})&=\vec{A}\times\vec{B}+\vec{A}\times\vec{C}\\ \vec{A}\cdot(\vec{B}\times\vec{C})&=-\vec{A}\cdot(\vec{C}\times\vec{B})=-\vec{C}\cdot(\vec{B}\times\vec{A})\\ \vec{A}\times(\vec{B}\times\vec{C})&=(\vec{A}\cdot\vec{C})\vec{B}-(\vec{A}\cdot\vec{B})\vec{C}\end{aligned} \tag{1.8}$$

1.3 Geometry described by vectors

Angle between lines:

$$\theta=\arccos\left(\frac{\vec{A}\cdot\vec{B}}{|\vec{A}||\vec{B}|}\right). \tag{1.9}$$

Equation of a line (A, B are points on the line):

$$\vec{r}=\vec{a}+\lambda(\vec{b}-\vec{a})=(1-\lambda)\vec{a}+\lambda\vec{b}. \tag{1.10}$$

it is clearly seen that the direction of this line is $\vec{\tau}=\frac{\vec{b}-\vec{a}}{|\vec{b}-\vec{a}|}$. The component form of equation of a lines is

$$\frac{x-a_x}{b_x-a_x}=\frac{y-a_y}{b_y-a_y}=\frac{z-a_z}{b_z-a_z}. \tag{1.11}$$

Using vector products:

$$(\vec{r}-\vec{a})\times(\vec{b}-\vec{a})=\vec{0}. \tag{1.12}$$

Equation of a plane (A is on the plane and $\hat{n}$ is a normal unit vector on the plane):

$$(\vec{r}-\vec{a})\cdot\hat{n}=0, \tag{1.13}$$

or, if p is the distance from our origin to the plane,

$$\vec{r}\cdot\hat{n}=p. \tag{1.14}$$

If A, B, C are points on the plane, a straightforward result is

$$(\vec{r}-\vec{a})=\alpha(\vec{b}-\vec{a})+\beta(\vec{c}-\vec{a}). \tag{1.15}$$

this results in

$$\vec{r} = \vec{a} + \alpha(\vec{b} - \vec{a}) + \beta(\vec{c} - \vec{a}) = (1 - \alpha - \beta)\vec{a} + \alpha\vec{b} + \beta\vec{c}. \tag{1.16}$$

Using vector products:

$$(\vec{r} - \vec{a}) \cdot [(\vec{b} - \vec{a}) \times (\vec{c} - \vec{a})] = 0 \tag{1.17}$$

Equation of a sphere, ρ is the radius and $\vec{a}$ is the center of the sphere.

$$|\vec{r} - \vec{a}| = \rho. \tag{1.18}$$

Equation of a cylinder, R is the radius and $\hat{n}$ is along the axis.

$$|\vec{r} - (\vec{r} \cdot \hat{n})\hat{n}| = R. \tag{1.19}$$

If the axis goes through a point A, replace $\vec{r}$ by $\vec{r} - \vec{a}$.
Equation of a cone, Q is the vertex and $\hat{n}$ is its axis, the semi-angle is α.

$$(\vec{r} - \vec{q}) \cdot \hat{n} = |\vec{r} - \vec{q}| \cos\alpha = \text{const.} \tag{1.20}$$

Distance from a point to a line. Suppose point Q has position $\vec{q}$,

$$d = \frac{|(\vec{q} - \vec{a}) \times (\vec{b} - \vec{a})|}{|\vec{b} - \vec{a}|}. \tag{1.21}$$

Distance from a point to a plane. Position of P is $\vec{p}$.

$$d = |(\vec{p} - \vec{a}) \cdot \hat{n}|. \tag{1.22}$$

Distance from a lint to a line. This does not always exists because lines can intersect. A is on line parallel to $\vec{t}$, B is on line parallel to $\vec{u}$.

$$d = \frac{|(\vec{b} - \vec{a}) \cdot (\vec{t} \times \vec{u})|}{\vec{t} \times \vec{u}}. \tag{1.23}$$

Condition for two lines to intersect is therefore

$$(\vec{b} - \vec{a}) \cdot (\vec{t} \times \vec{u}) = 0 \tag{1.24}$$

1.4 Basis

Linearly independent vectors can form a set of basis, it requires that

$$\vec{a} \cdot \vec{b} \times \vec{c} = [\vec{a}, \vec{b}, \vec{c}] \neq 0. \tag{1.25}$$

If $\vec{a}, \vec{b}, \vec{c}$ does form a basis, then the reciprocal basis is given by

$$\vec{A} = \frac{\vec{b} \times \vec{c}}{[\vec{a}, \vec{b}, \vec{c}]},\ \vec{B} = \frac{\vec{c} \times \vec{a}}{[\vec{a}, \vec{b}, \vec{c}]},\ \vec{C} = \frac{\vec{a} \times \vec{b}}{[\vec{a}, \vec{b}, \vec{c}]}. \tag{1.26}$$

This set of basis is special because

$$\vec{a} \cdot \vec{A} = 1,\ \vec{a} \cdot \vec{B} = \vec{a} \cdot \vec{C} = 0. \tag{1.27}$$

Note that

$$\vec{A} \cdot \vec{r} = \alpha\vec{A} \cdot \vec{a} + \beta\vec{A} \cdot \vec{b} + \gamma\vec{A} \cdot \vec{c} = \alpha, \tag{1.28}$$

where

$$\alpha = \frac{\vec{r} \cdot \vec{b} \times \vec{c}}{[\vec{a}, \vec{b}, \vec{c}]},\ \beta = \frac{\vec{r} \cdot \vec{c} \times \vec{a}}{[\vec{a}, \vec{b}, \vec{c}]},\ \gamma = \frac{\vec{r} \cdot \vec{a} \times \vec{b}}{[\vec{a}, \vec{b}, \vec{c}]}. \tag{1.29}$$

Similarly,

$$\vec{a} \cdot \vec{r} = \alpha. \tag{1.30}$$

1.5 Orthogonal coordinates

Doesn't need to write anything because we use them every day in physics.

1.6 Vector area

That is for some vector that is for integration over surface. A vector area has magnitude of that area and direction normal to that plane.
For example, **vector area of any closed body is zero**, and the vector area of a semi-sphere is πr^2.

Chapter 2

Complex Numbers

2.1 Representations

There are two ways of representing complex numbers.

$$\begin{aligned} c &= a + b\mathrm{i} \\ c &= r\mathrm{e}^{\mathrm{i}\phi} \end{aligned} \tag{2.1}$$

In a complex plane (**Argand diagram**), (a, b) is the coordinates in Cartesian system, but (r, ϕ) is that in polar system.

2.2 Algorithm of complex numbers

Rules cannot be simpler.

$$\begin{aligned} c_1 + c_2 &= (a_1 + a_2) + \mathrm{i}(b_1 + b_2) \\ c_1 c_2 &= r_1 r_2 \mathrm{e}^{\mathrm{i}(\phi_1 + \phi_2)} \end{aligned} \tag{2.2}$$

Modulus and **argument** angle are:

$$|c| = r\sqrt{a^2 + b^2}, \quad \arg(c) = \theta = \arctan\frac{b}{a}. \tag{2.3}$$

principal argument is the argument angle in $[-\pi, \pi]$.

2.3 De Moivre's Theorem

$$\cos(n\theta) + \mathrm{i}\sin(n\theta) = (\cos\theta + \mathrm{i}\sin\theta)^n. \tag{2.4}$$

If we want to sum the series $S_N = \sum_{k=0}^{N-1} \cos(k\theta)$. Then the thing to do is to write it as real parts of exponential, the question would reduce to geometric progression.

2.4 General powers

$$\begin{aligned} c_1^{c_2} &= \exp(c_2 \ln(c_1)) \\ &= \exp[(a_2 + \mathrm{i}b_2)(\ln|c_1| + \mathrm{i}(\theta + 2n\pi))] \\ &= \exp[(a_2 \ln|c_1| - b_2(\theta + 2n\pi)) + \mathrm{i}(b_2 \ln|c_1| + a_2(\theta + 2n\pi))] \\ &= \frac{|c_1|^{a_2}}{\exp[b_2(\theta + 2n\pi)]} \exp[\mathrm{i}(b_2 \ln|c_1| + a_2(\theta + 2n\pi)] \end{aligned} \tag{2.5}$$

2.5 Fundamental theorem of algebra

If $P(z)$ is a polynomial of degree $n \in \mathbb{Z},\ n \geq 1$,

$$P(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots + a_n z^n, \quad a_n \neq 0, \tag{2.6}$$

then $P(z) = 0$ has n roots. If z_1 is a root, then

$$P(z) = (z - z_1)Q(z) = 0 \ \Rightarrow\ Q(z) = 0. \tag{2.7}$$

Chapter 3

Hyperbolic Functions

3.1 Hyperbolic functions

We know that

$$\begin{aligned} \cos\theta &= \frac{\mathrm{e}^{\mathrm{i}\theta} + \mathrm{e}^{-\mathrm{i}\theta}}{2} \\ \sin\theta &= \frac{\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{-\mathrm{i}\theta}}{2\mathrm{i}} \\ \tan\theta &= \frac{1}{\mathrm{i}} \cdot \frac{\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{-\mathrm{i}\theta}}{\mathrm{e}^{\mathrm{i}\theta} + \mathrm{e}^{-\mathrm{i}\theta}} \end{aligned} \tag{3.1}$$

a different set of functions are

$$\begin{aligned} \cosh\phi &= \frac{\mathrm{e}^{\phi} + \mathrm{e}^{-\phi}}{2} \\ \sinh\phi &= \frac{\mathrm{e}^{\phi} - \mathrm{e}^{-\phi}}{2} \\ \tanh\phi &= \frac{\mathrm{e}^{\phi} - \mathrm{e}^{-\phi}}{\mathrm{e}^{\phi} + \mathrm{e}^{-\phi}} \end{aligned} \tag{3.2}$$

Relations between them are

$$\begin{aligned} \sin\mathrm{i}\theta &= \mathrm{i}\sinh\theta \\ \cos\mathrm{i}\theta &= \cosh\theta \\ \tan\mathrm{i}\theta &= \mathrm{i}\tanh\theta \end{aligned} \tag{3.3}$$

Clearly,

$$\cosh^2\theta - \sinh^2\theta = 1. \tag{3.4}$$

Also,

$$\begin{aligned} \cosh(A+B) &= \cosh A\cosh B + \sinh A\sinh B \\ \cosh(A-B) &= \cosh A\cosh B - \sinh A\sinh B \\ \sinh(A+B) &= \sinh A\cosh B + \cosh A\sinh B \\ \sinh(A-B) &= \sinh A\cosh B - \cosh A\sinh B \end{aligned} \tag{3.5}$$

$$\begin{aligned} 1 - \tanh^2\theta &= \frac{1}{\cosh^2\theta} \\ \coth^2\theta - 1 &= \frac{1}{\sinh^2\theta} \end{aligned} \tag{3.6}$$

