Mathematical Methods for Scientists

Dietrich Belitz
Department of Physics
University of Oregon

November 2, 2023

Contents

Acknowledgments

These notes are based on a one-quarter course taught various times at the University of Oregon. This course is the first quarter of a year-long sequence, with the following two quarters covering electromagnetism; the selection of the material covered partially reflects this. The notes were typeset by Wenqian Sun and Victor Rosolem based on handwritten notes. The TEX code is based on a template written for a different course by Joshua Frye, with help from Rebecca Tumblin and Brandon Schlomann.

Disclaimer

These notes are a work in progress. If you notice any mistakes, whether it's trivial typos or conceptual problems, please send email to `dbelitz@uoregon.edu`.

Chapter 1

Algebraic Structures

NOTATION

$\in$	is an element of, is in
$\notin$	is not an element of, is not in
$\Rightarrow$	implies
$\wedge$	logical and
$\vee$	logical or
$:=$	is defined to be
$\equiv$	identically equals
$\exists$	there exists
$\exists!$	there exists exactly one
$\forall$	for all
$\square$	end of proof
$\Longleftrightarrow$	if and only if
$\cong$	is isomorphic to

1 Sets and Mappings

1.1 Sets

Consider a collection of well-defined, distinct objects that can be either real or imagined, such as coins, cars, numbers, letters, or pieces of chalk.

Definition 1.
(a) A ***set*** M is defined by any property that each of the objects does or does not possess. If m is an object that has the property, then we say "m is an element of M" or "m is in M" and write $m \in M$. Otherwise, we write $m \notin M$.
(b) The set containing no elements is called ***empty set*** or ***null set*** and denoted by $\varnothing$.

Example 1. All pieces of blue chalk in a classroom form a set M_{bc}.

If a set M has elements $m_1, m_2, \ldots$, then we write $M = \{m_1, m_2, \ldots\}$. If p is the property that determines M, then we write $M = \{m; m \text{ has the property p.}\}$.

Example 2. Some common number sets are

the set of natural numbers denoted by $\mathbb{N} = \{1, 2, 3, \ldots\}$,

the set of integers denoted by $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$,

the set of rationals denoted by $\mathbb{Q} = \{p/q; p, q \in \mathbb{Z} \wedge q \neq 0\}$

the set of real numbers denoted by $\mathbb{R}$,

the set of complex numbers denoted by $\mathbb{C}$.

Remark 1. It is assumed that the reader has an intuitive understanding of these number sets. For a definition of $\mathbb{N}$, see Sec. 1.4 below; for a more recent definition, see, e.g., *Introduction to Mathematical Philosophy* by Bertrand Russell (1993). For a definition of $\mathbb{R}$, see *Algebra* by van der Waerden (1991). These books are listed on the class website.

Remark 2. If the objects themselves are sets, problems may result that we will ignore. See **Problem 1.1.1** (*Russell's Paradox*).

Definition 2. Let A and B be sets.
(a) A is called a ***subset*** of B ($A \subseteq B$) if $a \in A$ implies $a \in B$ ($a \in A \Rightarrow a \in B$).
(b) A and B are ***equal*** ($A = B$) if $A \subseteq B \wedge B \subseteq A$.
(c) A is called a ***proper subset*** of B ($A \subset B$) if $A \subseteq B \wedge A \neq B$.
(d) $\varnothing$ is a subset of any set.

Remark 3. The relation $A \subseteq B$ can be illustrated by a ***Venn diagram***, see Fig. 1.1.1.

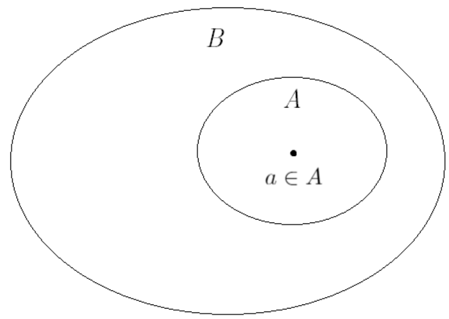


Fig. 1.1.1. $A \subseteq B$.

Remark 4. The relation $\subseteq$ is transitive, i.e., $\forall A, B, C,\ A \subseteq B \wedge B \subseteq C \Rightarrow A \subseteq C$. Fig. 1.1.2 depicts the transitive property.

