

FORMULAE ONE

ESSENTIAL
FORMULA SHEET
FOR EXAMS

Prepared by Soner Albayrak
 ©2025 soneralbayrak.com
 All rights reserved.

This document is prepared to help students in their examinations.

...it is originally designed to help students of the courses offered by the author at the Middle East Technical University.

This document is not meticulously checked, nor is it peer-reviewed.

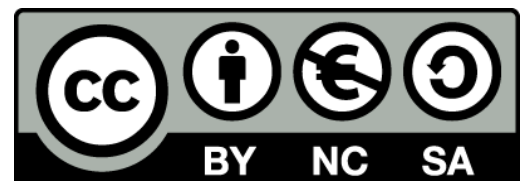
...if you are using it for an examination of the author, he will grade accordingly if there is an incorrect statement in this booklet.

...if not, no responsibility is taken for the consequences of any potentially incorrect information: use this booklet at your own risk!

This version of the document is prepared on November 11, 2025.

This document is licensed as CC BY-NC-SA, therefore you are free to share, copy, remix, redistribute, make derivatives as long as you attribute this work, use the material for non-commercial purposes only, and share under the same license.

<https://creativecommons.org/licenses/by-nc-sa/4.0/>



Contents

1	Mathematical notations and conventions	3
2	Special functions and definite integrals	10
3	Concept of classical probability	11
4	Fundamental definitions in Physics	14

1 Mathematical notations and conventions

This section sets our mathematical notations and conventions at a rigorous enough for a physics student level.

Set for our purposes is a collection of elements such that any given object either belongs or not to the chosen collection. In sets, the order of elements do not matter, i.e. $\{a, b\} = \{b, a\}$.

Ordered set for our purposes is same as set *except* the order of elements matter, i.e. $(a, b) \neq (b, a)$. One can construct ordered sets out of ordinary sets, $(a, b, c) := \{a, \{a, b\}, \{a, b, c\}\}$ and so on. Ordered sets with n elements is also called n -tuple.

Cartesian product of the sets $S_1, \dots, S_n$ is denoted as $S_1 \times S_2 \times \dots \times S_n$ and is inhabited by n -tuples, i.e.

$$S_1 \times \dots \times S_n = \{(x_1, \dots, x_n) \mid x_i \in S_i\} \quad (a)$$

Cartesian power (denoted S^n) is a special case of cartesian product such that $S^n := S \times S \times \dots \times S$.

Boolean (denoted $\mathbb{B}$) is a set with only two elements, usually chosen as True and False. Engineers also prefer 0 & 1.

A function f is a map from a set A (called *domain*) to another set B (called *codomain*) and is denoted $f : A \rightarrow B$. The output of the function, $f(x) \in B$, is uniquely determined by its input, $x \in A$; if we remove this condition, we get *relations* (also called *multivalued functions*).

Predicate is any function whose codomain is the set Boolean. For instance, $f : \mathbb{R} \rightarrow \mathbb{B}$ with

$$f(x) = (x > 2) \quad (b)$$

is a predicate with $f(3) = \text{True}$ and $f(1) = \text{False}$.

Truth functions are predicates whose domain is $\mathbb{B}^n$. In propositional logic, these functions are built through *logical connectives* among which the common ones are *and* ($\wedge$), *or* ($\vee$), *nand* ($\bar{\wedge}$), *nor* ($\bar{\vee}$), *equivalent* ($\Leftrightarrow$), and *nonequivalent* ($\nLeftrightarrow$). For instance, if a is True ($a = T$) and b is False ($b = F$), *their or is True* ($a \vee b = T$): the “truth table” below shows outputs of these connectives for all possible inputs:

a	b	$a \wedge b$	$a \bar{\wedge} b$	$a \vee b$	$a \bar{\vee} b$	$a \Leftrightarrow b$	$a \nLeftrightarrow b$
T	T	T	F	T	F	T	F
T	F	F	T	T	F	F	T
F	T	F	T	T	F	F	T
F	F	F	T	F	T	T	F

In engineering, *xor* / *xnor* are preferred instead of *nonequivalent* / *equivalent*; in any case, *functional completeness* of *nand* (*nor*) guarantees that any truth function can be built by only using $\bar{\wedge}$ ($\bar{\vee}$): in computer science, they are called *universal gates*. For instance, *or* can be constructed via *nand* as $a \vee b = (a \bar{\wedge} a) \bar{\wedge} (b \bar{\wedge} b)$; likewise, *negation* (denoted $\neg$) can be constructed via *nor* as $\neg a = a \bar{\vee} a$ where negation is defined with the property that $\neg \text{True} = \text{False}$ and $\neg \text{False} = \text{True}$.

Injective (surjective) functions (for our purposes) are functions, $f : A \rightarrow B$, whose action can always be cancelled by another function called *left (right) inverse* $g : B \rightarrow A$, i.e. $g \cdot f = \mathbb{I}_A$ ($f \cdot g = \mathbb{I}_B$) where $\mathbb{I}_C$ denotes identity function in C .

Bijjective functions (for our purposes) are those which are both injective and surjective, hence they have a unique inverse (denoted f^{-1}) such that $f \cdot f^{-1} = f^{-1} \cdot f = \mathbb{I}$.

Monomorphism, epimorphism, isomorphism

are generalizations of injective, surjective, and bijective functions beyond set theory. We will only use them as synonyms for their set-theoretic correspondents.

Set comprehension is a method to generate subsets from a big set via a predicate. The usual convention is $\{x \in S \mid P(x)\}$ which is a set whose elements are elements of S for which $P(x)$ is true. For instance, $\{x \in \mathbb{Z} \mid x^2 > 100\}$ is the set of integers whose square is greater than 100.

The non-negative integer power of an object A (denoted A^n) is defined recursively as

$$A^0 = \mathbb{I}, \quad A^n = A \cdot A^{n-1} \text{ for } n \geq 1 \quad (c)$$

with respect to the operation $\cdot$ (such as matrix multiplication or differentiation) and its identity object $\mathbb{I}$.

Exponentiation of an object A (denoted e^A) is

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n \quad (d)$$

where A^n is the n -th power of the object A .

Logarithm of an object A (denoted $\log A$) is defined as the inverse of the exponentiation. For objects for which the exponentiation is not a monomorphism (such as complex numbers), logarithm is a *relation* instead. Conventionally, one imposes restrictions on the domain to ensure that logarithm acts

as a function; for instance, for a complex number $z = re^{i\theta} \in \mathbb{C}$ with $(r, \theta) \in (\mathbb{R}^+, \mathbb{R})$, we can define $\log z = i\theta_p + \log r$ where $0 \leq \theta_p < 2\pi$ is called *the principal value of θ* .

The generalized power of an object A (denoted A^α) is defined as

$$A^\alpha = e^{\alpha \log A} \quad (e)$$

If exponentiation is not a monomorphism when acting on the domain of A , A^α is not a function but a relation *unless* a principle domain is selected (similar to the logarithm).

Generalized exponentiation of an object A (denoted α^A) is defined as

$$\alpha^A = e^{A \log \alpha} \quad (f)$$

Depending on the available values for $\log \alpha$, α^A may mean multiple different functions. However, each one is *still* a proper function, not a multi-valued function.

Trigonometric functions $\cos$, $\sin$, $\tan$, $\cot$, $\csc$, $\sec$ are defined in terms of the exponential via the equations

$$e^{\pm iA} = \cos(A) \pm i \sin(A), \quad \tan(A) = \frac{1}{\cot(A)} = \frac{\sin(A)}{\cos(A)} \quad (g)$$

$$\csc(A) \sin(A) = 1, \quad \sec(A) \cos(A) = 1 \quad (h)$$

Hyperbolic functions $\cosh$, $\sinh$, $\tanh$, $\coth$, $\csch$, $\sech$ are defined in terms of the exponential via equations

$$e^{\pm A} = \cosh(A) \pm \sinh(A), \quad \tanh(A) = \frac{1}{\coth(A)} = \frac{\sinh(A)}{\cosh(A)} \quad (i)$$

$$\csch(A) \sinh(A) = 1, \quad \sech(A) \cosh(A) = 1 \quad (j)$$

Inverse Trigonometric/Hyperbolic functions

are denoted with an *arc* prefix in their naming, i.e. $\arcsin(x) := \sin^{-1}(x)$. Like logarithm, these objects are *relations* (not functions) unless their domain is restricted.

The Kronecker symbol (Kronecker-delta) is defined

$$\delta : \{\mathbb{Z}, \mathbb{Z}\} \rightarrow \mathbb{Z} \quad (k)$$

$$\delta = \{i, j\} \rightarrow \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (l)$$

The Dirac-delta generalized function δ is (for all practical purposes of a Physicist) defined via the relation

$$\int_{\mathcal{A}} f(y) \delta(x - y) dy = \begin{cases} f(x) & \text{if } x \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases} \quad (m)$$

A useful representation of this generalized function is

$$\delta(x) = \int_{-\infty}^{\infty} e^{ikx} \frac{dk}{2\pi} \quad (n)$$

which can also be understood as Fourier transform of 1.

Heaviside generalized function θ is (for all practical purposes of a Physicist) defined via the relations

$$\int_a^b \theta(x) f(x) dx = \begin{cases} \int_a^b f(x) dx & \text{if } a \geq 0 \\ \int_0^b f(x) dx & \text{if } a < 0 \end{cases} \quad (o)$$

This definition implies that $\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ for $x < 0$; however, it *does not fix* $f(0)$. We choose *the convention* $f(0) = 1/2$; this ensures

$$\text{sgn}(x) = 2\theta(x) - 1 = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ -1 & \text{for } x < 0 \end{cases} \quad (p)$$

A particular permutation of n objects is denoted as $(i_1 i_2 \dots i_n)$ where $i_1 \neq i_2 \neq \dots \neq i_n \in \{1, \dots, n\}$. A permutation $(i_1 \dots i_n)$ is said to be an even (odd) permutation of $(k_1 \dots k_n)$ if the two are identical after the permutation of an even (odd) number of adjacent indices. For example, (2431) is an even permutation of (2143) and an odd one of (2134).

Levi-Civita symbol ϵ is defined as

$$\epsilon : \{\mathbb{Z}^+, \dots, \mathbb{Z}^+\} \rightarrow \mathbb{Z} \quad (q)$$

$$\epsilon = \{a_1, \dots, a_n\} \rightarrow \begin{cases} 1 & \text{if } (a_1 a_2 \dots a_n) \text{ is an even} \\ & \text{permutation of } (12 \dots n) \\ -1 & \text{if } (a_1 a_2 \dots a_n) \text{ is an odd} \\ & \text{permutation of } (12 \dots n) \\ 0 & \text{otherwise} \end{cases} \quad (r)$$

The determinant function (denoted $\det$) is defined

$$\det : \mathfrak{M}_{n \times n}(\mathcal{A}) \rightarrow \mathcal{A} \quad (s)$$

$$\det = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \sum_{i_1, \dots, i_n} \epsilon_{i_1 \dots i_n} a_{1i_1} \dots a_{ni_n} \quad (t)$$

where $\mathcal{A}$ is any field such that $a_{ij} \in \mathcal{A}$, $\forall i, j$.