3.2 Inverse hyperbolic functions

$$\begin{aligned} \sinh^{-1} x &= \ln\left(x + \sqrt{x^2+1}\right) \\ \cosh^{-1} x &= \ln\left(x \pm \sqrt{x^2-1}\right) \\ \tanh^{-1} x &= \ln\left(\sqrt{\frac{1+x}{1-x}}\right),\ x \in (-1,1) \end{aligned} \tag{3.7}$$

Chapter 4

Calculus and Elementary Analysis

4.1 Limits

Limit is given by

$$\lim_{x \to x_0} f(x) = K \tag{4.1}$$

if and only if for any $\varepsilon > 0$, $\exists \delta > 0$ such that

$$|f(x) - K| < \varepsilon \text{ for all } 0 < |x - x_0| < \delta. \tag{4.2}$$

Limit at infinity is given by

$$\lim_{x \to \infty} f(x) = K, \tag{4.3}$$

if and only if for any $\varepsilon > 0$, $\exists X < \infty$ such that

$$|f(x) - K| < \varepsilon \text{ for all } x > X. \tag{4.4}$$

Algebra of limits

$$\begin{aligned}
\lim(f(x) \pm g(x)) &= \lim f(x) \pm \lim g(x) \\
\lim(f(x)g(x)) &= \lim f(x) \lim g(x) \\
\lim\left(\frac{f(x)}{g(x)}\right) &= \begin{cases} \frac{\lim f(x)}{\lim g(x)} \text{ if } \lim g(x) \neq 0 \\ \text{Does not exist if } \lim f(x) \neq 0, \lim g(x) = 0 \\ \text{May exist if } \lim f(x) = 0, \lim g(x) = 0 \\ \text{May exist if } \lim f(x) = \infty, \lim g(x) = \infty \end{cases} \\
\lim[f(g(x))] &= f(\lim g(x)) \\
\lim[(f(x))^{g(x)}] &= (\lim f(x))^{\lim g(x)}
\end{aligned} \tag{4.5}$$

l'Hôpital's rule is for limits that are $0/0$ or ∞/∞.

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)} \tag{4.6}$$

Continuity at $x = a$ is defined by:

$$\begin{aligned}
&f(a) \text{ exists} \\
&\lim_{x \to a} f(x) = f(a)
\end{aligned} \tag{4.7}$$

Or alternatively using the (ε, δ) formalism: A real function $f(x)$ is continuous if for any $\varepsilon > 0$, $\exists \delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ for all $|x - a| < \delta$.

Differentiability at point $(x, f(x))$ is defined by:

$$\begin{aligned} &\text{the function must be continuous.} \\ &\text{the derivative } f' \text{ exists if it is finite and defined.} \\ &\Rightarrow \text{ the left} - \text{and} - \text{right} - \text{hand limits must be the same} \\ &\lim_{h\to 0}\left(\frac{f(x)-f(x-h)}{h}\right) = \lim_{h\to 0}\left(\frac{f(x+h)-f(x)}{h}\right). \end{aligned} \tag{4.8}$$

4.2 Differentiation

If we have a function of a single variable, $y = y(x)$, the derivative of y with respect to x is defined by the limiting process

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lim_{\delta x\to 0}\frac{y(x+\delta x)-y(x)}{\delta x} = \lim_{\delta x\to 0}\frac{\delta y}{\delta x}. \tag{4.9}$$

Chain rule is a way to differentiate a function of a function,

$$\frac{\mathrm{d}}{\mathrm{d}x}(f(u(x))) = \frac{\mathrm{d}f}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x}. \tag{4.10}$$

If the function is of multiple variables, such as $h = h(u(x,y), v(x,y))$, then

$$\frac{\partial h}{\partial x} = \frac{\partial h}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial h}{\partial v}\frac{\partial v}{\partial x}. \tag{4.11}$$

The product rule: if our function is a product of two independent functions, $y(x) = f(x)g(x)$,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}f}{\mathrm{d}x}g + f\frac{\mathrm{d}g}{\mathrm{d}x}. \tag{4.12}$$

Implicit differentiation: if $g(y) = h(x)$, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}h(x)}{\mathrm{d}x}\Big/\frac{\mathrm{d}g(y)}{\mathrm{d}y}. \tag{4.13}$$

Reciprocal rule: it seems trivial but actually not (for reasons that I do not understand).

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{\frac{\mathrm{d}y}{\mathrm{d}x}} \tag{4.14}$$

Leibnitz's formula is a great way to calculate high order differentiation.

$$(fg)^n = \sum_{m=0}^{n}\binom{n}{m}f^{(n-m)}g^{(m)}, \quad \text{where} \quad \binom{n}{m} = \frac{n!}{(n-m)!m!}, \tag{4.15}$$

note that

$$\binom{n}{m} + \binom{n}{m-1} = \binom{n+1}{m}. \tag{4.16}$$

We don't need to review stationary points because we do that in physics every day.

When stretching a graph, be aware of a number of things:

- Continuity and limits.
- Stationary points and intersections with axis.

4.3 Notation for approximation

Big O notation expresses the manner in which a function approaches a point (or infinity). For $f(x), g(x) \in \mathbb{R}$,

$$f(x) = \mathcal{O}(g(x)) \text{ as } x \to a, \tag{4.17}$$

if and only of $\exists$ constants ε, $K > 0$ such that

$$|f(x)| \leq K|g(x)| \text{ for all } |x - a| < \varepsilon. \tag{4.18}$$

Similar things can be done at infinity, for $f(x), g(x) \in \mathbb{R}$,

$$f(x) = \mathcal{O}(g(x)) \text{ as } x \to \infty, \tag{4.19}$$

if and only if $\exists$ constants $X, K > 0$ such that

$$|f(x)| \leq K|g(x)| \text{ for all } x > X. \tag{4.20}$$

For example, $x^2 + x = \mathcal{O}(x^2)$ as $x \to \infty$. We usually state the tightest upper bound.
Small o notation

Chapter 5

Series

5.1 Convergence

Given a sequence with an infinite number of terms, we define the n^{th} partial sum

$$S_n = \sum_{k=0}^{n} u_k \tag{5.1}$$

such that $S_0 = u_0$, $S_1 = u_0 + u_1$, etc.
Infinite series may **converge**, **diverge** or **oscillate**. If $\lim_{n\to\infty} S_n = S$, then for any ε, there exists a N such that

$$|S - S_n| < \varepsilon \quad \forall n > N, \tag{5.2}$$

the series is said to converge to S.
A **necessary condition** for the series to converge is that

$$u_k \to 0 \text{ as } k \to \infty \tag{5.3}$$

If $\sum_{k=0}^{\infty} |u_k|$ converges, the series is **absolutely convergent** and the original series is **necessarily convergent**, note that this applies even if $u_k \in \mathbb{C}$.
If $\sum_{k=0}^{\infty} u_k$ converges but $\sum_{k=0}^{\infty} |u_k|$ does not, then the series is **conditionally convergent**.
Harmonic series diverges:

$$\sum_{k=1}^{\infty} \frac{1}{k}. \tag{5.4}$$

Alternating harmonic series converges conditionally:

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = \ln 2. \tag{5.5}$$

Geometric progression

$$S_n = \sum_{k=0}^{n-1} r^k = \frac{1 - r^n}{1 - r} \tag{5.6}$$

5.2 Convergence tests

The comparison test and ratio test are for positive series, while the Leibnitz' criterion is used for alternating series. **Comparison test** is too trivial to mention.

Ratio test:

$$\begin{aligned} \lim_{k\to\infty} \frac{u_{k+1}}{u_k} < 1, &\quad \sum_k u_k \text{ converges} \\ \lim_{k\to\infty} \frac{u_{k+1}}{u_k} > 1, &\quad \sum_k u_k \text{ diverges} \\ \lim_{k\to\infty} \frac{u_{k+1}}{u_k} = 1, &\quad \sum_k u_k \text{ may converge} \end{aligned} \tag{5.7}$$

Leibnitz criterion: Alternating series have terms with alternating signs. This enhances converges. $S = \sum_k (-1)^{k+1} a_k$ with $a_k > 0$ converges if a_k is monotonically decreasing for large enough k and $\lim_{k\to\infty} a_k = 0$.

5.3 Taylor series

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n + R_{n+1}. \tag{5.8}$$

where

$$R_{n+1} = \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(\xi), \text{ where } \xi \in (a,x). \tag{5.9}$$

Taylor series about the origin is often called **Maclaurin series**, in which the remainder is

$$R_{n+1} = \frac{1}{(n+1)!}\left(\frac{1}{2}\right)^{n+1} \exp(\xi). \tag{5.10}$$

5.4 The binomial expansion

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)\cdots(\alpha-n+1)}{n!}x^n + \cdots . \tag{5.11}$$

Specifically, the term of order $N+1$ is:

$$a_{N+1}x^{N+1} = \frac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)\cdots(\alpha-N+1)(\alpha-N)}{(N+1)!}x^{N+1}. \tag{5.12}$$

Note that this is the general expansion, in which $\alpha \in \mathbb{R}$.

Chapter 6

Integration

6.1 Fundamental theorem of calculus

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(u)\mathrm{d}u = f(x). \tag{6.1}$$

Note the difference between **definite integrals** and **indefinite integrals**.
Partial fractions: just an example,

$$\frac{1}{x^2+x} = \frac{\alpha}{x} + \frac{\beta}{x+1} \quad \Rightarrow \quad \alpha = 1,\ \beta = -1. \tag{6.2}$$

6.2 Substitution

Substitution needs practice in order to spot the best route. There are some classic substitutions.

$$\begin{aligned} \sin x\mathrm{d}x &= -\mathrm{d}(\cos x) \\ \cos x\mathrm{d}x &= \mathrm{d}(\sin x) \\ \frac{1}{\cos^2 x}\mathrm{d}x &= \mathrm{d}(\tan x) \\ \frac{1}{1+x^2}\mathrm{d}x &= \mathrm{d}(\arctan x) \end{aligned} \tag{6.3}$$

Using half-angle formula

$$\begin{aligned} t &= \tan\frac{x}{2} \\ \sin x &= \frac{2t}{1+t^2} \\ \cos x &= \frac{1-t^2}{1+t^2} \\ \tan x &= \frac{2t}{1-t^2} \end{aligned} \tag{6.4}$$

Common substitutions on the **denominator** are:

Denominator involves	Substitution	Comments
a^2+x^2	$x=a\tan\theta$	
$\sqrt{a^2-x^2}$	$x=a\sin\theta$ or $x=a\cos\theta$	Need $\lvert x\rvert<\lvert a\rvert$
$\sqrt{x^2-a^2}$	$x=a\cosh\theta$	Need $\lvert x\rvert>\lvert a\rvert$
$\sqrt{a^2+x^2}$	$x=a\sinh u$	
a^2-x^2	$x=a\tanh u$	Need $\lvert x\rvert<\lvert a\rvert$
a^2-x^2	$x=a\coth u$	Need $\lvert x\rvert>\lvert a\rvert$
$a^2-x^2=(a+x)(a-x)$	Use partial fractions	
$a+bx+cx^2$	Factorise and partial fractions or **complete the square**	
$\sqrt{a+bx+cx^2}$	Complete the square	