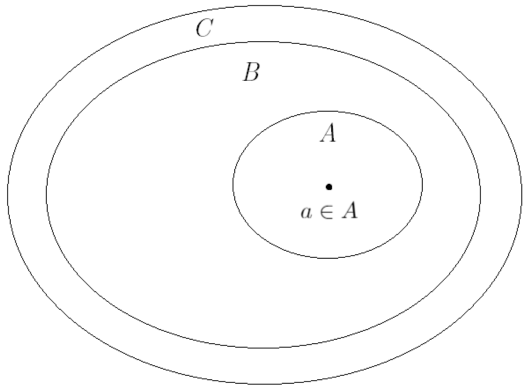
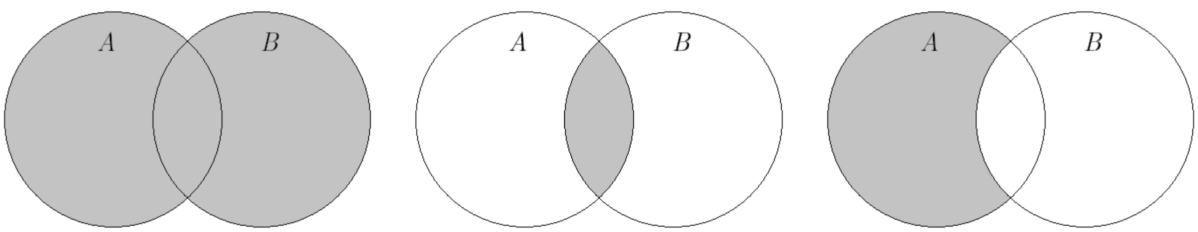


Fig. 1.1.2. The transitivity of $\subseteq$.

Definition 3. Let A and B be sets. We define
(a) the ***union*** of A and B by $A \cup B \coloneqq \{x; x \in A \vee x \in B\}$,
(b) the ***intersection*** of A and B by $A \cap B \coloneqq \{x; x \in A \wedge x \in B\}$ and
(c) the ***difference*** between A and B, or the ***complement*** of B in A, by $A \setminus B \coloneqq \{x; x \in A \wedge x \notin B\}$.

Remark 5. These relations can also be illustrated by Venn diagrams, see Fig. 1.1.3. They have distributive properties, see **Problem 1.1.2** (*Distributive Property of the Union and Intersection Relations*).



(a) $A \cup B$. (b) $A \cap B$. (c) $A \setminus B$.

Fig. 1.1.3. Illustration of (a) the union, (b) the intersection, and (c) the difference of two sets A and B.

Definition 4. Let A and B be sets. If $A \cap B = \varnothing$, then we say that A and B are ***disjoint***.

Definition 5. The ***Cartesian product*** of two sets A and B, denoted by $A \times B$, is the set of all possible ordered pairs with the first component of each pair an element of A, and the second one an element of B. We write $A \times B = \{(a, b); a \in A \wedge b \in B\}$.

Example 3. $\mathbb{R} \times \mathbb{R} \equiv \mathbb{R}^2$ is an algebraic representation of the Cartesian plane.

1.2 Mappings

Definition 1. Let X, Y be sets.

(a) Let φ be a prescription that associates with every $x \in X$ one and only one $y = \varphi(x) \in Y$. Then φ is called a ***mapping*** from X to Y, and we write $\varphi : X \to Y$.

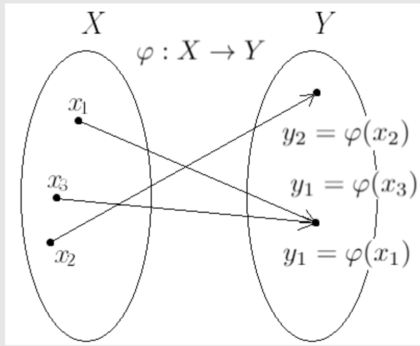
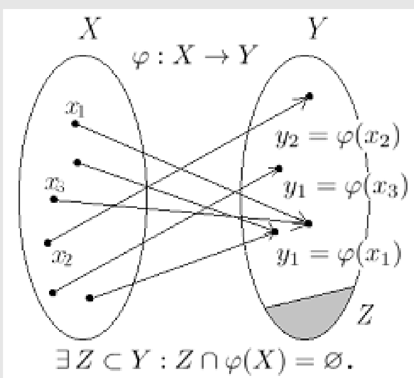


Fig. 1.2.1. A mapping.