The adjugate function (denoted adj) is defined as

$$\text{adj} : \mathfrak{M}_{n \times n}(\mathcal{A}) \rightarrow \mathfrak{M}_{n \times n}(\mathcal{A}) \quad (u)$$

$$\text{adj} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & & & \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix} \quad (v)$$

for

$$b_{k_n i_n} = \sum_{\substack{i_1, \dots, i_{n-1} \\ k_1, \dots, k_{n-1}}} \frac{\epsilon_{i_1 \dots i_n} \epsilon_{k_1 \dots k_n} a_{i_1 k_1} \dots a_{i_{n-1} k_{n-1}}}{(n-1)!} \quad (w)$$

where $\mathcal{A}$ is any field such that $a_{ij} \in \mathcal{A}$, $\forall i, j$.

Inverse of an object A is denoted as A^{-1} and is defined with respect to an operation “.” and its identity element $\mathbb{I}$ via the the equations $A \cdot A^{-1} = A^{-1} \cdot A = \mathbb{I}$. If “.” is matrix multiplication, then

$$A^{-1} = \frac{\text{adj}(A)}{\det A} \quad (x)$$

The trace function (denoted tr) is defined as

$$\text{tr} : \mathfrak{M}_{n \times n}(\mathcal{A}) \rightarrow \mathcal{A} \quad (\text{y})$$

$$\text{tr} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \sum_i a_{ii} \quad (\text{z})$$

where $\mathcal{A}$ is any field such that $a_{ij} \in \mathcal{A}$, $\forall i, j$.

Wronskian matrix of a set of functions $\{f_1(x), \dots, f_n(x)\}$ is defined as a square matrix where the first row is the set of the functions and the i -th row is $(i-1)$ -th derivative of the functions for all $n \geq i \geq 2$.

A complex number z is (for all practical purposes of a Physicist) a pair of two real numbers (x, y) where one can construct z via $z = x + iy$ (i is called *the imaginary unit* with the property $i^2 = -1$); conversely, one can extract x and y via $x = \text{Re}(z)$, $y = \text{Im}(z)$.

Complex conjugation (denoted $*$) is a function defined to act on complex numbers as

$$* : \mathbb{C} \rightarrow \mathbb{C} \quad (\text{aa})$$

$$* = z \rightarrow (z^* = \text{Re}(z) - i \text{Im}(z)) \quad (\text{ab})$$

Matrix transpose (denoted T) is a function defined

$$T : \mathfrak{M}_{n \times n}(\mathcal{A}) \rightarrow \mathfrak{M}_{n \times n}(\mathcal{A}) \quad (\text{ac})$$

$$T = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & & & \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix} \quad (\text{ad})$$

where $\mathcal{A}$ is any field such that $a_{ij} \in \mathcal{A}$, $\forall i, j$.

Hermitian conjugation (also called *conjugate transpose*, *adjoint*, or *dagger*) is a function defined as

$$\dagger : \mathfrak{M}_{n \times n}(\mathbb{C}) \rightarrow \mathfrak{M}_{n \times n}(\mathbb{C}) \quad (\text{ae})$$

$$\dagger = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} a_{11}^* & a_{21}^* & \dots & a_{n1}^* \\ a_{12}^* & a_{22}^* & \dots & a_{n2}^* \\ \dots & & & \\ a_{1n}^* & a_{2n}^* & \dots & a_{nn}^* \end{pmatrix} \quad (\text{af})$$

Characteristic polynomial of any square matrix A is

$$\det(A - \lambda_i \mathbb{I}) = 0 \quad (\text{ag})$$

Laplace transform (denoted $\mathcal{L}$) is an integral transform which converts a function $f : \mathbb{R} \rightarrow \mathbb{R}$ into another function $\hat{f} = \mathcal{L}(f)$ such that

$$\hat{f} : \mathbb{C} \rightarrow \mathbb{C}, \quad \hat{f}(s) = \int_0^\infty f(x) e^{-xs} dx \quad (\text{ah})$$

For *meromorphic* $\hat{f}$ (i.e. $\frac{\text{polynomial}}{\text{polynomial}}$), the inverse is computed by rewriting $\hat{f}(s)$ as a sum $\sum_i a_i (s + r_i)^{-n_i-1}$ which

is clearly (for some $c_{k,l}$) the Laplace transform of $f(x) = \sum_i e^{-r_i x} (c_{i,1} + c_{i,2}x + \dots c_{i,n_i}x^{n_i})$. Formally,

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad f(x) = \int_{\gamma-i\infty}^{\gamma+i\infty} \hat{f}(s) e^{xs} \frac{ds}{2\pi i} \quad (\text{ai})$$

where the *contour integral* in the complex plane is chosen appropriately based on the convergence.

Convolution of two functions f and g (denote $f * g$) is the operation that becomes multiplication in the Laplace domain, i.e. $\mathcal{L}(f * g) \equiv \mathcal{L}(f)\mathcal{L}(g)$; equivalently,

$$(f * g)(x) = \int_0^x f(y)g(x-y)dy \quad (\text{aj})$$

Fourier transforms are widely-used integral transformations, the simplest examples of harmonic analysis, and can be defined with any self-consistent convention. We choose

$$f : \mathbb{R} \rightarrow \mathbb{C}, \quad f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{ikx} \hat{f}(k) \quad (\text{ak})$$

$$\hat{f} : \mathbb{R} \rightarrow \mathbb{C}, \quad \hat{f}(k) = \int_{-\infty}^{\infty} dx e^{-ikx} f(x) \quad (\text{al})$$

— — —

$$f : [a, a+T] \rightarrow \mathbb{C}, \quad f(x) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{i\frac{2\pi n}{T}x} \hat{f}(n) \quad (\text{am})$$

$$\hat{f} : \mathbb{Z} \rightarrow \mathbb{C}, \quad \hat{f}(n) = \int_a^{a+T} dx e^{-i\frac{2\pi n}{T}x} f(x) \quad (\text{an})$$

— — —

$$f : \mathbb{Z} \rightarrow \mathbb{C}, \quad f(n) = \frac{1}{T} \int_a^{a+T} dx e^{i\frac{2\pi n}{T}x} \hat{f}(k) \quad (\text{ao})$$

$$\hat{f} : [a, a+T] \rightarrow \mathbb{C}, \quad \hat{f}(k) = \sum_{n=-\infty}^{\infty} e^{-i\frac{2\pi n}{T}k} f(n) \quad (\text{ap})$$

— — —

$$f : \mathbb{Z}_N \rightarrow \mathbb{Z}_N, \quad f(n) = \frac{1}{N} \sum_{m=0}^{N-1} e^{i\frac{2\pi nm}{N}} \hat{f}(m) \quad (\text{aq})$$

$$\hat{f} : \mathbb{Z}_N \rightarrow \mathbb{Z}_N, \quad \hat{f}(m) = \sum_{n=0}^{N-1} e^{-i\frac{2\pi nm}{N}} f(n) \quad (\text{ar})$$

where (ak), (am), (ao), and (aq) are called *Fourier Transform*, *Fourier Series*, *Discrete-time Fourier Transform*, and *Discrete Fourier Series* respectively while the rest are their inverses. We will stick to this naming but please be reminded that different communities (engineering, math, physics, etc.) use different naming conventions in general.

Rectangle function (also called *unit pulse*, *window / gate function*) is defined in terms of Heaviside function as

$$\text{rect}(x) = \theta\left(x + \frac{1}{2}\right) - \theta\left(x - \frac{1}{2}\right) \quad (\text{as})$$

Sine cardinal function (denoted sinc) is defined as $\text{sinc}(0) = 1$ and $\text{sinc}(x) = \sin(x)/x$ for $x \neq 0$. It is the Fourier transform of rect function, i.e. $\text{rect}(x) \leftrightarrow \text{sinc}(k/2)$.

“Even part of” and “odd part of” (denoted E and O) are higher order functions defined as

$$E : (\mathcal{A} \rightarrow \mathcal{A}) \rightarrow (\mathcal{A} \rightarrow \mathcal{A}) \quad (\text{at})$$

$$E = (x \rightarrow f(x)) \rightarrow \left(x \rightarrow f_E(x) = \frac{f(x) + f(-x)}{2} \right) \quad (\text{au})$$

$$O : (\mathcal{A} \rightarrow \mathcal{A}) \rightarrow (\mathcal{A} \rightarrow \mathcal{A}) \quad (\text{av})$$

$$O = (x \rightarrow f(x)) \rightarrow \left(x \rightarrow f_O(x) = \frac{f(x) - f(-x)}{2} \right) \quad (\text{aw})$$

with which any single-argument function satisfies $f = E \cdot f + O \cdot f$ (more commonly written $f(x) = f_E(x) + f_O(x)$).

Inner product between two functions f and g (for our purposes) is denoted $\langle f, g \rangle_\omega^{\mathcal{A}}$ for $\mathcal{A} \subseteq \mathbb{R}$ and is defined as:

$$\langle \cdot, \cdot \rangle_\omega^{\mathcal{A}} : (\mathcal{A} \rightarrow \mathbb{C}, \mathcal{A} \rightarrow \mathbb{C}) \rightarrow \mathbb{C} \quad (\text{ax})$$

$$\langle f, g \rangle_\omega^{\mathcal{A}} = \int_{\mathcal{A}} \left(f(x) \right)^* g(x) \omega(x) dx \quad (\text{ay})$$

Group is defined as a pair (S, o) where $S : \text{Set}$ and where $o : (S, S) \rightarrow S$ for which the following are true:

1. $(\exists e \in S)(\forall s \in S) o(e, s) = o(s, e) = s$
2. $(\forall s \in S) o(s, i(s)) = o(i(s), s) = e$
3. $(\forall a, b, c \in S) o(a, o(b, c)) = o(o(a, b), c)$

for a unique function $i : S \rightarrow S$.

Ring is defined as a triplet $(S, +, \cdot)$ where $S : \text{Set}$, and $+, \cdot : (S, S) \rightarrow S$ for which the following are true:

1. $(S, +) : \text{Commutative Group}$
2. $(\forall a, b, c \in S) a \cdot (b + c) = a \cdot b + a \cdot c$
3. $(\forall a, b, c \in S) (b + c) \cdot a = b \cdot a + c \cdot a$

Skew field is defined as a triplet $(S, +, \cdot)$ where $S : \text{Set}$, and $+, \cdot : (S, S) \rightarrow S$ for which following are true:

1. $(S, +, \cdot) : \text{Ring}$
2. $(S \setminus \{0\}, \cdot) : \text{Group}$

where 0 denotes the identity element with respect to $+$.

Field is defined as a triplet $(S, +, \cdot)$ where $S : \text{Set}$, and $+, \cdot : (S, S) \rightarrow S$ for which the following are true:

1. $(S, +, \cdot) : \text{Ring}$
2. $(S \setminus \{0\}, \cdot) : \text{Commutative Group}$

where 0 denotes the identity element with respect to $+$.