6.3 Integration by parts

$$\frac{\mathrm{d}}{\mathrm{d}x}(fg)=\frac{\mathrm{d}f}{\mathrm{d}x}g+f\frac{\mathrm{d}g}{\mathrm{d}x}. \tag{6.5}$$

Rearrange and it gives

$$\int_a^b f\frac{\mathrm{d}g}{\mathrm{d}x}\mathrm{d}x=[fg]_a^b-\int_a^b\frac{\mathrm{d}f}{\mathrm{d}x}g\mathrm{d}x \tag{6.6}$$

6.4 Differentiation of integrals

For first year students the key is to understand how to evaluate rather than derive the expression:

$$\frac{\mathrm{d}I}{\mathrm{d}q}=\int_{a(q)}^{b(q)}\frac{\partial f}{\partial q}\mathrm{d}x+f(b(q);q)\frac{\mathrm{d}b}{\mathrm{d}q}-f(a(q);q)\frac{\mathrm{d}a}{\mathrm{d}q}. \tag{6.7}$$

Differentiation with respect to the parameter: using partial derivative to differentiate a multiple variable function.

$$\begin{aligned} I(q,a(q),b(q))&=\int_{a(q)}^{b(q)}f(x,q)\mathrm{d}x\\ \frac{\mathrm{d}I}{\mathrm{d}q}&=\frac{\partial I}{\partial q}+\frac{\partial I}{\partial a}\frac{\mathrm{d}a}{\mathrm{d}q}+\frac{\partial I}{\partial b}\frac{\mathrm{d}b}{\mathrm{d}q}\\ &=\int_a^b\frac{\partial f}{\partial q}\mathrm{d}x+\frac{\mathrm{d}b(q)}{\mathrm{d}q}f(b(q);q)-\frac{\mathrm{d}a(q)}{\mathrm{d}q}f(a(q);q) \end{aligned} \tag{6.8}$$

Stirling's approximation is a good way to approximate sums as integrals.

$$\ln n!\simeq n\ln n-n=n(\ln n-1). \tag{6.9}$$

This is a good approximation with a fractional error

$$\varepsilon=\frac{\ln n!-n(\ln n-1)}{\ln n!}\to\mathcal{O}(1/n)\text{ as }n\to\infty. \tag{6.10}$$

6.5 Schwarz's inequality

Schwarz's inequality is a good way to compare things in statistics.

$$\left(\int_a^b P(x)Q(x)\mathrm{d}x\right)^2\le\int_a^b\left(P(x)\right)^2\mathrm{d}x\int_a^b\left(Q(x)\right)^2\mathrm{d}x \tag{6.11}$$

Multiple integral is integrals on a surface, volume, etc. Think about Gauss's Theorem if you have trouble visualizing it.

The Gaussian integral:

$$\int_0^\infty \mathrm{e}^{-ax^2+bx}\mathrm{d}x=\sqrt{\frac{\pi}{4a}}\mathrm{e}^{\frac{b^2}{4a}}. \tag{6.12}$$

Chapter 7

Elementary Probability Theory

7.1 Sets

The set of all possible outcomes of the experiment is called the **sample space**, S.

An **event** A is a **subset** of S, $A \subset S$. $A \cap B$ is the **intersection** of A and B; $A \cup B$ is the **union** of A and B; $\bar{A}$ is the **complement** of A; and an **empty set** is denoted by $\varnothing$.

$$\begin{aligned} A \cap B &= B \cap A \\ A \cup B &= B \cup A \\ (A \cap B) \cap C &= A \cap (B \cap C) = A \cap B \cap C \\ (A \cup B) \cup C &= A \cup (B \cup C) = A \cup B \cup C \\ A \cap (B \cup C) &= (A \cap B) \cup (B \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \end{aligned} \tag{7.1}$$

7.2 Probability

Probability of events follow these properties:

$$\begin{aligned} &0 \leq P(A) \leq 1 \\ &P(\bar{A}) = 1 - P(A) \\ &P(A \cup B) = P(A) + P(B) - P(A \cap B) \\ &P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned} \tag{7.2}$$

Conditional probability

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{7.3}$$

Bayes' Theorem is just a rearranged form of total probability theorem.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{7.4}$$

7.3 Permutations and combinations

Permutations is the total number of arrangements when the order matters.

$${}^{n}P_{m} = \frac{n!}{(n-m)!} \tag{7.5}$$

Combinations is the total number of arrangements when the order doesn't matter.

$$ {}^nC_m = \frac{n!}{m!(n-m)!} = \frac{{}^nP_m}{m!} \tag{7.6} $$

Combinations has a number of properties, some of them can be found in Yang Hui's triangle.

$$ {}^{n-1}C_m + {}^{n-1}C_{m-1} = {}^nC_m \tag{7.7} $$

7.4 Probability distributions

A random variable is a variable whose value is determined by the outcome of an 'experiment'. For example, suppose that we toss a coin three times and X is the number of heads.The variable X is called **discrete** because it can only take discrete values $0, 1, 2, 3$. in general, the discrete variable X can takes the values x_i with probability p_i for $i = 0, 1, 2, \cdots, n-1$. Then we can define a **probability function** $f(x)$ to be the probability that X takes the value x, the following function should be normalized.

$$ f(x) = P(X = x) = \begin{cases} p_i, \text{ if } x = x_i \\ 0, \text{ if } x \neq x_i \end{cases} . \tag{7.8} $$

Cumulative probability function states the probability that X takes a value which is less or equal to x,

$$ F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i). \tag{7.9} $$

The **mean**, or **expectation value**, is defined by

$$ E(X) = \sum_i x_i p_i. \tag{7.10} $$

The **standard deviation**, $D(x)$, is defined by the square root of **variance**:

$$ D^2(X) = E\left[(X - E(X))^2\right] = \sum_i (x_i - E(X))^2 p_i \tag{7.11} $$

Note that

$$ D^2(X) = E\left(X^2\right) - E^2(X). \tag{7.12} $$

7.5 Binomial distribution

$$ \begin{aligned} P(X = r) = f(r) &= {}^nC_r p^r (1-p)^{n-r} \\ E(X) &= \sum_{r=0}^{n} r \left[{}^nC_r p^r (1-p)^{n-r}\right] \\ &= \sum_{r=1}^{n} \frac{n!}{(r-1)!(n-r)!} p^r (1-p)^{n-r} \\ &= n \sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!} p^r (1-p)^{n-r} \\ &= np \sum_{r-1=0}^{n-1} \frac{(n-1)!}{(r-1)!(n-1-(r-1))!} p^{r-1} (1-p)^{n-1-(r-1)} \\ &= np(1-p+p)^{n-1} = np \end{aligned} \tag{7.13} $$

use similar ways for two times,

$$ E\left(X^2\right) = n(n-1)p^2 + np, \tag{7.14} $$

then

$$ D^2(X) = np(1-p). \tag{7.15} $$

7.6 Poisson distribution

$$P(X = r) = \frac{\lambda^r \mathrm{e}^{-\lambda}}{r!} \tag{7.16}$$

The mean is

$$\begin{aligned} E(X) &= \sum_{r=0}^{\infty} r f(r) \\ &= \exp(-\lambda) \sum_{r=1}^{\infty} \frac{\lambda^r}{(r-1)!} \\ &= \lambda \exp(-\lambda) \sum_{r=1}^{\infty} \frac{\lambda^{r-1}}{(r-1)!} \\ &= \lambda \exp(-\lambda) \exp(\lambda) = \lambda \end{aligned} \tag{7.17}$$

It is easy to find, using similar way

$$E\left(X^2 - X\right) = E(X(X-1)) = \lambda^2 \tag{7.18}$$

so

$$E\left(X^2\right) = \lambda^2 + \lambda \quad \Rightarrow \quad D^2(X) = \lambda. \tag{7.19}$$

7.7 Normal distribution

$$N(\mu, \sigma^2): \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2\right] \tag{7.20}$$

Part II

Lent Term

This note was, again, not in my plan, but rather a tool to develop live tex skills in practice. The note contains basic principles taught in the IA Natural Sciences Mathematics B Lent term course and, I promise, is a lot more detailed than the last one.

Chapter 8

Ordinary Differential Equations

8.1 Introduction

8.1.1 Examples

Find velocity $v(t)$ of a falling body of mass m under gravity (g) with a drag force $\propto v^2$ (with a constant proportion μ). Applying Newton's second law, $F = ma$,

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = mg - \mu v^2, \tag{8.1}$$

A second example is about concentrations in chemical reactions A + B = C:

$$\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{\mathrm{d}b}{\mathrm{d}t} = -\frac{\mathrm{d}c}{\mathrm{d}t} = \kappa ab \tag{8.2}$$

8.1.2 Classification

Standard notation: $y(x)$ is unknown function of x, where x is a independent variable and y is dependent variable. The nth derivative of $y(x)$ with respect to x will be written as $y^{(n)}(x)$.

The **general form** of the first-order ODE is

$$F(y', y, x) = 0. \tag{8.3}$$

If we can solve for y', we write

$$y' = f(x, y). \tag{8.4}$$

and solve y in the end.

Obviously, the nth order ODE

$$F(y^{(n)}, \cdots, y', y, x) = 0 \tag{8.5}$$

can be solved by reducing the total (highest) order of differentiation.

The second example we showed is a system of ODEs instead of a single ODE. Note that

$$\frac{\mathrm{d}b}{\mathrm{d}t} - \frac{\mathrm{d}a}{\mathrm{d}t} = 0, \tag{8.6}$$

therefore $(b-a)$ is a constant that can be obtained from any state within the validity of our equations. And, we can replace b by $a - a_0 + b_0$ and construct an equation for a and a alone.

ODEs can be solved analytical of numerical methods. How this works was a major part of the MIT course.

8.2 First-Order ODEs

8.2.1 Directly integrable (exact) equations

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x). \tag{8.7}$$

If we can find the integral of $f(x)$ over x the equation is solved. Note that the integral should be indefinite integral and the constant can be determined using any known state (y_0, x_0), not necessarily the initial state!

If we cannot determine the constant, our results would be a field with unit flux magnitude (but not direction) everywhere.

A more general form of ODEs is

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(y, x), \tag{8.8}$$

where y and x may be separable or not.

8.2.2 Separable equations

These are equations that can be written in the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{f(x)}{g(y)}. \tag{8.9}$$

Multiply both sides with $g(y)$ and integrate in x,

$$\int g(y)\mathrm{d}y = \int f(x)\mathrm{d}x. \tag{8.10}$$

Again, it is crucial to determine the constant. Another way to determine the constant is, if we know a state (y_0, x_0), we can do

$$\int_{y_0}^{y} g(y)\mathrm{d}y = \int_{x_0}^{x} f(x)\mathrm{d}x, \tag{8.11}$$

but this is not a uniform way.