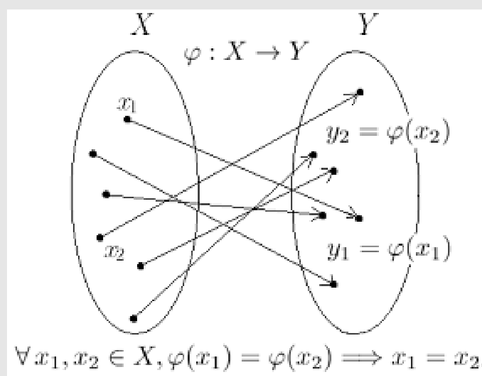
(b) $y = \varphi(x)$ is called the ***image*** of x under φ, and x is called a ***pre-image*** of y. We write $x \stackrel{\varphi}{\to} y$ or $\varphi : x \to y$.

(c) If every $y \in Y$ has at least one pre-image in X, then φ is called a ***surjective*** mapping. We write $Y = \varphi(X)$ and say that φ maps X ***onto*** Y.

(d) If every image $y \in Y$ has one and only one pre-image in X, then φ is called an ***injective*** or ***one-to-one*** mapping.



(a) A mapping that is not surjective.



(b) An injective mapping.

Fig. 1.2.2. Properties of mappings.

(e) A mapping that is both injective and surjective is called a ***bijective*** mapping.

(f) Let X be a set and let φ be a bijective mapping from $\mathbb{N}$ to X. Then X is called a ***countable*** set.

Example 1. $\mathbb{Z}$ and $\mathbb{Q}$ are countable sets. $\mathbb{R}$ is not countable.

Remark 1. No pre-image can have more than one image, and every $x \in X$ must be a pre-image of some $y \in Y$.

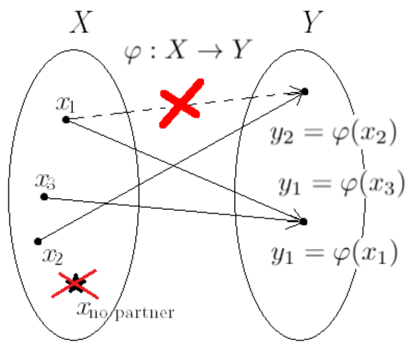


Fig. 1.2.3. A non-mapping.

Remark 2. An image can have multiple pre-images (See x_1 and x_3 in Fig. 1.2.1).

Example 2. Let $X = Y = \mathbb{R}$. $x \overset{\varphi}{\to} \sqrt{x}$ is not a mapping. But, if we choose $X = \{x; x \in \mathbb{R} \wedge x \geqslant 0\}$ and $Y = \mathbb{R}$, then $x \overset{\varphi}{\to} \sqrt{x}$ is a mapping. For more examples, see **Problem 1.1.3** (*Mappings*) and **Problem 1.1.4** (*Parabolic Mapping*).

Remark 3. If $\varphi : X \to Y$ is bijective, then there exists exactly one mapping $\varphi^{-1} : Y \to X$ such that $\varphi : x \to y$ implies that $\varphi^{-1} : y \to x$. φ^{-1} is called the ***inverse*** of φ. (This is plausible, but requires a proof, which we skip for now.)

Definition 2. Let X and Y be two identical sets. Let x be an arbitrary element of X. The mapping $\varphi : x \to x$ is called the ***identity mapping*** of X denoted by I_X or id_X.

Remark 4. It is obvious that id_X is bijective, and $\mathrm{id}_X{}^{-1} = \mathrm{id}_Y = \mathrm{id}_X$.

Remark 5. If X and Y are number sets, then mappings $f : X \to Y$ are called ***functions***, and we write $y = f(x)$. For functions, we sometimes relax the rule that no pre-image can have more than one image; functions violating the rule are called ***multivalued functions***, see Chapter 2.

Definition 3. Let X be a set. Let I be another set called ***index set***. We say that the images x_i of an arbitrary mapping $\varphi : i \in I \to x_i \in X$ are a system of elements of X that is ***labelled*** or ***indexed*** by I.