Linear space (also called *vector space*) over a field $F = (S, +, \cdot)$ shall be denoted as $V(F)$ and is defined as a triplet $(V, \oplus, \odot)$ with $V : \text{Set}$, $\oplus : (V, V) \rightarrow V$, and $\odot : (S, V) \rightarrow V$ for which the following are true:

1. $(V, \oplus) : \text{Commutative Group}$
2. $(\forall v \in V) 1 \odot v = v$ (1 is the identity element of $\cdot$)
3. $(\forall v \in V)(\forall s \in S) s \odot v \in V$
4. $(\forall v \in V)(\forall a, b \in S) (a \cdot b) \odot v = a \odot (b \odot v)$
5. $(\forall v \in V)(\forall a, b \in S) (a + b) \odot v = (a \odot v) \oplus (b \odot v)$
6. $(\forall v, w \in V)(\forall s \in S) s \odot (v \oplus w) = (s \odot v) \oplus (s \odot w)$

The elements of the set $S(V)$ are called *scalars (vectors)*.

Linear algebra (also called *vector algebra*) over a field $F = (S, +, \cdot)$ shall be denoted as $L(F)$ and is defined as a quadruple $(V, \oplus, \odot, \otimes)$ for which the following are true:

1. $(V, \oplus, \odot) : \text{Linear Space}$
2. $(\forall x, y, z \in V) x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z)$
3. $(\forall x, y, z \in V) (x \oplus y) \otimes z = (x \otimes z) \oplus (y \otimes z)$
4. $(\forall x, y \in V)(\forall a, b \in S) (a \odot x) \otimes (b \odot y) = (a \cdot b) \odot (x \otimes y)$

Lie algebra is a linear algebra $(V, \oplus, \odot, \otimes)$ with the additional condition that $(\forall x, y \in V) x \otimes y = -y \otimes x$.

Commutator is a higher order function which takes two functions $f, g : \mathcal{A} \rightarrow \mathcal{A}$ for any type $\mathcal{A}$, and gives a new function $[f, g] : \mathcal{A} \rightarrow \mathcal{A}$ by cascading their action. It is defined on an object $x \in \mathcal{A}$ as $[f, g](x) = f(g(x)) - g(f(x))$.

Basis B of a vector space V is “ $(B \supset V) : \text{Set}$ ” such that

1. $(\forall k \in \{1, 2, \dim B\})(\forall e_1, \dots, e_k \in B)(\forall c_1, \dots, c_k \in S) [c_1 = \dots = c_k = 0] \vee [c_1 e_1 + \dots + c_k e_k \neq 0]$
2. $(\forall v \in V)(\exists! a_1, \dots, a_{\dim B} \in S) v = a_1 e_1 + \dots + a_{\dim B} e_{\dim B}$

Normed vector space over a field F is a vector space $V(F)$ over which a function norm $: V \rightarrow \mathbb{R}$ exists with the notation $\text{norm} = x \rightarrow \|x\|$, for which following are true:

1. $(\forall v \in V)[\|v\| \neq 0] \vee [v = 0]$
2. $(\forall v \in V)(\forall s \in F)\|s \odot v\| = |s| \cdot \|v\|$
3. $(\forall v, w \in V)\|v \oplus w\| \leq \|v\| + \|w\|$

Inner product vector space over a field F is a vector space $V(F)$ over which a function $\langle \cdot \rangle : (V, V) \rightarrow \mathbb{C}$ exists for which following statements are true:

1. $(\forall v, w \in V) \langle v, w \rangle = \langle w, v \rangle^*$
2. $(\forall u, v, w \in V)(\forall a, b \in F) \langle au + bv, w \rangle = a \langle u, w \rangle + b \langle v, w \rangle$
3. $(\forall v \in V \setminus \{0\}) \langle v, v \rangle > 0$
4. $\langle 0, 0 \rangle = 0$

Dual of a vector space $V(F)$ is a vector space denoted as $V^*(F)$ whose elements are linear functions from the vector space $V(F)$ to the underlying field F .

Type (r, s) tensor on a vector space V is an element of vector space $\underbrace{V \otimes V \otimes \dots \otimes V}_r \otimes \underbrace{V^* \otimes V^* \otimes \dots \otimes V^*}_s$ where $\otimes$ is an associative bilinear map.

Tensor algebra $T(V)$ over a vector space V is the direct sum of all possible (r, s) tensor spaces, with the $\otimes$ being the natural product between different tensors.

Multivector (also called k -vector) is an element of the vector space whose elements are constructed via the associative antisymmetric *wedge product* $\wedge$ of the underlying vectors; e.g. $u \wedge v$ is a 2-vector if u and v are vectors.

Exterior algebra $\Lambda(V)$ over a vector space V is the direct sum of all possible multivectors, with wedge product $\wedge$ being the natural product between multivectors.

Covariant & Contravariant indices in our conventions refer to *downstairs* and *upstairs* indices of a tensor's components, hence are multiplied with basis vectors of V^* and V to yield the full tensor, e.g. $T = T^{ij}_k e_i \otimes e_j \otimes e^k$ with T^{ij}_k having one covariant and two contravariant indices where $\otimes$ is the associative binary operation of the algebra, e.g. $\otimes, \wedge, \dots$.

Contraction is the action of applying a dual vector ($V \rightarrow S$) to a vector (V), hence reducing a (r, s) -tensor to a $(r-1, s-1)$ -tensor. In an orthonormal basis with $e^i(e_k) = \delta^i_k$ (such as Cartesian coordinates), this amounts to summing over a covariant and a contravariant indices.

Manifold is (for our purposes) any space that resembles $\mathbb{R}^d$ near its every point, for instance the sphere S^2 .

(Co)tangent space to a manifold M at a point x is $\mathbb{R}^d$ centered at x and is denoted as $T_x M$ ($T^*_x M$). The (co)tangent space is inhabited by the (co)vectors at $x \in M$, with the basis vectors usually chosen as $\frac{\partial}{\partial x^i} (dx_i)$.

(Co)tangent bundle is *disjoint union* of (co)tangent spaces of a manifold M , and is denoted as TM (T^*M).

Musical isomorphism between a tangent and cotangent bundle is initiated with two functions: $\flat : TM \rightarrow T^*M$ and $\sharp : T^*M \rightarrow TM$, hence for instance $(x^i e_i)^\flat = (x_i e^i)$, and $(x_i e^i)^\sharp = (x^i e_i)$

Field in Physicist terminology broadly refers to any map from a manifold M to *something* ($\mathbb{R}, TM, \dots$). The field is *named* appropriately depending on the output: scalar field ($M \rightarrow \mathbb{R}$), vector field ($M \rightarrow TM$), tensor field ($M \rightarrow (TM \otimes TM \otimes T^*M \otimes \dots)$), and so on.

Differential forms (or forms for short) are functions that takes a point x from a Manifold M and yields a multi(co)vector from the exterior algebra of the (co)tangent space of M at x , e.g. $\omega = (x, y) \rightarrow dx + y dy$.

Hodge dual of a multivector or a form α is denoted as $\star \alpha$, and its components in $\mathbb{R}^d$ are related for $\alpha = \alpha_{i_1 \dots i_k} e^{i_1} \wedge \dots \wedge e^{i_k}$ and $\star \alpha = (\star \alpha)_{i_{k+1} \dots i_d} e^{i_{k+1}} \wedge \dots \wedge e^{i_d}$ as

$$(\star \alpha)_{i_{k+1} \dots i_d} = \frac{1}{(d-k)!} \alpha_{i_1 \dots i_k} \epsilon^{i_1 \dots i_k i_{k+1} \dots i_d} \delta_{i_{k+1} \ell_{k+1}} \dots \delta_{i_d \ell_d}$$

Exterior derivative takes a p -form ω to $p+1$ form $d\omega$;

with the basis vectors $\{dx^i\}$, it reads as

$$\omega = \omega_{i_1 \dots i_p} dx^{i_1} \wedge \dots \wedge dx^{i_p} \quad (\text{az})$$

$$d\omega = \frac{\partial \omega_{i_1 \dots i_p}}{\partial x^k} dx^k \wedge dx^{i_1} \wedge \dots \wedge dx^{i_p} \quad (\text{ba})$$

Gradient (denoted grad) is a function $\text{Scalar Field} \rightarrow \text{Vector Field}$, defined as $\text{grad} = f \rightarrow (df)^\sharp$. ∇f is also used as a notation for $\text{grad}(f)$. In Cartesian coordinates,

$$\begin{aligned} \text{grad} &= ((x_1, \dots, x_d) \rightarrow f(x_1, \dots, x_d)) \\ &\rightarrow \left((x_1, \dots, x_d) \rightarrow \frac{\partial f(x_1, \dots, x_d)}{\partial x_i} \hat{x}_i \right) \end{aligned} \quad (\text{bb})$$

Divergence (denoted div) is a function $\text{Vector Field} \rightarrow \text{Scalar Field}$, defined as $\text{grad} = v \rightarrow (\star d \star v^\flat)^\sharp$. $\nabla \cdot v$ is also used as a notation for $\text{div}(v)$. In Cartesian coordinates,

$$\begin{aligned} \text{div} &= ((x_1, \dots, x_d) \rightarrow v^i(x_1, \dots, x_d) \hat{x}_i) \\ &\rightarrow \left((x_1, \dots, x_d) \rightarrow \frac{\partial v^i(x_1, \dots, x_d)}{\partial x^i} \right) \end{aligned} \quad (\text{bc})$$

Curl (denoted curl) is a function $\text{Vector Field} \rightarrow (d-2) - \text{Vector Field}$, defined as $\text{curl} = v \rightarrow (\star dv^\flat)^\sharp$. In $d=3$, $\nabla \times v$ is also used as a notation for $\text{curl}(v)$; in Cartesian coordinates,

$$\nabla \times v = \left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z} \right) \hat{x} + \left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x} \right) \hat{y} + \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} \right) \hat{z} \quad (\text{bd})$$

Laplacian (for our purposes) is a function $\text{Tensor Field} \rightarrow \text{Tensor Field}$, denoted as Δ (sometimes also as ∇^2), and is defined in Cartesian coordinates as

$$\begin{aligned} R &: TM \otimes \dots \otimes TM \otimes T^*M \otimes \dots \otimes T^*M \\ R &= R^{i_1 \dots i_r}_{k_1 \dots k_s} \frac{\partial}{\partial x^{i_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{k_1} \otimes \dots \otimes dx^{k_s} \\ \Delta R &: TM \otimes \dots \otimes TM \otimes T^*M \otimes \dots \otimes T^*M \\ \Delta R &= \frac{\partial^2 R^{i_1 \dots i_r}_{k_1 \dots k_s}}{\partial x^m \partial x_m} \frac{\partial}{\partial x^{i_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{k_1} \otimes \dots \otimes dx^{k_s} \end{aligned} \quad (\text{be})$$

For instance, in three dimensional Euclidean space

$$\Delta \vec{V}(x, y, z) = \frac{\partial^2 V_x(x, y, z)}{\partial x^2} \hat{x} + \frac{\partial^2 V_y(x, y, z)}{\partial y^2} \hat{y} + \frac{\partial^2 V_z(x, y, z)}{\partial z^2} \hat{z} \quad (\text{bf})$$

when acting on a vector field $\vec{V}(x, y, z)$.

Differentiation identities follow from the following three properties of the exterior algebra: $d^2 = 0$, $(a^\flat)^\sharp = (a^\sharp)^\flat = a$, and $\star \star a = a$. These imply

$$\text{curl}(\text{grad}(f)) = \text{div}(\text{curl}(v)) = 0 \quad (\text{bg})$$

for any f and v , which in three dimensional Euclidean spaces can also be denoted as

$$\nabla \times \nabla f = \nabla \cdot \nabla \times v = 0 \quad (\text{bh})$$

Product rule for differentiation for vector fields follow from Leibniz rule for the exterior derivative, i.e.