Note: What will happen at the point $g(y) = 0$?

8.2.3 Initial conditions

Our constant of integration leads to different solutions, the solution including the arbitrary constant is called **general solution** (GS).

8.2.4 Linear differential equations

A linear ODE is

$$\frac{\mathrm{d}y}{\mathrm{d}x} + p(x)y = f(x), \tag{8.12}$$

where y and y' appear linearly. If $f(x) = 0$, the equation is homogeneous but if $f(x) \neq 0$, the equation is not homogeneous.

The homogeneous linear equation is clearly separable and will always give a solution in form of $C\mathrm{e}^{-P(x)}$ where $P(x)$ is the antiderivative of $p(x)$ and C is an arbitrary constant.

Now we focus on the non-homogeneous case. It can be integrated if we multiply by a suitable function $\mu(x)$ (integration factor).

$$\begin{aligned} \mu(x)\frac{\mathrm{d}y}{\mathrm{d}x} + \mu(x)p(x)y &= \mu(x)f(x) \\ \frac{\mathrm{d}}{\mathrm{d}x}\left[y\mu(x)\right] &= \mu(x)f(x) \end{aligned} \tag{8.13}$$

Observing the equation, it is clear that we want to choose a $\mu(x)$ that satisfies

$$\frac{\mathrm{d}\mu}{\mathrm{d}x} = \mu(x)p(x), \tag{8.14}$$

which is a separable ODE whose general solution is

$$\mu(x) = \exp\left[\int p(x)\mathrm{d}x\right]. \tag{8.15}$$

Note

- $\mu(x)$ is reciprocal of the solution of the homogeneous equation.

- Do not include an arbitrary multiplication constant, it does not matter which constant we choose.

Having out choice of μ, the non-homogeneous equation looks like

$$y\mu(x) = \int \mu(x)f(x)\mathrm{d}x \tag{8.16}$$

which gives

$$y = \frac{1}{\mu(x)}\int \mu(x')f(x')\mathrm{d}x'. \tag{8.17}$$

Quick Trick When $p(x)$ is a constant α, $\mu(x) = \mathrm{e}^{\alpha x}$.

8.2.5 Solution by substitution

The goal is to find substitution that reduces ODE to a separable or linear ODE.

We will talk about two kinds of equations as examples: homogeneous equation and Bernoulli's equation.

Homogeneous equation:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f\left(\frac{y}{x}\right) \tag{8.18}$$

Substitute y/x by u and it gives

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{f(u)-u}{x}, \tag{8.19}$$

which is a separable equation. But keep an eye on the values of x, especially when it is equal to or less than 0.

Bernoulli's equation:

$$\frac{\mathrm{d}y}{\mathrm{d}x} + p(x)y = q(x)y^n, \quad n \in \mathbb{Z},\ n \neq 0, 1. \tag{8.20}$$

Use substitution $z = y^{1-n}$ and it gives

$$\frac{\mathrm{d}z}{\mathrm{d}x} = (1-n)y^n\frac{\mathrm{d}y}{\mathrm{d}x} = (1-n)y^n\left[-p(x)y + q(x)y^n\right] \tag{8.21}$$

therefore,

$$\frac{\mathrm{d}z}{\mathrm{d}x} = (1-n)\left[-p(x)z + q(x)\right] \tag{8.22}$$

which is a linear equation that can be solved.

Example Bernoulli's equation with $n = 2$:

$$\frac{\mathrm{d}y}{\mathrm{d}x} + xy = x^3y^2. \tag{8.23}$$

Use substitution $z = 1/y$,

$$\frac{\mathrm{d}z}{\mathrm{d}x} = -\frac{1}{y^2}\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{-1}{y^2}\left[-xy + x^3y^2\right] = xz - x^3, \tag{8.24}$$

which is a linear inhomogeneous equation with $p(x) = -x$, $f(x) = x^3$. The integration factor is therefore

$$\mu(x) = \exp\left[\int -x\mathrm{d}x\right] = \exp\left[-\frac{x^2}{2}\right]. \tag{8.25}$$

As mentioned before, the constant in integration factor can be ignored. Multiply our ODE by $\mu(x)$,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\exp\left[-\frac{x^2}{2}\right]z\right] = -\exp\left[-\frac{x^2}{2}\right]x^3. \tag{8.26}$$

Integrate by parts:

$$z = x^2 + 2 + C = x^2 + C\exp\left(\frac{x^2}{2}\right) + 2 \quad \Rightarrow \quad y = \frac{1}{x^2 + C\exp\left(\frac{x^2}{2}\right) + 2} \tag{8.27}$$

Example

$$\frac{\mathrm{d}y}{\mathrm{d}x} = F(ax + by + c) \tag{8.28}$$

suggests the substitution $u(x) = ax + by + c$,

$$\frac{\mathrm{d}u}{\mathrm{d}x} = a + b\frac{\mathrm{d}y}{\mathrm{d}x} = a + bF(u), \tag{8.29}$$

which is separable.

Example

$$\frac{\mathrm{d}y}{\mathrm{d}x} = yf(\mathrm{e}^{\alpha x}y^{\beta}),\ y > 0. \tag{8.30}$$

suggests substitution $u = \ln y$

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}(\ln y) = f\left(\mathrm{e}^{\alpha x + \beta \ln y}\right). \tag{8.31}$$

Introduce new notation: $g(z) = f(\mathrm{e}^x)$, we find

$$\frac{\mathrm{d}u}{\mathrm{d}x} = g(\alpha x + \beta u), \tag{8.32}$$

which goes back to the previous example.

Example

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{y^4\mathrm{e}^x}{y^3\mathrm{e}^x + 1},\ y > 0. \tag{8.33}$$

Use knowledge from the last example, the solution should be

$$\frac{\mathrm{e}^x}{y} = -\frac{1}{4y^4} + C. \tag{8.34}$$

8.2.6 Additional remarks

Not all non-linear equations can be solved analytically, such as $y\prime = x + y^3$, but we can always interprate ODE geometrically. $F(y\prime, y, x)$ defines the slope of the solution curve on x-y plane, that returned to what we learned in the MIT course.

For example, consider a chemical reaction $A \to B \to C$,

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\kappa_a a, \quad \frac{\mathrm{d}b}{\mathrm{d}t} = \kappa_a a - \kappa_b b, \quad \frac{\mathrm{d}c}{\mathrm{d}t} = \kappa_b b. \tag{8.35}$$

We can solve this equation by substitutions but can also use numerical methods to plot the functions.

Now consider A and B only, the first equation reads

$$a(t) = a_0 \exp(-k_a t), \tag{8.36}$$

where $a_0 = a(0)$. Substitute into the equation of $b(t)$:

$$\frac{\mathrm{d}b}{\mathrm{d}t} + k_b b = k_a a_0 \exp(-k_a t). \tag{8.37}$$

Considering the fact that $b(0) = 0$,

$$b(t) = \frac{k_a a_0}{k_b - k_a}[\exp(-k_a t) - \exp(-k_b t)]. \tag{8.38}$$

At large times one of the exponential terms dominates and $b(t)$ decays exponentially at the slower of the rates k_a and k_b. The ratio is given by

$$\frac{b}{a} = \frac{k_a}{k_b - k_a}[1 - \exp((k_a - k_b)t)], \tag{8.39}$$

which is independent of a_0 and therefore can be used to measure the formation time of rocks using the ratio of linked isotopes.

$$t = -\left(\frac{1}{k_b - k_a}\right) \ln\left[1 - \left(\frac{k_b - k_a}{k_a}\right)\frac{b}{a}\right] \tag{8.40}$$

8.3 Second-Order ODEs

8.3.1 Linear second-order ODEs

The general equation is

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + p(x)\frac{\mathrm{d}y}{\mathrm{d}x} + q(x)y = f(x). \tag{8.41}$$

Again, the equation is said to be homogeneous if $f(x) = 0$ and inhomogeneous if not.

We define a linear differential operator $\mathcal{L}$ which is

$$\mathcal{L} = \frac{\mathrm{d}^2}{\mathrm{d}x^2} + p(x)\frac{\mathrm{d}}{\mathrm{d}x} + q(x). \tag{8.42}$$

The operator is said to be linear because it satisfies:

$$\begin{aligned} \mathcal{L}(\alpha u) &= \alpha \mathcal{L}(u) \\ \mathcal{L}(u + v) &= \mathcal{L}(u) + \mathcal{L}(v) \\ \mathcal{L}(\alpha u + \beta v) &= \alpha \mathcal{L}(u) + \beta \mathcal{L}(v) \end{aligned} \tag{8.43}$$

8.3.2 Principle of superposition

If the equation is homogeneous, $\mathcal{L}y = 0$, and we have two solutions u and v, we can find the rest of solutions by superposition

$$\mathcal{L}(\alpha u + \beta v) = 0, \text{ for any } \alpha, \beta = \text{const}. \tag{8.44}$$

If the equation is inhomogeneous, $\mathcal{L}y = f$. A particular integral y_p is any solution of $\mathcal{L}y = f$ and the complimentary function y_c is the general solution of $\mathcal{L}y = 0$. The general solution of the inhomogeneous equation is

$$y = y_p + y_c. \tag{8.45}$$

Why? Just substitute $y_p + y_c$ into the equation

$$\mathcal{L}(y_p + y_c) = \mathcal{L}y_p + \mathcal{L}y_c = f + 0 = f. \tag{8.46}$$

but this does not prove that $y_p + y_c$ is the general solution[1]. Use contradiction method: assume that there exists some function $\bar{y}$ which is not belong to out set of general solution but satisfies the linear inhomogeneous equation. Write

$$\mathcal{L}(\bar{y} - y_p) = \mathcal{L}(\bar{y}) - \mathcal{L}y_p = f - f = 0 \tag{8.47}$$

and note that $\bar{y} - y_p = y_c$, which forms a contradiction. Therefore any solution of our equation can be written in the form of our general solution.

8.3.3 Homogeneous equation with constant coefficients

Let $p(x) = 2\gamma$, $q(x) = \omega_0^2$, which are both real constants.

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\gamma\frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2 x = 0 \tag{8.48}$$

Try to guess the form of the solution, it is clear that exponent is a good choice. Substitute $x = \mathrm{e}^{\lambda t}$,

$$\lambda^2 + 2\gamma\lambda + \omega_0^2 = 0. \tag{8.49}$$

We can ignore the exponent because $\mathrm{e}^{\lambda t}$ can not be zero at any time. The two solutions of λ are:

$$\lambda_{1,2} = -\gamma \pm \sqrt{\gamma^2 - \omega_0^2} \tag{8.50}$$

There are obviously three conditions that are related to three conditions in physics:

Light damping If $\gamma < \omega_0$, both λ are complex. Defining the constant $\omega = \sqrt{\gamma^2 - \omega_0^2}$, the general solution is:

$$x = \mathrm{e}^{-\gamma t}\left[(C_1 + C_2)\cos(\omega t) + \mathrm{i}(C_1 - C_2)\sin(\omega t)\right]. \tag{8.51}$$

In math we care only about the real part of the solution but in some cases we should consider both. Note that it is easy mistake to write

$$x \doteq C\mathrm{e}^{-\gamma t}\cos(\omega t + \phi) \quad \text{[Not always right]}, \tag{8.52}$$

this only holds in some particular cases where your initial condition of oscillation is very friendly with computing.