Remark 6. We often choose $I = \mathbb{N}$. However, this is not necessary; in general, I does not even have to be countable.

Example 3. Counting is an example of indexing objects with $I = \mathbb{N}$.

Example 4. Consider rotations ρ in the Cartesian plane. We can label each ρ with the corresponding angle of rotation α. This uses the uncountable set $I = [0, 2\pi[$ to label rotations: $\varphi : \alpha \in I \to \rho_\alpha$.

Remark 7. We can use I to index sets. This allows us to generalize our previous concepts of union and intersection: for more than two sets, the union of these sets (labelled by I) can be defined by

$$\bigcup_{i \in I} X_i \coloneqq \{x; \exists\, i \in I : x \in X_i\},$$

and the intersection by

$$\bigcap_{i\in I} X_i := \{x; \forall\, i \in I, x \in X_i\}.$$

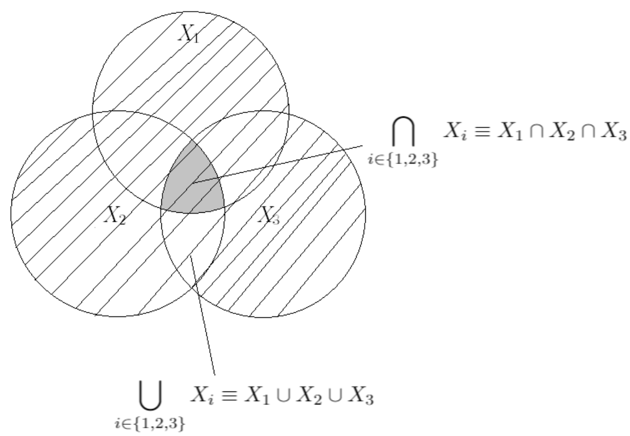


Fig. 1.2.4. Union and intersection of three sets.

Remark 8. We can also generalize the Cartesian product with the help of I, e.g., $\underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ times}} \equiv \mathbb{R}^n := \{(x_1, x_2, \ldots, x_n); \forall\, i \in [1, n] \cap \mathbb{N}, x_i \in \mathbb{R}\}$.

Definition 4. Let X, Y, Z be sets. Let $f : X \to Y$ and $g : Y \to Z$ be mappings. The relation that connects each $x \in X$ to some $g(f(x)) \in Z$ defines another mapping $g \circ f : X \to Z$ called the ***composition*** of f and g. We say "g after f", or "g follows f".

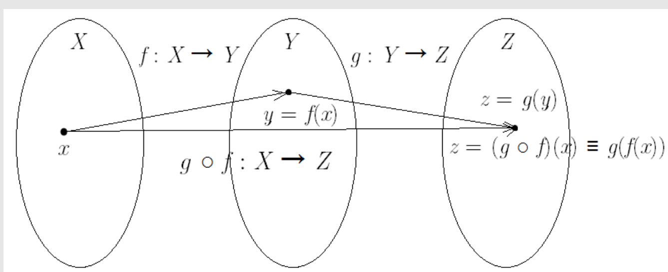


Fig. 1.2.5. The composition of f and g.

Proposition 1. *Let X_1, X_2, X_3, X_4 be sets. Let $f_1 : X_1 \to X_2$, $f_2 : X_2 \to X_3$, $f_3 : X_3 \to X_4$ be mappings. Then $f_3 \circ (f_2 \circ f_1) = (f_3 \circ f_2) \circ f_1 \equiv f_3 \circ f_2 \circ f_1$.*

Proof. Let x be an arbitrary element of X_1. Then $(f_3 \circ (f_2 \circ f_1))(x) = f_3\Big((f_2 \circ f_1)(x)\Big) = f_3\Big(f_2(f_1(x))\Big)$. We also have $((f_3 \circ f_2) \circ f_1)(x) = (f_3 \circ f_2)\Big(f_1(x)\Big) = f_3\Big(f_2(f_1(x))\Big)$. The equality $f_3 \circ (f_2 \circ f_1) = (f_3 \circ f_2) \circ f_1$ is thus established: for all $x \in X_1$, both of these two mappings map x to $f_3\Big(f_2(f_1(x))\Big) \in X_4$. We say that the operation $\circ$ is ***associative*** and write $f_3 \circ f_2 \circ f_1$. □

Remark 9. In general, the operation $\circ$ is not commutative, i.e., $f_2 \circ f_1 \neq f_1 \circ f_2$.