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^p \omega \wedge d\eta \quad (\text{bi})$$

for the p -form ω . For 3d Euclidean spaces, this implies

$$\begin{aligned} \nabla \cdot (ab) &= \nabla a \cdot b + a \nabla \cdot b, & \nabla \times (ab) &= \nabla a \times b + a \nabla \times b \\ \nabla(a \cdot b) &= (a \cdot \nabla)b + (b \cdot \nabla)a + a \times (\nabla \times b) + b \times (\nabla \times a) \\ \nabla \cdot (a \times b) &= b \cdot (\nabla \times a) - a \cdot (\nabla \times b) \\ \nabla \times (a \times b) &= a(\nabla \cdot b) - b(\nabla \cdot a) + (b \cdot \nabla)a - (a \cdot \nabla)b \end{aligned} \quad (\text{bj})$$

A harmonic tensor field R (a harmonic function being a special case as a harmonic type $(0,0)$ tensor field) is an element of the kernel of the Laplacian, i.e. $\Delta R = 0$.

Helmholtz decomposition of a 3d vector field E is a way of rewriting it in terms of its *scalar potential* Φ (related to the divergence of vector field) and its *vector potential* V (related to the curl of vector field): $E = c - \nabla\Phi + \nabla \times V$ for a constant vector c where

$$\begin{aligned} \Phi(r) &= \frac{1}{4\pi} \int_{\text{manifold}} \frac{\nabla' \cdot E(r')}{|r - r'|} dV' - \frac{1}{4\pi} \oint_{\text{boundary}} \frac{\hat{n}' \cdot E(r')}{|r - r'|} dS' \\ V(r) &= \frac{1}{4\pi} \int_{\text{manifold}} \frac{\nabla' \times E(r')}{|r - r'|} dV' - \frac{1}{4\pi} \oint_{\text{boundary}} \frac{\hat{n}' \times E(r')}{|r - r'|} dS' \end{aligned} \quad (\text{bk})$$

Vectors in 3d Euclidean Space satisfy following formula

$$a \cdot (b \times c) = b \cdot (c \times a) = c \cdot (a \times b) \quad (\text{bl})$$

$$a \times (b \times c) = (a \cdot c)b - (a \cdot b)c \quad (\text{bm})$$

$$(a \times b) \cdot (c \times d) = (a \cdot c)(b \cdot d) - (a \cdot d)(b \cdot c) \quad (\text{bn})$$

Arc-length is the length of a curve (denoted by s), which satisfies $s(t) = \int_{t_0}^t \left| \frac{dx(t')}{dt'} \right| dt'$. In this equation, $x(t)$ is the position of a point on the curve, t is the parametrization parameter, and t_0 is the value of t at the starting point of the curve. The arc-length itself can be used to parametrize the curve.

Tangent vector to a curve in the arc-length parametrization is the function $t(s) = \frac{dx(s)}{ds}$. It has unit norm, and can be likened to the ratio velocity per speed.

Curvature of a curve κ is a function of the arc-length whose value is $\kappa(s) = \left| \frac{dt(s)}{ds} \right|$.

Principle normal of a curve n is a function of the arc-length whose value is $n(s) = \frac{1}{\kappa(s)} \frac{dt(s)}{ds}$. It has unit norm, and can be likened to the acceleration unit vector.

Binormal vector of a curve b is a function of the arc-length whose value is $b(s) = t(s) \times n(s)$ ($|b(s)| = 1$).

Torsion of a curve τ is a function of the arc-length whose value is $\tau(s) = -n(s) \cdot \frac{db(s)}{ds}$.

The Frenet-Serret equations is a closed system of equations which completely determine the properties of a curve as a function of the curvature and torsion functions:

$$\begin{aligned} \frac{dt(s)}{ds} &= \kappa(s)n(s), & \frac{db(s)}{ds} &= -\tau(s)n(s), \\ \frac{dn(s)}{ds} &= \tau(s)b(s) - \kappa(s)t(s) \end{aligned} \quad (\text{bo})$$

Generalized Stokes theorem equates the integration of a p -form ω over the boundary of a manifold ∂M to the integration of the exterior derivative of the p -form $d\omega$ over the manifold M : $\int_{\partial M} \omega = \int_M d\omega$.

Integral theorems are special cases of the generalized Stokes theorem. For a volume $V \in \mathbb{R}^3$, a surface $S \in \mathbb{R}^3$, a curve $\gamma \in \mathbb{R}^3$, and a region $D \in \mathbb{R}^2$ (and for the notation ∂A being boundary of A), we have

$$\begin{aligned} \int_V \nabla \cdot F dV &= \oint_{\partial V} F \cdot dS, & \int_S (\nabla \times F) \cdot dS &= \oint_{\partial S} F \cdot d\Gamma \\ \int_\gamma \nabla f \cdot dr &= f \Big|_{\text{initial}}^{\text{final}}, & \int_D \left(\frac{\partial M(x, y)}{\partial x} - \frac{\partial L(x, y)}{\partial y} \right) dx dy \\ & & = \oint_{\partial D} (L(x, y)dx + M(x, y)dy) \end{aligned} \quad (\text{bp})$$

Further identities can be derived by imposing $F = \phi(x, y, z)c$ for the constant vector field c or similar constraints.

Spherical and Cylindrical coordinates in $\mathbb{R}^3$ are defined in terms of the Cartesian coordinates (x, y, z) as

$$z = r \cos(\theta), \quad y = x \tan(\phi) \quad (\text{bq})$$

$$x = z \tan(\theta) \cos(\phi), \quad (\text{br})$$

for the spherical coordinates (r, θ, ϕ) and as

$$x = r \cos(\theta), \quad y = r \sin(\theta) \quad (\text{bs})$$

for the cylindrical coordinates (r, θ, z) .

Polar coordinates in $\mathbb{R}^d$ $(r, \theta_1, \dots, \theta_{d-1})$ are defined in terms of the Cartesian coordinates $(x_1, \dots, x_d)$ as

$$x_1 = r \cos(\theta_1), \quad x_d = x_{d-1} \tan(\theta_{d-1}) \quad (\text{bt})$$

$$x_i = x_{i-1} \tan(\theta_{i-1}) \cos(\theta_i) \quad \text{for } 1 < i < d \quad (\text{bu})$$

In two-dimensions, this reduces to the familiar polar coordinates $(x, y) = (r \cos \theta, r \sin \theta)$; in 3 dimensions, it becomes the familiar spherical coordinates for $(x_1, x_2, x_3) = (z, x, y)$ and $(\theta_1, \theta_2) = (\theta, \phi)$. In higher dimensions, polar coordinates are also called *hyperspherical coordinates*.

Cylindrical coordinates in $\mathbb{R}^d$ $(r, \theta_1, \dots, \theta_{n-1}, x_n, x_{n+1}, \dots, x_d)$ is a coordinate system such that a subset $\mathbb{R}^n$ of the total space $\mathbb{R}^d$ (for $n < d$) is converted into the polar coordinates. For instance, if we convert $\mathbb{R}^2$ of $\mathbb{R}^3$ into polar coordinates, we obtain the familiar 3d *cylindrical coordinates*, i.e. $(x, y, z) = (r \cos \theta, r \sin \theta, z)$.

(Anti)holomorphic function of a complex variable is a function f for which the derivative with respect to z ($\bar{z}$) is uniquely defined, i.e.

$$\begin{aligned} \frac{df(x, y)}{dz} &:= \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{f(x + \Delta x, y + \Delta y) - f(x, y)}{\Delta x + i\Delta y} \\ \left(\frac{df(x, y)}{d\bar{z}} &:= \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{f(x + \Delta x, y + \Delta y) - f(x, y)}{\Delta x - i\Delta y} \right) \end{aligned} \quad (\text{bv})$$

is well-defined and independent of the order of limits (this condition leads to *Cauchy-Riemann equations*). As any antiholomorphic function can be written as *complex conjugate of a holomorphic function*, one usually focuses on the analysis of holomorphic functions alone.

An analytic function is a function expandable as a convergent power series. Cauchy's integral formula ensures that a *complex analytic function* (with a series expansion in z) is equivalent to a *holomorphic function*.

Cauchy's integral formula for a function f that is complex-analytic in the region $D \subset \mathbb{C}$ can be written as

$$f(z) = \oint_{\partial D} \frac{f(\omega)}{\omega - z} \frac{d\omega}{2\pi i} \quad (\text{bw})$$

Laurent series of a function that is complex analytic for $R_1 < |z - a| < R_2$ is the convergent series expansion $f(z) = \sum_{n=-\infty}^{\infty} c_n(z - a)^n$.

A pole of a complex analytic function f is a point $a \in \mathbb{C}$ such that $f(a)$ is singular and S is a non-empty set for $S = \{m \in \mathbb{Z} \mid (z - a)^m f(z) \text{ is analytic at } a\}$. $\min(S)$ is called the order of the pole.

A zero of a function f is the value a such that $f(a) = 0$. $\min\{m \in \mathbb{Z} \mid \lim_{z \rightarrow a} (z - a)^{-m} f(z) \neq 0\}$ is called order of zero.

Riemann sphere (denoted $\hat{\mathbb{C}}$) is the *compactification* of the complex plane $\mathbb{C}$. More simply, it is the inclusion of *infinity* as a single point to the complex plane, $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, such that we get a complete symmetry between large numbers and small numbers: large numbers are points close to the *north pole* (by convention), whereas small numbers are points close to the *south pole*; the map $z \rightarrow \frac{1}{z}$ sends such numbers to each other and switches north and south poles (hence $\infty \leftrightarrow 0$).

A meromorphic function f in a domain D is a holomorphic function in D except a set of points at which f has a pole. For example, $z \rightarrow \frac{1}{\sin(z)}$, $z \rightarrow \frac{e^z}{z}$ are meromorphic functions in $D = \mathbb{C}$. If we also include infinity ($D = \mathbb{C} \cup \{\infty\}$), they are no longer meromorphic as they are singular at infinity but that singularity is not a pole. In fact, *the only meromorphic functions in $D = \mathbb{C} \cup \{\infty\}$* are rational functions, e.g. $z \rightarrow \frac{(z-1)(z+i)}{(2z+\pi)(z-i+1)}$.