Critical damping If $\gamma = \omega_0$, the solution is surprisingly not just $x = A\mathrm{e}^{-\gamma t}$. That is because the $\lambda = -\gamma$ is a double solution, so we must find the second solution of this degenerate case. Try in the form $x = t\mathrm{e}^{\lambda t}$, which gives

$$(\lambda^2 + 2\gamma\lambda + \omega_0^2)t\mathrm{e}^{\lambda t} + (2\lambda + 2\gamma)\mathrm{e}^{\lambda t} \doteq 0 \tag{8.53}$$

Note that $\lambda = -\gamma$, the equation holds. Therefore the general solution is

$$x = (C_1 + C_2 t)\mathrm{e}^{-\gamma t} \tag{8.54}$$

Why does $t\mathrm{e}^{\lambda t}$ work? Let's assume that $\lambda_1 \neq \lambda_2$ but choose $\lambda_{1,2} = -\gamma \pm \varepsilon$. The solution is therefore

$$\begin{aligned} x &= C_1\mathrm{e}^{(-\gamma+\varepsilon)t} + C_2\mathrm{e}^{(-\gamma-\varepsilon)t} \\ &= \mathrm{e}^{-\gamma t}\left[C_1\mathrm{e}^{\varepsilon t} + C_2\mathrm{e}^{-\varepsilon t}\right] \\ &= \mathrm{e}^{-\gamma t}\left[(C_1 + C_2) + t\varepsilon(C_1 - C_2) + \frac{t^2\varepsilon^2}{2}(C_1 + C_2) + \cdots\right] \\ &= \mathrm{e}^{-\gamma t}\left[C_1' + tC_2' + \mathcal{O}(\varepsilon^2)\right] \end{aligned} \tag{8.55}$$

[1] Any solution can be written in this form

Over damping If $\gamma > \omega_0$, both λ are real, therefore the general solution of our equation is

$$x = C_1 \mathrm{e}^{\lambda_1 t} + C_2 \mathrm{e}^{\lambda_2 t}. \tag{8.56}$$

No damping If $\gamma = 0$, there is no damping acting on the system and it will naturally give simple harmonic motion.

No motion If $\lambda_{1,2} = 0$, the particle, or whatever, is just stationary or moving at a constant speed. Obviously, "No motion" means there exists a inertial frame in which the particle does not move at all.

8.3.4 Initial (boundary) conditions

We must have both $x(t_0)$ and $\frac{\mathrm{d}x}{\mathrm{d}t}\Big|_{t_0}$ to determine the actual solution of our second order ODE. Boundary conditions can be satisfied at all times, but should also present in similar forms.

8.3.5 Inhomogeneous equation with constant coefficients

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + 2\gamma \frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2 x = f(t). \tag{8.57}$$

To solve this, we have to find x_p, which has something to do with $f(t)$, the driven force[2]. The shortcut is to consider properties of $\mathcal{L}$.

$$\begin{aligned}
\mathcal{L}1 &= \omega_0^2 \\
\mathcal{L}t &= 2\gamma + \omega + 0^2 t \\
\mathcal{L}t^2 &= 2 + 4\gamma t + \omega_0^2 t^2 \\
\mathcal{L}t^n &= n(n-1)t^{n-2} + 2\gamma n t^{n-1} + \omega_0^2 t^n \\
\mathcal{L}\mathrm{e}^{\alpha t} &= (t^2 + 2\gamma t + \omega_0^2)\mathrm{e}^{\alpha t} \\
\mathcal{L}\cos(\omega t) &= (\omega_0^2 - \omega^2)\cos(\omega t) - 2\gamma t \sin(\omega t) \\
\mathcal{L}\sin(\omega t) &= (\omega_0^2 - \omega^2)\sin(\omega t) + 2\gamma t \cos(\omega t)
\end{aligned} \tag{8.58}$$

Also, principle of superposition, $\mathcal{L}(\alpha x) = \alpha f(t)$, tells us that the solution depends linearly on forcing. Additionally, we can construct different kinds of functions at the LHS to meet the $f(t)$ at the RHS.

Note that, when applying $\mathcal{L}$ to a multiplication of two functions, there exists a cross product term

$$\begin{aligned}
\mathcal{L}(fg) &= \frac{\mathrm{d}^2}{\mathrm{d}t^2}(fg) + 2\gamma \frac{\mathrm{d}}{\mathrm{d}t}(fg) + \omega_0^2 (fg) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}f}{\mathrm{d}t}g + f\frac{\mathrm{d}g}{\mathrm{d}t}\right) + 2\gamma\left(\frac{\mathrm{d}f}{\mathrm{d}t}g + f\frac{\mathrm{d}g}{\mathrm{d}t}\right) + \omega_0^2 fg \\
&= \frac{\mathrm{d}^2 f}{\mathrm{d}t^2}g + f\frac{\mathrm{d}^2 g}{\mathrm{d}t^2} + 2\frac{\mathrm{d}f}{\mathrm{d}t}\frac{\mathrm{d}g}{\mathrm{d}t} + 2\gamma\left(\frac{\mathrm{d}f}{\mathrm{d}t}g + f\frac{\mathrm{d}g}{\mathrm{d}t}\right) + \omega_0^2 fg \\
&= \mathcal{L}(f)g + f\mathcal{L}(g) + 2\frac{\mathrm{d}f}{\mathrm{d}t}\frac{\mathrm{d}g}{\mathrm{d}t} - \omega_0^2 fg
\end{aligned} \tag{8.59}$$

Back to what we use for y_p to solve the equation, here is a summary:

$f(t)$	Functions to try (x_p)
Polynomial of degree n	Polynomial of degree n (Do not forget the constant at the end!)
$C\mathrm{e}^{kt}$	$d\mathrm{e}^{kt}$
$C_1\cos(\omega t) + C_2 \sin(\omega t)$	$D_1\cos(\omega t) + D_2 \sin(\omega t)$
Linear combination of all above	Linear combination of all above (Principle of superposition)

Sometimes these fail, for example in the $f(t) = C\mathrm{e}^{kt}$ case, if $\mathcal{L}(\mathrm{e}^{kt}) = 0$ we can not use our x_p anymore, instead, just shift an order and try $dx\mathrm{e}^{kt}$. If it fails again, shift order for another time.

[2]Note that here we are taking a standard approximation that elastic wave transmits at a speed of infinity.

Summary We can solve a inhomogeneous equation with constant coefficients following three steps:

1. Find the general solution of the corresponding homogeneous equation.
2. Find y_p and find the general solution of the original inhomogeneous equation.
3. Identify two constants using boundary or initial conditions.

8.4 Connections of First and Second-Order ODEs

For a first-order equation, it can also be solved by finding y_c and y_p. It is obvious in the view of computers.

For second-order equation, they can be solved by finding the complimentary integral.

$$\begin{aligned} \frac{\mathrm{d}^2y}{\mathrm{d}x^2} + 2a\frac{\mathrm{d}y}{\mathrm{d}x} + by &= f(x) \\ \lambda^2 + 2a\lambda + b = (\lambda - \lambda_1)(\lambda - \lambda_2) &= 0 \\ \Rightarrow \left(\frac{\mathrm{d}}{\mathrm{d}x} - \lambda_1\right)\left(\frac{\mathrm{d}}{\mathrm{d}x} - \lambda_2\right) y &= f(x) \end{aligned} \tag{8.60}$$

Therefore, if I introduce $z(x) = (\mathrm{D} - \lambda_2)y$, then

$$(\mathrm{D} - \lambda_1)z(x) = f(x). \tag{8.61}$$

What we have done is we replaced the second-order ODE with two coupled first-order ODEs. Solve for $z(x)$ and for y using integrating factors.

8.5 Driven Oscillator

A driven oscillator usually has a driven force $f(t) = f_0\cos(\omega t)$,

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\gamma\frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2 x = f_0\cos\omega t. \tag{8.62}$$

The characteristic equation is still $\lambda^2 + 2\gamma\lambda + \omega_0^2 = 0$ and the solutions $\lambda_{1,2}$ are the same. Suppose that our special solution has form $x_p = K\mathrm{e}^{\mathrm{i}\omega t}$, plug in and we have

$$x_p = \frac{f}{\sqrt{(\omega_0^2 - \omega^2)^2 - 4\gamma^2\omega^2}}\cos(\omega t - \phi), \quad \phi = \arctan\left(\frac{2\gamma\omega}{\omega_0^2 - \omega^2}\right). \tag{8.63}$$

Then the final solution is naturally

$$x(t) = A\mathrm{e}^{\lambda_1 t} + B\mathrm{e}^{\lambda_2 t} + \frac{f}{\sqrt{(\omega_0^2 - \omega^2)^2 - 4\gamma^2\omega^2}}\cos(\omega t - \phi), \quad \phi = \arctan\left(\frac{2\gamma\omega}{\omega_0^2 - \omega^2}\right). \tag{8.64}$$

You can discuss the variation of amplitude, phase difference and other things on your own.

8.6 nth-Order ODEs

8.6.1 General nth-order ODEs

You can always guess, once your guess is right, you have found the solution[3].

[3] This is obviously pure nonsense.

8.6.2 nth-order linear inhomogeneous ODEs with constant coefficients

A general nth-order ODE would be a complete mess, however, if it is in linear forms with constant coefficients, it is possible for us to write down a general analytical solution. Our equation is namely

$$\sum_{i=0}^{n} b_i \frac{\mathrm{d}^i y}{\mathrm{d}x^i} = f(x), \tag{8.65}$$

where b_i are constants. Again, we use the complementary function method, suppose $y_c = \mathrm{e}^{\alpha x}$, our homogeneous equation becomes

$$\sum_{i=0}^{n} b_i \alpha^i = 0. \tag{8.66}$$

There are naturally n solutions of α, these can be found by rearranging the equation to

$$\prod_{i=1}^{n} (\alpha - \alpha_i) = 0. \tag{8.67}$$

Then, for α_i that are all different and independent of each other,

$$y_c = \sum_{i=1}^{n} C_i \mathrm{e}^{\alpha_i x}. \tag{8.68}$$

For each root α_i that repeated k_i times:

$$y_c = \sum_{i=1}^{n} \left(\sum_{\ell=0}^{k_i} C_{i,\ell} x^{\ell-1} \right) \mathrm{e}^{\alpha_i x} \tag{8.69}$$

Then we can use the method of variation of parameters to find y_p, however, this is usually hard work.

Chapter 9

Multi-variable Functions

9.1 Introduction

A quantity can be a function of several variables, for example, electric potential is a function of position, $\phi = \phi(x, y, z)$. We introduce partial differentiation to deal with the change of these variables.

9.2 Partial Differentiation

The partial differentiation of $f(x_1, x_2, \cdots)$ with respect to x_i is defined by

$$\frac{\partial f}{\partial x^i} = \partial_i f = \lim_{\delta x^i \to 0} \frac{f(x^i + \delta x^i, \cdots) - f(x^i, \cdots)}{\delta x^i} \tag{9.1}$$

The gradient of $f(x_i)$, in Cartesian coordinates, is defined by

$$\nabla f(x_i) = \sum_i \frac{\partial f}{\partial x^i} \hat{i}. \tag{9.2}$$

Note that[1] this is a map $\mathbb{R} \longmapsto \mathbb{R}^n$.

For second-order differentiation (of smooth functions),

$$\partial_x \partial_y f = \partial_y \partial_x f. \tag{9.3}$$

which is great because we now know that partial derivatives commute.

Definition The function $f(x, y)$ is differentiable at point (x, y) if

$$\frac{\varepsilon}{\sqrt{\delta x^2 + \delta y^2}} \to 0 \quad \text{as} \quad \sqrt{\delta x^2 + \delta y^2} \to 0 \tag{9.4}$$

where

$$\varepsilon = f(x + \delta x, y + \delta y) - f(x, y) - \alpha \delta x - \beta \delta y, \quad \text{where} \quad \alpha, \beta = \text{const.} \tag{9.5}$$

The total derivative of $f(x, y)$ is

$$\mathrm{d}f(x, y) = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y + \mathcal{O}(\mathrm{d}x\mathrm{d}y). \tag{9.6}$$

Therefore, if x and y depend on another variable t,

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}. \tag{9.7}$$

One should be careful with the signs d and δ. δ is a sign for small, but not infinitesimal changes. From now on, d can only be used when it is representing the total derivative.

[1] See vector calculus notes.

9.2.1 Taylor series

Similar to single-variable calculus,

$$f(x+\delta x, y+\delta y) = f(x,y) + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\delta x^2 + \frac{\partial^2 f}{\partial x \partial y}\delta x \delta y + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}\delta y^2 + \cdots \tag{9.8}$$

Note that binomial theorem is used on the RHS.

9.2.2 Chain Rule

If we re-parameterise a function, it can be written as

$$f(x,y) = f(x(u,v), y(u,v)) = g(u,v). \tag{9.9}$$

Note that

$$\begin{aligned} \mathrm{d}x &= \frac{\partial x}{\partial u}\mathrm{d}u + \frac{\partial x}{\partial v}\mathrm{d}v \\ \mathrm{d}y &= \frac{\partial y}{\partial u}\mathrm{d}u + \frac{\partial y}{\partial v}\mathrm{d}v \end{aligned} \tag{9.10}$$

therefore

$$\mathrm{d}f = \left[\frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}\right]\mathrm{d}u + \left[\frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}\right]\mathrm{d}v. \tag{9.11}$$

A simpler case is that x and y are all parameterised by t, this gives

$$\mathrm{d}f = \left(\frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}\right)\mathrm{d}t. \tag{9.12}$$

9.2.3 Reciprocity and cyclic relations

Simply

$$\frac{\partial x}{\partial y}\frac{\partial y}{\partial x} = 1, \tag{9.13}$$

and

$$\frac{\partial x}{\partial z}\frac{\partial y}{\partial x}\frac{\partial z}{\partial y} = -1. \tag{9.14}$$

9.2.4 Exact differentials

The general expression

$$\omega = P(x,y)\mathrm{d}x + Q(x,y)\mathrm{d}y \tag{9.15}$$

is called a differential form in the variables x and y. We say that ω is an exact differential if there is a function $f(x,y)$ such that

$$\mathrm{d}f = P(x,y)\mathrm{d}x + Q(x,y)\mathrm{d}y. \tag{9.16}$$

Since we know that

$$\mathrm{d}f = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y, \tag{9.17}$$

we have the relation

$$\frac{\partial f}{\partial x} = P, \quad \frac{\partial f}{\partial y} = Q. \tag{9.18}$$

Using the symmetrical relation

$$\begin{aligned} \frac{\partial}{\partial y}\frac{\partial f}{\partial x} &= \frac{\partial}{\partial x}\frac{\partial f}{\partial y} \\ \frac{\partial P}{\partial y} &= \frac{\partial Q}{\partial x} \end{aligned} \tag{9.19}$$

Note that it works in the opposite direction. $P(x,y)\mathrm{d}x + Q(x,y)\mathrm{d}y$ is an example is an exact differential if and only if $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$.

9.2.5 Integrating factors for differential forms

The function $\mu(x,y)$ is called an integrating factor for the differential form $\omega = P(x,y)\mathrm{d}x + Q(x,y)\mathrm{d}y$ if the differential form

$$\mu(x,y)\left[P(x,y)\mathrm{d}x + Q(x,y)\mathrm{d}y\right] \text{ is exact.} \tag{9.20}$$

The condition for this is straightforward:

$$\frac{\partial}{\partial y}\mu P = \frac{\partial}{\partial x}\mu Q. \tag{9.21}$$

9.2.6 Maxwell's Relations

Partial derivatives and differentials are frequently encountered in thermodynamics, and the classic application is to gases. Suppose a unit mass of gas has pressure p, volume V and temperature T.

The internal energy is defined to be

$$\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V. \tag{9.22}$$

This differential relation is a statement of the first law of thermodynamics. In an infinitesimal process, the increase in internal energy is equal to the amount of heat added to the gas (by adding entropy) plus the amount of work done on it (by compressing it).

Regarding U as a function of (S,V),

$$\mathrm{d}U = \left(\frac{\partial U}{\partial S}\right)_V \mathrm{d}S + \left(\frac{\partial U}{\partial V}\right)_S \mathrm{d}V, \tag{9.23}$$

therefore

$$\left(\frac{\partial U}{\partial S}\right)_V = T, \quad \left(\frac{\partial U}{\partial V}\right)_V = -p. \tag{9.24}$$

Symmetry of partial differentiation gives

$$\frac{\partial^2 U}{\partial S \partial V} = \frac{\partial^2 U}{\partial V \partial S} \quad \Rightarrow \quad \left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial p}{\partial S}\right)_V \tag{9.25}$$

We can derive similar relations by making a change of variables known as a Legendre transformation.

The Helmholtz free energy is defined by $F = U - TS$, therefore

$$\mathrm{d}F = -S\mathrm{d}T - p\mathrm{d}V. \tag{9.26}$$

Similar equation can be derived:

$$\begin{gathered}\left(\frac{\partial F}{\partial T}\right)_V = -S, \quad \left(\frac{\partial F}{\partial V}\right)_T = -p \\ \left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V\end{gathered} \tag{9.27}$$

For enthalpy $H = U + pV$,

$$\begin{gathered}\mathrm{d}H = T\mathrm{d}S + V\mathrm{d}p \\ \left(\frac{\partial H}{\partial S}\right)_p = T, \quad \left(\frac{\partial H}{\partial p}\right)_S = V \\ \left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p\end{gathered} \tag{9.28}$$

For Gibbs free energy $G = H - TS$,

$$\begin{gathered}\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}p \\ \left(\frac{\partial G}{\partial T}\right)_p = -S \quad \left(\frac{\partial G}{\partial p}\right)_T = V \\ \left(\frac{\partial S}{\partial p}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_p\end{gathered} \tag{9.29}$$

Each Maxwell relations imply to others, for example, if we assume relation (9.27) then we can deduce (9.28)

$$\begin{aligned}\left(\frac{\partial S}{\partial V}\right)_T &= -\left(\frac{\partial S}{\partial T}\right)_V \left(\frac{\partial T}{\partial V}\right)_S \\ &= -\left(\frac{\partial S}{\partial p}\right)_V \left(\frac{\partial p}{\partial T}\right)_V \left(\frac{\partial T}{\partial V}\right)_S \\ &= -\left(\frac{\partial p}{\partial T}\right)_V \left(\frac{\partial T}{\partial V}\right)_S \Big/ \left(\frac{\partial p}{\partial S}\right)_V \\ &= \left(\frac{\partial p}{\partial T}\right)_V \left(\frac{\partial T}{\partial V}\right)_S \Big/ \left(\frac{\partial T}{\partial V}\right)_S \\ &= \left(\frac{\partial p}{\partial T}\right)_V .\end{aligned} \tag{9.30}$$

A different type of relation can also be derived from the equation $\mathrm{d}U = T\mathrm{d}S = p\mathrm{d}V$. By considering the two thermal variables (T, S) we are led to write

$$\begin{aligned}\mathrm{d}U &= T\mathrm{d}S - p\mathrm{d}V \\ &= T\mathrm{d}S - p\left[\left(\frac{\partial V}{\partial T}\right)_S \mathrm{d}T + \left(\frac{\partial V}{\partial S}\right)_T \mathrm{d}S\right] \\ &= -p\left(\frac{\partial V}{\partial T}\right)_S \mathrm{d}T + \left[T - p\left(\frac{\partial V}{\partial S}\right)_T\right]\mathrm{d}S.\end{aligned} \tag{9.31}$$

Symmetry of the mixed partial derivatives implies

$$\begin{gathered}\left(\frac{\partial}{\partial S}\right)_T \left[-p\left(\frac{\partial V}{\partial T}\right)_S\right] = \left(\frac{\partial}{\partial T}\right)_S \left[T - p\left(\frac{\partial V}{\partial S}\right)_T\right] \\ -\left(\frac{\partial p}{\partial S}\right)_T \left(\frac{\partial V}{\partial T}\right)_S - p\frac{\partial^2 V}{\partial S \partial T} = 1 - \left(\frac{\partial p}{\partial T}\right)_S \left(\frac{\partial V}{\partial S}\right)_T - p\frac{\partial^2 V}{\partial T \partial S} \\ \left(\frac{\partial p}{\partial T}\right)_S \left(\frac{\partial V}{\partial S}\right)_T - \left(\frac{\partial p}{\partial S}\right)_T \left(\frac{\partial V}{\partial T}\right)_S = 1\end{gathered} \tag{9.32}$$

9.3 Stationary Points

Equipped with techs above, we can finally go to our topic: Multi-variable functions.