Residue of a complex function at an isolated singularity a is defined as

$$\text{Res} : (\mathbb{C} \rightarrow \mathbb{C}, \mathbb{C}) \rightarrow \mathbb{C} \quad (\text{bx})$$

$$\text{Res}(f, a) = \frac{1}{2\pi i} \oint_{C_a} f(z) dz \quad (\text{by})$$

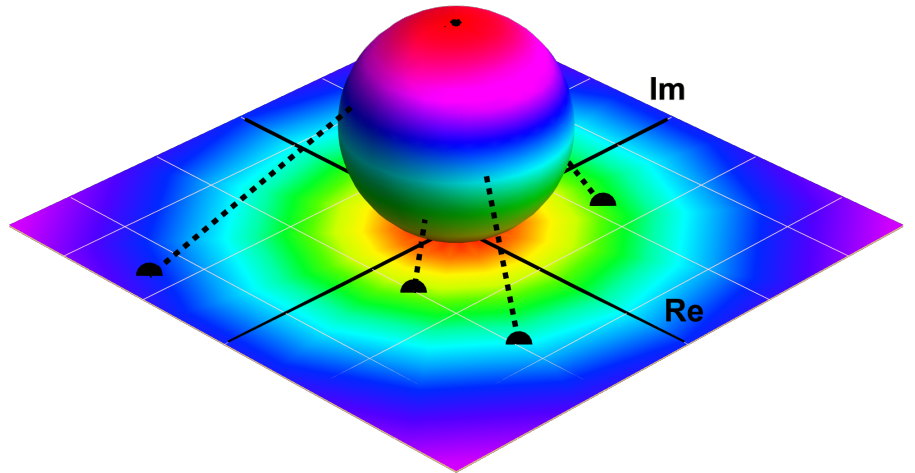
for an infinitesimal closed contour C_a centered at a .

Cauchy's principal value (denoted p.v.) is for our purposes defined via the relation

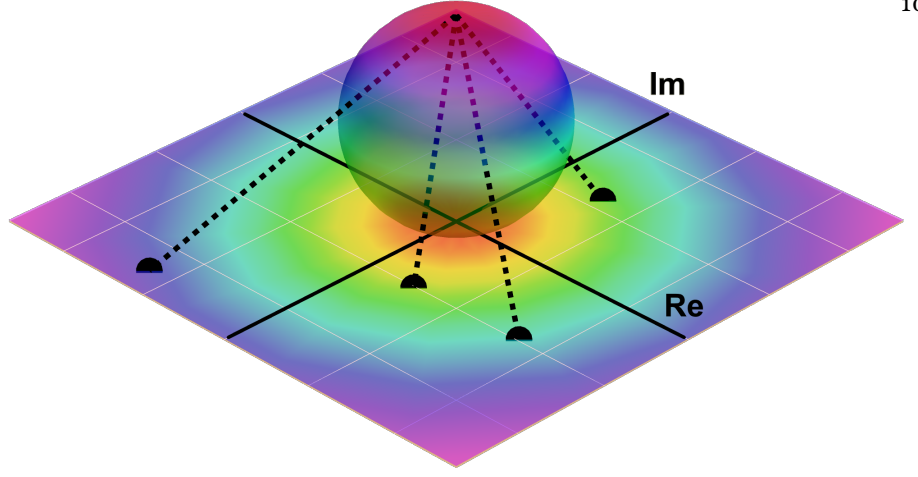
$$\text{p.v.} \int_a^c f(x) dx = \lim_{\epsilon \rightarrow 0} \left[\int_a^{b-\epsilon} f(x) dx + \int_{b+\epsilon}^c f(x) dx \right] \quad (\text{bz})$$

for $a < b < c$, where $f(x)$ is assumed to be analytic in $[a, c] \setminus \{b\}$. If $f(x)$ is analytic at b , the principle value gives the same result with the ordinary integral; on the other hand, if $f(x)$ is not analytic at b , the principle value assigns a well-defined value to the integral which would be otherwise ill-defined as a function.

Conformal transformation (for our purposes) is any mapping $x \rightarrow x'$ of the coordinates for which the angles between (co)tangent vectors do not change, e.g. $\frac{\langle dx, dy \rangle}{\sqrt{|dx||dy|}} = \frac{\langle dx', dy' \rangle}{\sqrt{|dx'||dy'|}}$; for instance, translation $x' = x + a$, rotation $x' = e^{i\theta} x$ or scaling $x' = \lambda x$ are so.



Stereographic projection is a conformal mapping between $\mathbb{R}^d$ (d -dimensional plane) and S^d (d -sphere); however, we are only interested in the map between $\mathbb{R}^2$ (the complex plane) and S^2 (the Riemann sphere). Geometrically, the mapping can be readily applied as follows: **(1)** embed S^2 and $\mathbb{R}^2$ into $\mathbb{R}^3$ such that the origin of $\mathbb{R}^2$ and south pole of S^2 coincide; **(2)** draw a line from north pole to a point $z \in \mathbb{R}^2$; **(3)** the intersection of the line with S^2 gets mapped to z .



2 Special functions and definite integrals

This section contains a brief list of some important special functions used in Physics and how these functions are related to certain definite integrals. You should understand that these functions are very well-studied and there exists a substantial literature regarding their properties and numerical values; therefore, one should try to rewrite any given integral in terms of these special functions if possible.

Floor (ceiling) function is denoted as $\lfloor x \rfloor$ ($\lceil x \rceil$) and gives the largest (smallest) integer less (greater) than or equal to x .

(Incomplete) Gamma function (denoted Γ) are $(\mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C})$ $\mathbb{C} \rightarrow \mathbb{C}$ functions defined via the integrals

$$\begin{aligned} \Gamma(z) &:= \int_0^\infty t^{z-1} e^{-t} dt \quad \text{for } \operatorname{Re}(z) > 0 \\ \Gamma(z, a) &:= \int_a^\infty t^{z-1} e^{-t} dt \quad \text{for } \operatorname{Re}(z) > 0 \end{aligned} \quad (\text{ca})$$

where value of $\Gamma(z)$ at anywhere else on $\mathbb{C}$ (except the negative integers, i.e. its poles) can be computed iteratively using

$$\Gamma(z) = \Gamma(z+1)/z \quad (\text{cb})$$

and analogously for the incomplete gamma function.

(Double) Factorial are $\mathbb{C} \rightarrow \mathbb{C}$ functions, which for positive integer input satisfy the relations $n! = n \times (n-1) \times \dots \times 1$ and $n!! = n \times (n-2) \times \dots \times (1 \text{ or } 2)$. For more general input

$$z! := \Gamma(z+1) \quad (\text{cc})$$

whereas the analog for $z!!$ is too complicated for our purposes.

Pochhammer symbol is a binary function defined as

$$(a)_b := \frac{\Gamma(a+b)}{\Gamma(a)} \quad (\text{cd})$$

whenever the right hand side is well defined. Although multiple contradicting definitions (see *falling and rising factorials*) exist, we will work with the conventions of `Mathematica`.

Binomial symbol is a binary function defined as

$$\binom{n}{k} := \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \quad (\text{ce})$$

whenever the right hand side is well defined. For positive integers with $n \geq k$, this has the combinatorial meaning of “number of different ways to choose k elements out of n elements”.

Legendre polynomial of type 1 is defined for $|z| \leq 1$:

$$P_\ell(z) = \begin{cases} 2^{-\ell} \sum_{k=0}^{\lfloor \ell/2 \rfloor} (-1)^k \binom{\ell}{k} \binom{2\ell-2k}{\ell} z^{\ell-2k} & \text{if } \ell \in \mathbb{N} \\ P_{-\ell-1}(z) & \text{if } \ell \in \mathbb{Z}^- \end{cases} \quad (\text{cf})$$

with their role in the expansion of norms of vector differences:

$$\frac{1}{|\mathbf{x} - \mathbf{y}|} = \sum_{\ell=0}^{\infty} \frac{r_{<}^\ell}{r_{>}^{\ell+1}} P_\ell(\hat{\mathbf{x}} \cdot \hat{\mathbf{y}}) \quad \text{for } \begin{cases} r_{>} = \max(|\mathbf{x}|, |\mathbf{y}|) \\ r_{<} = \min(|\mathbf{x}|, |\mathbf{y}|) \end{cases} \quad (\text{cg})$$

Associated Legendre polynomial of type 1 is defined for $|z| \leq 1$ as

$$P_\ell^m(z) = \frac{(1+z)^{m/2}}{(1-z)^{m/2}} \sum_{k=0}^{\ell} \frac{(-\ell)_k (\ell+1)_k}{\Gamma(k-m+1)k!} \left(\frac{1-z}{2} \right)^k \quad \text{if } \ell \in \mathbb{N} \quad (\text{ch})$$

with the relation $P_\ell^0(z) = P_\ell(z)$ and the orthogonality property

$$\int_{-1}^1 P_\ell^m(x) P_n^m(x) dx = \frac{2}{2\ell+1} \frac{(\ell+m)!}{(\ell-m)!} \delta_{n,\ell} \quad (\text{ci})$$

(Associated) Legendre polynomial of type 2 & 3 are analytic continuations of $P_\ell(z)$ and $P_\ell^m(z)$ beyond $|z| \leq 1$.

Spherical harmonics are defined for $\ell, |m| \in \mathbb{N}$ & $|m| \leq \ell$:

$$Y_\ell^m(\theta, \phi) = \sqrt{\frac{(2\ell+1)(\ell-m)!}{4\pi(\ell+m)!}} e^{im\phi} P_\ell^m(\cos(\theta)) \quad (\text{cj})$$

with the orthonormality and completeness conditions as

$$\int_0^{2\pi} d\phi \int_0^\pi \sin(\theta) d\theta \left[Y_{\ell'}^{m'}(\theta, \phi) \right]^* Y_\ell^m(\theta, \phi) = \delta_{\ell, \ell'} \delta_{m, m'}$$

$$\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[Y_\ell^m(\theta', \phi') \right]^* Y_\ell^m(\theta, \phi) = \delta(\phi - \phi') \delta(\cos(\theta) - \cos(\theta')) \quad (\text{ck})$$

Polygamma function (denoted Ψ) is defined as

$$\Psi^{(n)}(z) = \frac{d^{n+1}}{dz^{n+1}} \log(\Gamma(z)) \quad (\text{cl})$$

which is equivalent to the *Laplace transform* of $\frac{(-1)^{n+1} t^n}{1 - e^{-t}}$:

$$\Psi^{(n)}(s) = \int_0^\infty \frac{(-1)^{n+1} t^n}{1 - e^{-t}} e^{-st} dt \quad \text{for } \text{Re}(s) > 0 \quad (\text{cm})$$

Among many other applications, polygamma functions directly appear for expectation values of observables in statistical mechanics, with partition functions being combinations of gamma functions.

Beta function (denoted B) is defined as

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (\text{cn})$$

which is equivalent to the following integral

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \quad \text{for } \text{Re}(a), \text{Re}(b) > 0 \quad (\text{co})$$

Besides its appearance in advanced areas (regularization in quantum field theory, scattering amplitudes in string theory, etc.), beta function also arises in d -dimensional angular integrals in polar coordinates as for instance

$$\int_0^\pi (\sin \theta)^n d\theta = B\left(\frac{n+1}{2}, \frac{1}{2}\right) \quad \text{for } \text{Re}(n) > -1 \quad (\text{cp})$$

3 Concept of classical probability

This brief section aims to alleviate the absence of a formal probability training for the physics students. The content is far less rigorous than a proper measure-theoretical approach but is sufficient for all undergraduate and most (if not all) graduate students. The title says classical because an extension of these concepts to quantum world is possible via the analysis of von Neumann algebra's (in fact, classical probability is the special case where the von Neumann algebra is abelian) but it is beyond the scope of these notes.

Probability is a somewhat ambiguous term as different interpretations of the concept assign different meanings: *frequentists* relate probability to empirical statistics of repeated events, *subjectivists* to personal beliefs, *objective Bayesians* to universal rational beliefs constrained by principles, among others. In these notes, we will take an *axiomatic approach* (based on Kolmogorov axioms) with a tendency towards Bayesian interpretation, but will gloss over any philosophical subtleties.