9.3.1 Functions of one variable

Consider a function $f(x) : \mathbb{R} \longmapsto \mathbb{R}$, near any point, we can use Taylor expansion to approximate

$$f(x) = f(x_) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2, \tag{9.33}$$

provided that the first and second order derivatives exist. Local minimum and maximum are easy to define:

$$\begin{aligned}&x = x_0 \text{ is a local maximum if } f''(x_0) < 0 \\ &x = x_0 \text{ is a local minimum if } f''(x_0) > 0\end{aligned} \tag{9.34}$$

If both $f'(x_0)$ and $f'' = 0$ then we would need to inspect higher-order terms in the Taylor expansion to determine whether the function has a minimum, maximum or neither. For example, if $f''' \neq 0$, that indicates neither a minimum nor a maximum but rather a stationary point of inflection[2].

[2](In general, a point of inflection is a point on a curve at which the curvature changes sign. The curvature is $f''(1 - f'^2)^{-3/2}$

9.3.2 Functions of multi variables

Consider, for example a differentiable function $f(x, y)$ of two variables. Near any point (x_0, y_0) we have

$$f(x, y) = f(x_0, y_0) + \partial_x f(x_0, y_0)(x - x_0) + \partial_y f(x_0, y_0)(y - y_0). \tag{9.35}$$

The dominant variation is linear in the displacement $\delta x = x - x_0$ and $\delta y = y - y_0$.

If both $\partial_x f(x_0, y_0) = 0$ and $\partial_y f(x_0, y_0) = 0$ then $f(x, y)$ is said to have a stationary point at (x_0, y_0) because then it does not vary to first order displacements.

In terms of the gradient vector

$$f(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot \delta\vec{x}, \tag{9.36}$$

The condition for a stationary point is naturally $\nabla f(\vec{x}_0) = \vec{0}$.

For a two-variable function there can be local maximum, local minimum and saddle points.

Applying a Taylor expansion to a function that has a stationary point at (x_0, y_0), we obtain the approximation

$$f(x, y) = f(x_0, y_0) + \frac{1}{2}\partial_x\partial_x f(x_0, y_0)(x-x_0)^2 + \partial_x\partial_y f(x_0, y_0)(x-x_0)(y-y_0) + \frac{1}{2}\partial_y\partial_y f(x_0, y_0)(y-y_0)^2. \tag{9.37}$$

To determine whether $f(x, y)$ has a local minimum or maximum, or neither, let us first consider the general homogeneous quadratic function of two variables

$$Q(x, y) = Ax^2 + 2Bxy + Cy^2, \tag{9.38}$$

we can complete the square

$$Q(x, y) = \frac{1}{A}\left[(Ax + By)^2 + (AC - B^2)y^2\right] \tag{9.39}$$

Then everything is straightforward:

$$\begin{aligned}
&\text{if } A > 0 \text{ and } AC > B^2 \text{ which requires } C > 0\\
&\text{then } Q(x, y) > 0 \text{ for all } (x, y) \neq (0, 0) \text{ and we have a minimum}\\
&\text{if } A < 0 \text{ and } AC > B^2 \text{ which requires } C < 0\\
&\text{then } Q(x, y) < 0 \text{ for all } (x, y) \neq (0, 0) \text{ and we have a maximum}\\
&\text{if } AC < B^2 \text{ then } Q(x, y) \text{ can be either positive or negative and we have a saddle point.}
\end{aligned} \tag{9.40}$$

Although this analysis fails when $A = 0$, similar conclusions can be drawn from the alternative way of completing the square,

$$Q(x, y) = \frac{1}{C}\left[(Bx + Cy)^2 + \left(AC - B^2\right)x^2\right], \tag{9.41}$$

which is valid provided that $C \neq 0$. (If $A = C = 0$, then $Q = 2Bxy$, which has a saddle point, unless B also equals to zero.) Applying this analysis to the approximation for f, we conclude that

- f has local minimum if $\partial_x\partial_x f\partial_y\partial_y f > (\partial_x\partial_y f)^2$ with $\partial_x\partial_x f > 0$ and $\partial_y\partial_y f > 0$;
- f has local maximum if $\partial_x\partial_x f\partial_y\partial_y f > (\partial_x\partial_y f)^2$ with $\partial_x\partial_x f < 0$ and $\partial_y\partial_y f < 0$;
- f has a saddle point if $\partial_x\partial_x f\partial_y\partial_y f < (\partial_x\partial_y f)^2$.

We will not discuss any other special cases[3].

Hessian matrix

$$H = \begin{pmatrix} \partial_x\partial_x f & \partial_x\partial_y f \\ \partial_y\partial_x f & \partial_y\partial_y f \end{pmatrix} \tag{9.42}$$

The symmetry of mixed partial derivatives implies that the Hessian matrix is symmetric. It therefore has real eigenvalues and orthogonal eigenvectors. A stationary point is a minimum if the eigenvalues of the Hessian matrix are both positive, a maximum if they are both negative, and a saddle point if one is negative and the other positive.

[3] In special intermediate cases, higher-order terms in the Taylor expansion are needed to determine the nature of the stationary point.

9.4 Conditional Stationary Values

9.4.1 Two independent variables, one constraint

The stationary points found by $\nabla f = 0$ are called unconstraint. We are also interested in the cases where x and y cannot vary independently, but rather are related by a condition or a constraint of the form $g(x, y) = 0$, which means that the points (x, y) lie on a curve.

In some complicated cases we need some particular method to solve the stationary points, a classical one is called the method of Lagrange multipliers.

Lagrange multipliers Suppose the maximum of f on the curve $g = 0$ is at (x_0, y_0). Consider the variation of $f(x, y)$ and $g(x, y)$ near this point. We have

$$\begin{aligned} \mathrm{d}f &= \partial_x f \mathrm{d}x + \partial_x f \mathrm{d}y \\ \mathrm{d}g &= \partial_x g \mathrm{d}x + \partial_y g \mathrm{d}y \end{aligned} \tag{9.43}$$

where we recall that these differentials express the first-order variations. In order to stay on the curve $g = 0$, the displacements $\mathrm{d}x$ and $\mathrm{d}y$ are constrained by

$$\partial_x g \mathrm{d}x + \partial_y g \mathrm{d}y = 0. \tag{9.44}$$

This means nothing but the vector displacement $(\mathrm{d}x, \mathrm{d}y)$ is tangent to the curve $g = 0$. Solving this equation for either $\mathrm{d}x$ or $\mathrm{d}y$ and substituting into the expression for $\mathrm{d}f$, we find

$$\partial_x f \partial_y g - \partial_y f \partial_x g = 0. \tag{9.45}$$

Or, equivalently,

$$\frac{\partial_x f}{\partial_x g} = \frac{\partial_y f}{\partial_y g} = \lambda. \tag{9.46}$$

λ is called the Lagrange multiplier.

Lagrangian function To find the stationary points of $f(x, y)$ subject to the constraint, solve the simultaneous equations

$$f_x = \lambda g_x, \quad f_y = \lambda g_y, \quad g = 0, \tag{9.47}$$

for x, y and λ, where λ is a Lagrange multiplier.

We construct the Lagrangian function[4]

$$L(x, y, \lambda) = f(x, y) - \lambda g(x, y). \tag{9.48}$$

Then the method of Lagrange multipliers is equivalent to solving the equations

$$L_x = L_y = L_\lambda = 0. \tag{9.49}$$

9.4.2 More than two independent variables, more than one constraints

For a Lagrangian function with three independent variables, and two constraints, what you need to do is simply

$$L = f - \lambda g - \mu h. \tag{9.50}$$

Then following this order you know what to do for any general case.

[4]Sometimes $L = f + \lambda g$ is used instead but the results are the same.

9.5 The Boltzmann Distribution as an Example

The classic example of maximisation subject to constraints comes from statistical mechanics. In ideal gas, the total number of particles is

$$N = \sum_{i=1}^{n} N_i, \tag{9.51}$$

the total energy of the system is

$$E = \sum_{i=1}^{n} N_i E_i. \tag{9.52}$$

The distribution of particles among states is described by the numbers $(N_1, N_2, \cdots, N_n)$. A given distribution can be achieved in many different ways, the number of possible ways is simply

$$W = \frac{N!}{N_1! N_2! \cdots N_n!}. \tag{9.53}$$

The function we want to maximise is $\ln W$, which is proportional to the total entropy of the system.

$$\ln W = \ln(N!) - \sum_{i=1}^{n} \ln(N_i!). \tag{9.54}$$

The relevant constraints are $N = \text{const}$ and $E = \text{const}$, which gives

$$L = \ln(N!) - \sum_{i=1}^{n} \ln(N_i!) - \alpha \left(\sum_{i=1}^{n} N_i - N \right) - \beta \left(\sum_{i=1}^{n} N_i E_i - E \right). \tag{9.55}$$

Considering Stirling's approximation

$$x! \approx \sqrt{2\pi x} x^x \mathrm{e}^{-x}, \tag{9.56}$$

the stationary point of L therefore occurs where

$$N_i = N \mathrm{e}^{-\alpha} \mathrm{e}^{-\beta E_i}, \tag{9.57}$$

which gives

$$N_i = \frac{N \mathrm{e}^{-\beta E_i}}{\sum_{j=1}^{n} \mathrm{e}^{-\beta E_j}}. \tag{9.58}$$

With degenerate states of degeneracy g_i,

$$N_i = \frac{N g_i \mathrm{e}^{-\beta E_i}}{\sum_{j=1}^{n} g_j \mathrm{e}^{-\beta E_j}}. \tag{9.59}$$

Chapter 10

Vector Calculus and Fourier Series

Please refer to the 'IA Vector Calculus Notes'. All contents, including tensors, are included in that file.

We shall skip all unnecessary introduction of Fourier series, because it has already been widely used in all subjects of science. This chapter will be concise.

Part III

Easter Term

Chapter 11

Linear Algebra

11.1 Vectors and Vector Spaces

For the complete definition of vector spaces and vectors, refer to the Differential Geometry & General Relativity note.

A vector space is a set of objects with properties (addition and scaling), these objects might be vectors but they don't have to be.

In a vector space, a scalar is simply a real number. If we relate every point in our space with a scalar, a scalar field is created.

We say "V is a vector space over the field K" if certain properties hold. The most important properties for us are still addition and scaling[1].

- If $\vec{u}$ and $\vec{v}$ are elements of V, then so is $\vec{u}+\vec{v}$.
- If $\vec{u}$ is an element of V then so is $\lambda\vec{u}$, where $\lambda \in K$.

The familiar example is that $\mathbb{R}^n$ is a vector space over $\mathbb{R}$. Polynomial functions also satisfies the properties thus they are vector space over $\mathbb{C}$. This shows that elements of vector spaces do not have to be vectors.

The span of $\vec{u}_1, \cdots, \vec{u}_m$ is defined by

$$S = \text{span}(\vec{u}_1, \cdots, \vec{u}_m) = a_1\vec{u}_1 + \cdots + a_m\vec{u}_m. \tag{11.1}$$

If $\vec{u}_1, \cdots, \vec{u}_m$ are elements of vector space V over K, then S is also an vector space over K. If elements of S are linearly independent then the dimension of S is defined to be m[2]. In this case $(\vec{u}_1, \cdots, \vec{u}_m)$ is a set of basis for V. Then any element in V can be written as

$$\vec{v} = \sum_{i=1}^{m} \lambda_i \vec{u}_i. \tag{11.2}$$

And $(\lambda_1, \cdots, \lambda_2)$ are called the coordinates of $\vec{v}$ in the basis $(\vec{u}_1, \cdots, \vec{u}_m)$.

For example, Legendre Polynomials are a basis for polynomials.

11.2 Matrices

"A matrix is a rectangular array of numbers." It looks like

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix}. \tag{11.3}$$

[1] Refer to the discussion on groups in the DG & GR note

[2] In this part of the course we consider only the finite-dimensional vector spaces. But infinite dimensions are important, especially in QM.

11.2.1 Basic operations

The addition of matrices is simply done by adding each term separately

$$C_{ij} = A_{ij} + B_{ij}. \tag{11.4}$$

If a real number λ is applied to a matrix, the resultant matrix is made up of terms

$$D_{ij} = \lambda A_{ij}. \tag{11.5}$$

If A is a $m \times n$ matrix and B is a $n \times p$ matrix, the matrix product of them has elements

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}, \quad i, j \in \mathbb{N}, \quad i \in [1, m], \quad j \in [1, p]. \tag{11.6}$$

The summed indices are called dummy indices, and the others are called free indices. We must have the same free indices on the left hand side and the right hand side. Also, in equations for matrix multiplication, every dummy index should appear exactly twice.

Unfortunately matrices do not generally commute. Often the commutator

$$[A, B] = AB - BA \tag{11.7}$$

is useful in physics.

11.2.2 More definitions

The transpose of a matrix is obtained by swapping rows and columns,

$$A^T_{ij} = A_{ji}. \tag{11.8}$$

The transpose of products is clearly

$$\begin{aligned} (AB)^T_{ij} &= \left(\sum_k A_{ik} B_{kj} \right)^T \\ &= \sum_k A_{jk} B_{ki} \\ &= \sum_k B_{ki} A_{jk} \\ &= \sum_k \left(B^T\right)_{ik} \left(A^T\right)_{kj} \\ &= \left(B^T A^T\right)_{ij} \end{aligned} \tag{11.9}$$

The Hermitian conjugate of a matrix (usually containing complex terms) is defined by

$$A^\dagger = \left(A^T\right)^* = \left(A^*\right)^T, \tag{11.10}$$

and the property $(AB)^\dagger = B^\dagger A^\dagger$ follows naturally. If a square matrix satisfies $A^\dagger = A$, it is called Hermitian, and $A^\dagger = -A$ for anti-Hermitian.

If a matrix satisfies $A = A^T$, it is symmetric. And anti-symmetric matrices satisfy $A = -A^T$[3]. Diagonal matrices have $A_{ij} = 0$ if $j \neq i$. Unit matrix of order n is simply $I = \mathrm{diag}(1, \cdots, 1)$[4].

The trace of a matrix is defined by the sum of its diagonal components

$$\mathrm{Tr}(A) = \sum_k A_{kk}. \tag{11.11}$$

[3] They have zeros diagonally.

[4] Or using the Kronecker delta to represent unit matrices δ_{ij}

Of course $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$, and following the cyclic permutation,

$$\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB). \tag{11.12}$$

For column vectors with complex elements, it is conventional to define the scalar product as $\vec{a} \cdot \vec{b} = \vec{a}^\dagger \vec{b}$. This ensures that $\vec{a} \cdot \vec{a}$ is always a real number (either positive or zero), and $\sqrt{\vec{a} \cdot \vec{a}}$ can be interpreted as the length of the vector.

11.2.3 Determinants

The determinant of a square matrix A is a scalar, denoted by $\det(A)$, for 2×2 matrices,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \det(A) = ad - cb. \tag{11.13}$$

But for $n \geq 3$ matrices,

$$\det(A) = \sum_k (-1)^{1+k} M_{1,k} A_{1k} = \sum_k C_{1,k} A_{1k}, \tag{11.14}$$

where $M_{1,k}$ is the minor of A_{1k} and $C_{1,k}$ is the corresponding factor.

The Levi-Civita symbol (also known as alternating tensor or completely antisymmetric tensor) is

$$\varepsilon_{ijk} = \begin{cases} 1 & \text{if } (ijk) = (123) \\ -1 & \text{if } (ijk) = (321) \\ 0 & \text{otherwise} \end{cases}, \tag{11.15}$$

then for 3×3 matrices the determinant could be written as

$$\det(A) = \sum_{ijk} \varepsilon_{ijk} A_{1i} A_{2j} A_{3k}. \tag{11.16}$$

Similarly, for higher order square matrices,

$$\det(A) = \sum_{ijk\cdots n} \varepsilon_{ijk\cdots n} A_{1i} A_{2j} A_{3k} \cdots A_{mn}. \tag{11.17}$$

11.2.4 Properties of determinants

Properties related to row operations:

- If we swap two rows, the determinant should be multiplied by -1.
- If we multiply any row with λ, the determinant should be multiplied by λ.
- Adding any row to another leaves the determinant unchanged.
- If the rows are linearly dependent, the determinant should be zero (this can be proved by combining the two properties above).

Other properties seem obvious from the definition in the last subsection.

- $\det(A) = \det(A^T)$.
- $\det(AB) = \det(A)\det(B)$.
- $\det(\mathrm{diag}(a_1, a_2, \cdots)) = a_1 a_2 \cdots$.

11.2.5 Inverse

The inverse matrix A^{-1} is defined by

$$A^{-1}A = I, \tag{11.18}$$

where I is the unit matrix with the same dimension.

The components of A^{-1} can be found by taking the determinant,

$$A_i^{-1}j = \frac{C_{j,i}}{\det(A)}. \tag{11.19}$$

Some properties of the inverse, which are again obvious from definition:

- Although matrices don't generally commute, $[A, A^{-1}] = 0$.
- $(A^{-1})^T = (A^T)^{-1}$.
- The inverse is unique.

11.2.6 Simultaneous equations

It is easy to write simultaneous equation in linear form, $Ax = y$. If we want to solve x just take the inverse of A and write in the form $x = A^{-1}y$.

It is interesting to consider the solutions to $Ax = y$, since it does not always have unique solutions.

$Ax = 0$ The kernel of A is a set of vectors x that obey $Ax = 0$, denoted as $\ker(A)$.

It is obvious to say that we can always take $x = 0$, and $Ax = 0$ has non-zero solutions if and only if the columns of A are linearly dependent.

If the only possible x is 0, we say that the kernel is trivial. And $Ax = 0$ has a unique solution ($x = 0$) in this case.

$Ax = y$ Now we move on to the general case, we shall see first the existence of solutions then the uniqueness.

Suppose we choose a set of b_i that

$$\sum_i b_i(i\text{th equation}) = 0, \tag{11.20}$$

on the RHS we have

$$0 = \sum_i b_i y_i. \tag{11.21}$$

If the RHS is not zero then the equations are inconsistent, which means that no solutions exist. Solutions exist if and only if the rows of A are linearly dependent.

If solutions exist we shall move on to consider if it is unique. Suppose z is a non-trivial component of $\ker(A)$, then

$$A(x + \lambda z) = y, \tag{11.22}$$

which means that $x + \lambda z$ is another set of solutions. Hence, if $\ker(A)$ is not trivial, we can never have a unique solution.

If the columns of A are linearly dependent then the kernel is non-trivial so $Ax = y$ can't have a unique solution (either multiple solutions or no solutions). Also, if the columns are linearly independent then there can't be multiple solutions (either no solutions or a unique solution).

11.3 Linear Maps

Consider two vector spaces V and W. A map from V to W is a function f with the following property: if $v \in V$ then $f(v) \in W$. A linear map from V to W is a mapping with one additional property: if $v_1, v_2 \in V$, then

$$f(v_1 + v_2) = f(v_1) + f(v_2), \tag{11.23}$$

which is called linearity.

For example, if A is a $m \times n$ matrix, then it is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$.

11.3.1 Orthogonal matrices

A square matrix with $O^TO = I$ is called an orthogonal matrix. Using the properties of inverse matrices, this means that $O^{-1} = O^T$ and $OO^T = O^TO = I$.

Also, the determinant is special since

$$1 = \det(I) = \det(OO^T) = \det(O^TO) = \det(O)\det(O^T) = \det(O)^2, \tag{11.24}$$

which implies that $\det(O) = \pm 1$.

Now it is clear to say that the columns of any orthogonal matrix are an orthonormal set of vectors (basis).

11.3.2 Change of coordinates

It is not necessary to discuss it here because it is often used in physics.

11.4 Eigenvalues and Eigenvectors

For a square matrix, if there exists a pair of (λ, v) which satisfies

$$Av = \lambda v, \tag{11.25}$$

then v is called an eigenvector of A and λ is the corresponding eigenvalue of A.[5]

To find out the eigenvectors and eigenvalues, we should consider in general the equation:

$$(A - \lambda I)v = 0. \tag{11.26}$$

The equation has non-zero solutions if and only if $\det(A - \lambda I) = 0$. We define

$$P_A(\lambda) = \det(A - \lambda I), \tag{11.27}$$

as the characteristic polynomial of A. Then the eigenvalues of A are the roots of P_A, which are the solutions of $P_A(\lambda) = 0$.

An $n \times n$ matrix should have n eigenvalues. If we have solved all of them, we can find their corresponding eigenvectors by directly plugging λ into $(A - \lambda I)v = 0$.

It is easy to find, following the properties of polynomials, that

$$\mathrm{tr}(A) = \sum \lambda, \quad \det(A) = \prod \lambda. \tag{11.28}$$

Recalling the eigenvalues and eigenvectors of Hessian matrix, if it has two positive elements, the function has a local minimum; two negative elements a local maximum; and positive plus negative a saddle point. In most cases, these results can be easily obtained by taking the trace and determinant, since we do not need the exact value.

11.4.1 For real symmetric matrices

Eigenvalues and eigenvectors for real symmetric matrices have some nice properties, lead by the linearity and the real, symmetric features.

Firstly, eigenvalues for real symmetric matrices are real, this could be proved by taking the complex conjugate of the eigenvalue definition and contract with the original equation. The symmetric property $(A = A^T)$ is also used.

Secondly, eigenvectors corresponding to different eigenvalues are orthogonal. This can be proved by listing two different eigenvalues of A, writing as $w^T\lambda v = (\gamma w)^T v$ and rearranging.

Finally, if A is $n \times n$ then a set of n orthogonal eigenvectors always exists.

[5] In quantum mechanics we know that each eigenvector is a physical state and the eigenvalue is the measured physical quantity value of that state.

11.4.2 Diagonalisation of real symmetric matrices

The transformation from the original matrix A to the diagonal matrix D can be written as

$$D = S^T AS, \tag{11.29}$$

where S is built by listing rows of eigenvectors.

For partial differential equations:

Please refer to the book written by Evans, there is literally nothing to add.