Sample space (denoted Ω), for our purposes, is the set of all possible outcomes for a physical (or a thought) experiment.

Error function (denoted erf) is a $\mathbb{C} \rightarrow \mathbb{C}$ function defined

$$\text{erf}(z) := \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad (\text{cq})$$

with the limit $\lim_{z \rightarrow \infty} \text{erf}(z) = 1$, meaning

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi} \quad (\text{cr})$$

Error function directly appears in *cumulative distribution function* for normal (also called Gaussian) probability distributions.

Fresnel integrals are $\mathbb{C} \rightarrow \mathbb{C}$ functions defined as

$$S(z) := \int_0^z \sin\left(\frac{\pi t^2}{2}\right) dt, \quad C(z) := \int_0^z \cos\left(\frac{\pi t^2}{2}\right) dt \quad (\text{cs})$$

which are linear combinations of the error function via Euler's formula. This leads to the following numerical results

$$\int_0^\infty \sin\left(\frac{\pi t^2}{2}\right) dt = \int_0^\infty \cos\left(\frac{\pi t^2}{2}\right) dt = \frac{1}{2} \quad (\text{ct})$$

Exponential integral (denoted E) is a $\mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}$ function

$$E_\nu(z) := \int_1^\infty \frac{e^{-zt}}{t^\nu} dt \quad \text{for } \text{Re}(z) > 0 \quad (\text{cu})$$

Event space (denoted F) is, loosely speaking, a set of all possible events that can be assigned a probability. For a finite size sample space Ω , the event space F is the set of all possible subsets of Ω ; for infinite sets, the rigorous treatment requires σ -algebras which we do not need to know. For example, for sample space $\Omega = \{\text{AA}, \text{BA}, \text{BB}, \text{CB}, \text{CC}, \text{DC}, \text{DD}, \text{FD}, \text{FF}, \text{NA}\}$, the element $x = \{\text{FD}, \text{FF}, \text{NA}\}$ satisfies $x \in F$ and can be interpreted as the event “*getting a failing grade*”.

Probability measure (denoted $\mathbb{P}$) is a function $F \rightarrow [0, 1]$ (from event space to a unit interval element), which satisfies **(1)**

$P(\emptyset) = 0$ & $P(\Omega) = 1$, and **(2)** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Random variable (denoted with latin capital letters, e.g. $X, Y, \dots$) is technically a function from sample space Ω to a measurable field (we will take as $\mathbb{R}$). In most situations of interest to Physics, Ω itself is a subset of $\mathbb{R}$, hence one loosely calls the *output of the function* X as the random variable instead of the function itself; for instance:

(a) Consider $\Omega = \{AA, BA, BB, CB, CC, DC, DD, FD, FF, NA\}$, with $X(AA) = 4$, $X(BA) = 3.5$, and so on: the function X is a random variable as originally defined above.

(b) Consider $\Omega = \{1, 2, 3, 4, 5, 6\}$, i.e. potential outcomes of a dice roll. Instead of the function $X : \Omega \rightarrow \mathbb{R}$, one usually calls $X \in \Omega \subset \mathbb{R}$ “the random variable”, with X taking values $1, 2, \dots$.

In either convention, random variables X_1, X_2 can be added or scaled; as functions, $X_1 + X_2$ and λX simply denote the maps $\omega \rightarrow X_1(\omega) + X_2(\omega)$ and $\omega \rightarrow \lambda X(\omega)$ respectively. This generalizes to arbitrary functions of random variables $f(X_1, \dots, X_n)$.

We will also abuse the notation and denote probabilities with the notation $P(X > x)$ which technically is a shorthand notation for $P(\{\omega \in \Omega \mid X(\omega) > x\})$, and similarly for the sign $<$.

Probability density function (PDF) of a random variable X (denoted f_X) is, for our purposes, a generalized function (also called distribution) which tells us how likely a particular output of the random variable X is. Probability measure on Ω uniquely fixes the probability density function of X once the random variable $X : \Omega \rightarrow \mathbb{R}$ is chosen, i.e. $\int_{\Omega}(\dots)dP(\omega) \sim \int_{\mathbb{R}}(\dots)f_X(x)dx$. For instance, for a fair dice, we have $f_X(x) = \frac{1}{6} \sum_{n=1}^6 \delta(x - n)$ for Dirac-Delta distribution δ . Conservation of total probability ensures that for any PDF f ,

$$f_{aX}(x) = \frac{1}{|a|} f_X\left(\frac{x}{a}\right), \quad f_{X_1+X_2}(s) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x, s-x) dx \quad (\text{cv})$$

for the joint probability density function f_{X_1, X_2} .

Joint probability density function of random variables $X_1, \dots, X_n$ (denoted $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$) is, loosely speaking, the probability of observing the values x_i for the random variables $X_i : \Omega \rightarrow \mathbb{R}$ with $\int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n = 1$. The event space Ω may or may not be of form $\Omega_1 \times \Omega_2 \times \dots \times \Omega_n$.

(Joint) Marginal probability density function of random variables $X_1, \dots, X_k$ (denoted $f_{X_1, \dots, X_k}(x_1, \dots, x_k)$) is, obtained from a joint probability distribution $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ for $n > k$ by integrating remaining random variables:

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) := \int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_{k+1} \dots dx_n \quad (\text{cw})$$

(Joint) (Marginal) Cumulative density function of random variables $X_1, \dots, X_k$ (denoted $F_{X_1, \dots, X_k}$, of type $\mathbb{R}^k \rightarrow \mathbb{R}$, abbreviated as CDF) is defined for our purposes via

$$F_{X_1, \dots, X_k}(x_1, \dots, x_k) := \int_{-\infty}^{x_1} dy_1 \dots \int_{-\infty}^{x_k} dy_k f_{X_1, \dots, X_k}(y_1, \dots, y_k) \quad (\text{cx})$$

Conditional (joint) probability density function

of random variables $X_1, \dots, X_n$ is the probability of observing $x_1, \dots, x_k$ *provided that* $x_{k+1}, \dots, x_n$ have been observed:

$$f_{X_1, \dots, X_k \mid X_{k+1}, \dots, X_n}(x_1, \dots, x_k \mid x_{k+1}, \dots, x_n) := \begin{cases} \frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}{f_{X_{k+1}, \dots, X_n}(x_{k+1}, \dots, x_n)} & \text{if } f_{X_{k+1}, \dots, X_n}(x_{k+1}, \dots, x_n) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{cy})$$

for the joint pdf $f_{X_1, \dots, X_n}$ and the joint marginal pdf $f_{X_{k+1}, \dots, X_n}$.

(Conditional) Independence of two sets of random variables is the idea that the measurements of these sets are not affected by one another (possibly conditioned on a third set of random variables). Let X, Y, Z denote three ordered sets of random variables, i.e. $X \doteq (X_1, \dots, X_n)$. Then, if for all z ,

$$f_{X, Y \mid Z}(x, y \mid z) = f_{X \mid Z}(x \mid z) \times f_{Y \mid Z}(y \mid z) \quad (\text{cz})$$

then “ x is conditionally independent of y , given z ”. When there is no condition, we say “ x is independent of y ” for

$$f_{X, Y}(x, y) = f_X(x) \times f_Y(y) \quad (\text{da})$$

(Conditional) Expectation value (denoted E) of the function $g(X_1, \dots, X_n)$ for the random variables $X_1, \dots, X_n$ is its integration against the (conditional) joint pdf:

$$E[g(X_1, \dots, X_n) \mid Z_1 = z_1, \dots, Z_k = z_k] := \int_{\mathbb{R}^n} g(x_1, \dots, x_n) \times f_{X_1, \dots, X_n \mid Z_1, \dots, Z_k}(x_1, \dots, x_n \mid z_1, \dots, z_k) dx_1 \dots dx_n \quad (\text{db})$$

The simplest such expectation value is “expectation value of the random variable X ” which is simply $E[X] = \int_{\mathbb{R}} x f_X(x) dx$.

(Mixed) Moment of order k ($\vec{k}$) of random variable(s) X ($X_1, \dots, X_n$) is denoted m and is the expectation value $m_k := E[X^k]$ ($m_{\vec{k}} := E[X_1^{k_1} X_2^{k_2} \dots X_n^{k_n}]$).

Mean of a random variable X is the $m_1 = E[X]$.

Central (mixed) moment of order k ($\vec{k}$) of random variable X ($X_1, \dots, X_n$) is $\mu_k := E[(X - E[X])^k]$ ($\mu_{\vec{k}} := E[(X_1 - E[X_1])^{k_1} \dots (X_n - E[X_n])^{k_n}]$).

Variance of a random variable X is the μ_2 for X .

Standard deviation (denoted σ) is defined as $\sigma := \sqrt{\mu_2}$.

Covariance of the random variables X_1 and X_2 is the central moment of order $(1, 1)$, i.e. $\mu_{1,1}$.

Correlation coefficient between two random variables (denoted ρ) is defined as $\rho(X, Y) := \frac{\mu_{1,1}(X, Y)}{\sigma(X)\sigma(Y)}$ with $|\rho| \leq 1$.

Uncorrelated random variables are those for which $\rho(X, Y) = 0$. Independent random variables *are always* uncorrelated but not all uncorrelated variables are independent.

(Mixed) Moment generating function (MGF) of a random variable X ($X_1, \dots, X_n$) is defined as

$$M_{X_1, \dots, X_k}(t_1, \dots, t_k) := \sum_{k_1=0}^{\infty} \dots \sum_{k_n=0}^{\infty} \frac{t_1^{k_1} \dots t_n^{k_n}}{k_1! \dots k_n!} m_{\mathbf{k}} = \mathbb{E} \left[e^{t_1 X_1 + \dots + t_k X_k} \right] \quad (\text{dc})$$

whenever this sum converges. There are probability distributions (such as Log-normal or Laplace) where all moments exist and are finite, but MGF is nevertheless undefined.

It is interesting to note that MGF's, apart from their direct use in Physics, mathematically resemble (and generalize) to the partition functions in statistical mechanics and vacuum generating functions in quantum field theory.

(Mixed) Cumulant generating function (CGF) is defined as $K_{X_1, \dots, X_n}(t_1, \dots, t_n) := \log \left(M_{X_1, \dots, X_n}(t_1, \dots, t_n) \right)$.

(Mixed) Cumulant of order k ($\vec{k}$) of random variable X ($X_1, \dots, X_n$) is denoted κ and can be defined as

$$\kappa_{\vec{k}} := \left(\frac{d}{dt_1} \right)^{k_1} \dots \left(\frac{d}{dt_1} \right)^{k_n} K_{X_1, \dots, X_n}(t_1, \dots, t_n) \Big|_{t_1=\dots=t_n=0} \quad (\text{dd})$$

This definition requires K (hence M) to be well-defined, but we can still *define* cumulants by combining moments through Bell polynomials even if MGF and CGF do not exist.

Just like moments (likewise MGF and CGF), the cumulants are also the mathematical analogs of certain Physical quantities; interestingly, *connected Feynman diagrams* are generalizations of cumulants to the quantum field theoretical context.

Independent and identically distributed (iid)

random variables form the relevant framework for the modeling of results in repeated experiments of assumed time-invariant nature (a dice rolled three times), or of cumulative results of multiple setups assumed to be independent (three dices rolled at once).

Tail of a distribution describes the behavior as the absolute value of the random variable goes to infinity. One can quantitatively describe it via an index α (called *tail index*) as

$$\alpha := \log \left(\lim_{x \rightarrow \infty} \frac{\mathbb{P}(|X| > x)}{\mathbb{P}(|X| > ex)} \right) \quad (\text{de})$$

Stable distributions are probability distributions such that the random variable $Y := \frac{X_1 + X_2 + \dots + X_n - a_n}{b_n}$ (for some ordinary real numbers a_n and b_n) has the same PDF with X_i where X_i are iid random variables. With some computation beyond the scope of these notes (start with characteristic function, use Gil-Pelaez inversion formula to get CDF, differentiate to get PDF), one can show that all stable distributions have to have PDF of the form

$$f_Y(y) = \frac{1}{\pi} \int_0^{\infty} e^{-\gamma^\alpha t^\alpha} \cos(-\gamma^\alpha t^\alpha \beta \Phi_\alpha(t) + (\delta - y)t) dt \quad (\text{df})$$

for tail index $0 < \alpha \leq 2$, and $-1 \leq \beta \leq 1$, $\gamma > 0$, $\delta \in \mathbb{R}$ where

$$\Phi_\alpha(t) = \begin{cases} \frac{2}{\pi} \log(t) & \text{if } \alpha = 1 \\ \tan\left(\frac{\pi\alpha}{2}\right) & \text{elsewhere} \end{cases} \quad (\text{dg})$$

Stable distributions act like *attractors* for other distributions; in particular, central limit theorem (discussed below in detail) guarantees that the probabilistic behavior of the average of many iid random variables almost always approaches to a stable distribution with $\alpha = 2$: such stable distributions are called *Gaussian distribution* (discussed below in detail).

Central Limit Theorem for iid (for physicists) basically means that the probability distribution describing (possibly modified) average of iid's most likely converges to a *stable distribution*. We can quantify this by considering the tail index α and the parameter v defined as

$$v := \lim_{x \rightarrow \infty} x^2 \mathbb{P}(|X| > x) \quad (\text{dh})$$

which measures *the importance of extreme values* (for finite v , extremes are negligible): for any given α and v , one of the following situation arises for (possibly modified) average of iid's:

	Finite v	Infinite v
$\alpha = 0$ (slowly varying tail)	Mathematically Impossible	No stable limit (extremes dominates)
$0 < \alpha < 2$ (regularly varying tail)	Mathematically Impossible	α -Stable distribution
$\alpha = 2$ (regularly varying tail)	Gaussian distribution	Gaussian distribution
$2 < \alpha \leq \infty$ (light tail)	Gaussian distribution	Mathematically Impossible
α does not exist! (oscillatory tail)	Gaussian distribution	No stable limit (extremes dominates)

Central Limit Theorem for non-iid is basically this: some independent distributions would still average to a Gaussian distribution even if they are not identically distributed. In particular, if there are n independent random variables X_i with tail index $\alpha_i > 2$ (hence each random variable has a light tail),

$$Y \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1) \quad \text{for } Y = \left(\sum_{i=1}^n \mu_2^{(i)} \right)^{-1/2} \sum_{i=1}^n \left(X_i - m_1^{(i)} \right) \quad (\text{di})$$

where $m_1^{(i)}$ and $\mu_2^{(i)}$ are the mean and the variance of X_i respectively. If the sum includes a random variable with $\alpha \leq 2$, one then needs to check what is called the *Lindeberg's condition* to verify if central limit theorem still holds.

Gaussian distribution (also called normal distribution, denoted $\mathcal{N}$) is the unique stable distribution with finite variance. It is described by only two parameters, its mean m_1 and its variance μ_2 :

$$f_{\mathcal{N}(m_1, \mu_2)}(x) = \frac{\exp\left(-\frac{(x-m_1)^2}{2\mu_2}\right)}{\sqrt{2\pi\mu_2}} \quad (\text{dj})$$

which can be obtained from the general form of stable distributions in (df) by setting $\alpha = 2$, $\delta = m_1$, and $\gamma^2 = \mu_2/2$. The classical central limit theorem states that for iid random variables $X_{1,\dots,n}$ with finite mean m_1 and finite variance μ_2 , we have

$$Y \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \mu_2) \text{ for } Y = \frac{X_1 + X_2 + \dots + X_n - nm_1}{\sqrt{n}} \quad (\text{dk})$$

Zeta distribution is an example of a distribution with infinite variance that still approaches to a Gaussian distribution under a non-standard scaling. For the pdf

$$f_X(x) = \frac{1}{\zeta(3)} \sum_{i=1}^{\infty} \frac{\delta(x-i)}{x^3} \quad (\text{dl})$$

the variance of X is infinite, yet Central Limit Theorem holds:

$$Y \xrightarrow{n \rightarrow \infty} \mathcal{N}\left(0, \frac{1}{\zeta(3)}\right) \text{ for } Y = \frac{X_1 + X_2 + \dots + X_n - nm_1}{\sqrt{n \log(n)}} \quad (\text{dm})$$

where m_1 is mean of X and ζ is the Riemann-Zeta function.

Pareto distribution is an example of a distribution whose average does not approach to a Gaussian but an α -stable distribution. For the parameter range $1 < \alpha < 2$ and $t > 0$,

$$f_X(x) = \begin{cases} \alpha t^\alpha x^{-1-\alpha} & x \geq t \\ 0 & x < t \end{cases} \quad (\text{dn})$$

it has finite mean ($m_1 = \frac{\alpha t}{\alpha-1}$) but infinite variance. We have

$$Y \xrightarrow{n \rightarrow \infty} \text{Stable}(\alpha, \beta, \delta, \gamma) \text{ for } Y = \frac{X_1 + X_2 + \dots + X_n - nm_1}{n^{1/\alpha}} \quad (\text{do})$$

where α is the same one appearing in $f_X(x)$, $\delta = 0$ as Y has zero mean, and β & γ are to be determined explicitly.

Pareto distribution describes phenomena where *extreme events dominate the statistics*; for instance, high magnitude earthquakes are less probable but such rare events happen at

a power-law (not exponentially) suppressed frequency, meaning that their probability actually follows a Pareto distribution (this goes by the name Gutenberg–Richter law). In fact, we can see Pareto distribution in all natural phenomena where power law statistics take over, such as phase transitions in matter, turbulence in fluids, or galaxy formation in cosmology.

Cauchy distribution is an example of a probability distribution whose average does not approach to a Gaussian as it is itself an α -stable distribution. It has the pdf

$$f_X(x) = \frac{1}{\pi} \frac{b}{(x-a)^2 + b^2} \quad (\text{dp})$$

where variance is infinite and mean is undefined. To see this, change the codomain of the random variable from $\mathbb{R}$ to $[-\ell, r]$ with $\ell, r > 0$. The pdf for this *truncated region* then reads as

$$\hat{f}_X(x) = \frac{1}{\tan^{-1}\left(\frac{a+\ell}{b}\right) - \tan^{-1}\left(\frac{a-r}{b}\right)} \frac{b}{(x-a)^2 + b^2} \quad (\text{dq})$$

which integrates to 1 and satisfies $\lim_{\ell, r \rightarrow \infty} \hat{f}_X(x) = f_X(x)$. If we now compute the mean with this pdf and then take the limit, we would observe that the limits $\lim_{\ell \rightarrow \infty}$ and $\lim_{r \rightarrow \infty}$ do not commute: *the mean for Cauchy distribution does not exist!*

Cauchy distribution describes phenomena where the fluctuations are so strong that the mean is simply undefined. For instance, many biomedical imaging methods such as optical coherence tomography rely on photon transport models with non-standard photon diffusion, and some such models stabilize to Cauchy distribution under repeated sampling.

Law of large numbers is roughly the formulation of the intuition that the statistical average obtained through sampling a random variable would converge to the expectation value of the random variable as the sampling count goes to infinity. Obviously, this law does not apply if there is no mean to converge, with Cauchy distribution being an example.

4 Fundamental definitions in Physics

This section briefly reviews empirical foundations of fundamental physics. Unless stated otherwise, all definitions in this section are made in SI units.

Observable and physical (for our purposes) are to be understood in their colloquial meaning, and we will use them interchangeably (although they are not!). Whether something can be *directly observed* or *inferred from direct observables* and whether indirect observables are physical are beyond our scope and are in the realm of philosophy of science.

A (physical) object or a system is difficult to define because its existence assumes a way to unambiguously distinguish entities, i.e. we should be able to tell in principle (not necessarily in practice) if something is a part of a given system or not (which is tricky, especially within quantum information theory). We will use its loose self-evident definition and describe it through set theory to ensure it being well-defined; for

more relaxed depictions, one might look into *category theory*.

Phenomenological X is meant as X being defined more in line with observables than in line with mathematical description at the assumed fundamental level. Philosophically, for our purposes, fundamental descriptions tend to rely on *constructionism* and *reductionism* whereas phenomenological definitions tend to be free of such prerequisites.

Dimension of an observable A is a label that indicates the subset of all observables with which A is commensurable. For instance, *length* and *electric charge* are dimensions.

Unit of a dimension is a name indicating a standardized amount for the given dimension. Different units for the same dimension might describe different amounts, but are related by

a real number proportionality constant. For an observable A , $[A]$ denotes its dimension in any chosen unit, i.e. $[5 \text{ kg}] = \text{kg}$.

System of units is the pair $(S, \mathcal{F})$ where S is a set that contains a unique unit for each and all possible dimensions of all observables of interest and $\mathcal{F}$ is set of its fundamental units.

Fundamental units is any set $\mathcal{F}$ which is a subset of S such that *known physical laws* ensure an isomorphism between $(S, \mathcal{F})$ and (a vector space V , a basis of V).

Let's illustrate this. Assume that the observables of interest are such that we have only four dimensions: forces, lengths, durations, and masses, which are commensurable only among themselves. We choose four units for each: *Newton*, *meter*, *second*, and *kilogram* (abbreviated N, m, s, kg). Then, $S = \{\text{N, m, s, kg}\}$; and if we do not know any empirical physical law connecting these, we can choose $\mathcal{F} = S$ as well. However, Newton's second law creates a dependency among them, i.e. $[F = ma]$ means $\text{N} = \text{kg} \cdot \text{m} \cdot \text{s}^{-2}$. With this information, we can choose only 3 fundamental units, for instance $\mathcal{F} = \{\text{N, kg, m}\}$.

Derived units (denote D) is the complement set of $\mathcal{F}$ in S .

Dimensional analysis is loosely the study of physical phenomena through dimensions of the variables with the rules:

- $A + B$ is meaningless unless $[A] = c[B]$ for $c \in \mathbb{R}$
- $[A + B] = [A]$ unless $A + B$ is meaningless
- $[xy] = [x] \cdot [y]$
- $[x^n] = [x]^n$ for $n \in \mathbb{R}$

The first condition rules out power series, hence exponentiation, hence logarithm, hence generalized powers of dimensional quantities: $\log(5 \text{ N})$ or $(2 \text{ kg})^{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}$ do not make sense.

In addition to the mathematical requirements above, there is actually a stronger physical constraint: once we choose the fundamental units (say $f_1, \dots, f_n$), the unit of any observable is always of form $m_0 f_1^{m_1} f_2^{m_2} \dots f_n^{m_n}$ for $m_i \in \mathbb{R}$ if we accept that physics is independent of our choice of scales (i.e. mile or km).

Système international d'unités (SI units) is a system of units for all known natural phenomena. The physical laws discovered so far ensures that all experimental observables can be understood with seven fundamental units. In most of the physics, we only need the following five though:

Time	Length	Mass	Current	Temperature
second (s)	meter (m)	kilogram (kg)	ampere (A)	kelvin (K)

Remember that future discoveries may change the dimensions and fundamental units! For instance, a discovery of a new abelian gauge field may need its own fundamental current unit just like the electric current unit ampere.

Conversion between unit systems with the same number of fundamental units is simply a basis transformation. For instance, consider the *speed of light* c , *reduced Planck constant* $\hbar$, and *gravitational constant* G as our new fundamental

units. In SI units, they read as

$$\begin{aligned} c &= 299,792,458 \text{ m s}^{-1} \\ \hbar &= 662,607,015 \times 10^{-42} \text{ kg m}^2 \text{ s}^{-1} \\ G &= 66,743 \times 10^{-15} \text{ kg}^{-1} \text{ m}^3 \text{ s}^{-2} \end{aligned} \quad (\text{dr})$$

which can be inverted to yield

$$\begin{aligned} \text{kg} &\approx 5 \times 10^7 c^{1/2} \hbar^{1/2} G^{-1/2} \\ \text{m} &\approx 6 \times 10^{34} c^{-3/2} \hbar^{1/2} G^{1/2} \\ \text{s} &\approx 2 \times 10^{43} c^{-5/2} \hbar^{1/2} G^{1/2} \end{aligned} \quad (\text{ds})$$

We can now convert between these unit system; for instance, the gravitational acceleration g on the surface of Earth reads

$$g \approx 9.8 \text{ m s}^{-2} \approx 2 \times 10^{-51} c^{7/2} \hbar^{-1/2} G^{-1/2} \quad (\text{dt})$$

Natural units is the unit system where a subset of fundamental units are chosen to be constants of nature; in its most general form, an arbitrary SI unit $\text{s}^{a_1} \text{m}^{a_2} \text{kg}^{a_3} \text{A}^{a_4} \text{K}^{a_5}$ is rewritten as $e^{\alpha_1} c^{\alpha_2} \hbar^{\alpha_3} G^{\alpha_4} k_B^{\alpha_5}$ for the *proton charge* e and the *Boltzmann constant* k_B , where different variations are also known in literature (Planck natural units, atomic natural units, etc.). After the conversion, the units are often dropped; e.g., equation (dt) reads as $g \approx 2 \times 10^{-51}$ in the natural units of $(c, \hbar, G)$.

Position of an object A will be assumed to be an observable and taken as an element of the set $\mathbb{R}^3$. This is compatible with non-relativistic quantum mechanics but needs modifications in quantum field theory and general relativity. A mathematical extension of the word *position* is loosely used with any manifold $\mathcal{M}$ to denote an element of that manifold, such as "position in phase space $\mathbb{R}^6$ ".

Time (denoted t) is a notoriously difficult concept to define: we will simply take it as a variable. One usually takes it to be real, but its complexification is also useful in quantum mechanics. It can be combined with space to yield the manifold $\mathbb{R}^4$ (Galilean spacetime) or $\mathbb{R}^{3,1}$ (Lorentzian spacetime).

Velocity (acceleration) is an observable defined as the first (second) derivative of position with respect to time.

Force on an object A is a phenomenological idealization of the interactions between A and its environment. We will take the force to be physical as it is measurable through the direct observable "length": in principle, we can always measure the force on an object A if we connect it to another object B via a (not necessarily ideal) spring, put this couple in empty space, and measure the change in the length of the spring due to the existence of the force.

Mass / Inertia of an object A (denoted m / I) is a phenomenological description of the object's resistance to translation / rotation. If we apply a force $F_A(\mathbf{v})$ to an object A (of the velocity $\mathbf{v}$) and measure its acceleration $a_A(\mathbf{v})$, we can define its mass as

$$m_A = \lim_{\mathbf{v} \rightarrow 0} \frac{F_A(\mathbf{v})}{a_A(\mathbf{v})} \quad (\text{du})$$

whenever this limit exists. One can define the components of the inertia tensor analogously.

Gravitational mass of an object A is the measure of the strength of the gravitational interactions A experiences. In Newtonian physics, it is not necessarily same with the *rest mass* (or simply mass as defined above). In Einstein's General Relativity, rest mass and gravitational mass are postulated to be same (this is called *equivalence principle*). In various modern approaches, equivalence principle emerges as output of mathematical requirements (such as diffeomorphism invariance).

Electric charge and electric current (denote q_e and $\bar{I}_e$) are two interrelated phenomena, with electric current being time rate of flow of electric charge. Until 2019, electric current was defined in terms of the observable force and distance, and the electric charge was defined in terms of the electric current. Currently, we define these objects not through the macroscopic observables but through the mathematical modeling via quantum electrodynamics: indeed, we assume *existence of a proton* and *that charge of the proton is universal and time-invariant*, and call that charge "elementary charge e " which is accepted to be a fundamental constant of nature for all practical purposes. We then use it to define the SI units Coloumb & Ampere and all other related quantities.

Electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$ are defined in terms of the physical quantities *force*, *velocity*, and *electric charge* as follows. The electric field $\mathbf{E}$ at position $\mathbf{r}$ is defined as

$$\mathbf{E}(\mathbf{r}) = \lim_{q_e \rightarrow 0} \lim_{\mathbf{v} \rightarrow 0} \frac{\mathbf{F}(\mathbf{r}, \mathbf{v})}{q_e} \quad (\text{dv})$$

where $\mathbf{F}(\mathbf{r}, \mathbf{v})$ is the force on an object of electric charge q_e which is at the position $\mathbf{r}$ and has the velocity $\mathbf{v}$. With this, the magnetic field $\mathbf{B}$ at position $\mathbf{r}$ can be defined as

$$\mathbf{B}(\mathbf{r}) = \lim_{q_e \rightarrow 0} \left[\left(\hat{\mathbf{y}} \cdot \frac{\mathbf{F}(\mathbf{r}, \hat{\mathbf{z}}) - q_e \mathbf{E}(\mathbf{r})}{q_e} \right) \hat{\mathbf{x}} + \left(\hat{\mathbf{z}} \cdot \frac{\mathbf{F}(\mathbf{r}, \hat{\mathbf{x}}) - q_e \mathbf{E}(\mathbf{r})}{q_e} \right) \hat{\mathbf{y}} + \left(\hat{\mathbf{x}} \cdot \frac{\mathbf{F}(\mathbf{r}, \hat{\mathbf{y}}) - q_e \mathbf{E}(\mathbf{r})}{q_e} \right) \hat{\mathbf{z}} \right] \quad (\text{dw})$$

where $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ are unit vectors forming a right-handed Cartesian coordinate system. The definition of $\mathbf{B}$ should be modified if we want to work in a *left-handed* coordinate system because $\mathbf{B}$ is actually not a *true vector* but a *pseudo (axial) vector*.

Lorentz force can be defined as the corollary of the definitions above, i.e. $\lim_{q \rightarrow 0} [\mathbf{F} - q(\mathbf{E} + \mathbf{v} \times \mathbf{B})] = 0$. If we focus on finite q , the self-interaction of the charge enters into the picture, i.e. " $\mathbf{F} - q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = \text{self-interaction terms}$ ". Such extension of the Lorentz equation has been carried out in the literature, and the most advanced one (to my knowledge) goes by the name *Abraham-Lorentz-Dirac-Langevin equation*, which is way beyond the scope of undergraduate physics. Instead, the common approach is to state Lorentz force as

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (\text{dx})$$

where the force is understood to be on a *test particle* whose charge q is infinitesimal.

Maxwell's equations in vacuum form a collection of

empirical observations: in SI units and the vector notation,

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} & (\text{Gauss's law}) \\ \nabla \cdot \mathbf{B} &= 0 & (\text{Gauss's law for magnetism}) \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} & (\text{Faraday's law}) \\ \nabla \times \mathbf{B} &= \mu_0 \left(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) & (\text{Ampère-Maxwell law}) \end{aligned} \quad (\text{dy})$$

where ρ , $\mathbf{J}$, ϵ_0 , and μ_0 are *charge density*, *current density*, *vacuum permittivity*, and *vacuum permeability* respectively.

Vacuum permittivity/permeability (denoted ϵ_0/μ_0) are *defined* as experimental coefficients in SI units, fitted through Maxwell equations where charge/current and $\mathbf{E}/\mathbf{B}$ fields are independently defined. In other unit systems (such as Gaussian units), they may have theoretically fixed values. However, regardless of the unit system, they satisfy $\epsilon_0 \mu_0 c^2 = 1$ where c is the speed of light in vacuum.

Polarization/magnetization of a medium (denoted $\mathbf{P}/\mathbf{M}$) is the reaction of medium to external $\mathbf{E}/\mathbf{B}$ field.

Bound charge and bound current are charge and current induced by polarization and magnetization of medium:

$$\rho_b = -\nabla \cdot \mathbf{P}, \quad \mathbf{J}_b = \nabla \times \mathbf{M} + \frac{\partial \mathbf{P}}{\partial t} \quad (\text{dz})$$

Free charge/current (denoted $\rho_f/\mathbf{J}_f$) is any charge/current that is not a bound charge/current.

Electric displacement field / magnetic flux density (denoted $\mathbf{D}/\mathbf{H}$) are "effective" electric and magnetic field:

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}(\mathbf{E}), \quad \mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M}(\mathbf{H}) \quad (\text{ea})$$

Maxwell's equations in media read as

$$\nabla \cdot \mathbf{D} = \rho_f, \quad \nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \times \mathbf{H} = \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t} \quad (\text{eb})$$

Linear medium is one in which $\mathbf{P}/\mathbf{M}$ are linear functions, hence $\mathbf{D} = \epsilon \mathbf{E}$ and $\mathbf{B} = \mu \mathbf{H}$ with ϵ & μ denoting *permittivity* and *permeability of the medium*.

Flux of a vector field $\mathbf{V}$ (denoted Φ) is the integral of its orthogonal component over a given surface; for instance, electric flux over a surface A would be $\Phi_E = \int_A \mathbf{E} \cdot d\mathbf{S}$.

Energy density of electromagnetic fields is

$$u(\mathbf{r}) = \frac{\epsilon_0}{2} |\mathbf{E}(\mathbf{r})|^2 + \frac{1}{2\mu_0} |\mathbf{B}(\mathbf{r})|^2 \quad (\text{ec})$$

Its integration gives electromagnetic energy in that volume.

Energy flux density of electromagnetic field is

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B} \quad (\text{ed})$$

This vector represents the energy transported by electromagnetic fields per unit area in unit time.

Momentum density of electromagnetic field is

$$\mathbf{g} = \frac{1}{c^2} \mathbf{S} \quad (\text{ee})$$