Example 5. Consider two real functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x + 1$ and $g(x) = x^2$, respectively. Then $g \circ f \neq f \circ g$: for an arbitrary $x \in \mathbb{R}$, $(g \circ f)(x) \equiv g(f(x)) = (x+1)^2 \neq x^2 + 1 = f(g(x)) \equiv (f \circ g)(x)$.

1.3 Ordered Sets

Definition 1. Let X be a set. An ***order*** on X is defined as a relation $x \sim y$ between components of ordered pairs $(x, y) \in X \times X$ that possesses the following properties: $\forall\, x, y, z \in X$,

1. $x \sim x$; (reflexivity)
2. $(x \sim y \wedge y \sim x) \Rightarrow x = y$;
3. $(x \sim y \wedge y \sim z) \Rightarrow x \sim z$. (transitivity)

If, in addition,

4. $\forall\,(x, y) \in X \times X, x \sim y \vee y \sim x$,

then we call the order ***linear***.

Example 1. Let $m, n \in \mathbb{N}$. The relation "m divides n" is an order on $\mathbb{N}$. It is not linear since, e.g., 2 does not divide 3 and 3 does not divide 2.

Example 2. The relation "Person 1 is the mother of Person 2" is not an order on the set of all people, since reflexivity is not satisfied.

Example 3. The ordinary "less or equal" relation $\leqslant$ is a linear order on $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$, and $\mathbb{R}$.

Remark 1. $\leqslant$ is often used to denote general orders.

Remark 2. Also of interest are *equivalence relations*, see **Problem 1.1.5** (*Equivalence Relations*).

Definition 2.

(a) Let X be a set and $\leqslant$ an order on X. Let $Y \subseteq X$. Let $\mathrm{b} \in X$ have the property: $\forall\, y \in Y, y \leqslant \mathrm{b}$ $(\mathrm{b} \leqslant y)$. Such a b is called an ***upper*** (a ***lower***) ***bound*** of Y, and we say that Y is ***bounded above*** (***below***) by b.

(b) Let B be the set of upper (lower) bounds of Y. Let $\mathrm{b}_0 \in B$ have the property: $\forall\, \mathrm{b} \in B, \mathrm{b}_0 \leqslant \mathrm{b}$ $(\mathrm{b} \leqslant \mathrm{b}_0)$. We call b_0 the ***least upper bound*** (***greatest lower bound***) or the ***supremum*** (***infimum***) of Y, and we write $\mathrm{b}_0 = \sup Y$ $(\mathrm{b}_0 = \inf Y)$.

Remark 3. The supremum or infimum of Y, if it exists, is not necessarily an element of Y.

Example 4. Consider $Y = [0, 1[\subset \mathbb{R} = X$. We have $\sup Y = 1 \notin Y$, and $\inf Y = 0 \in X$.

1.4 Natural Numbers, and the Principle of Mathematical Induction

Remark 1. An *axiom* is a statement that is postulated to be true within a given logical framework and cannot be proven within that framework. Axioms serve as a foundation for provable statements.

Definition 1. The natural numbers $\mathbb{N}$ can be defined by ***Peano's axioms***:
(1) the number $1 \in \mathbb{N}$;
(2) for all $n \in \mathbb{N}$, there exists a unique successor $n^+ \in \mathbb{N}$;
(3) for all $n \in \mathbb{N}$, $n^+ \neq 1$, i.e., 1 is not the successor of any number;
(4) if $m^+ = n^+$, then $m = n$, i.e., every natural number except for 1 is the unique successor of one and only one number;
(5) Let $M \subseteq \mathbb{N}$. If M satisfies
(a) $1 \in M$ and
(b) $\forall m \in M, \exists!\, m^+ \in M$,
then $M = \mathbb{N}$.

Remark 2. The successor n^+ is usually denoted by $n + 1$. We write $1^+ \equiv 1 + 1 = 2$, $2^+ = 3$, $3^+ = 4$, etc.

Remark 3. Axiom (5) is called the ***principle of mathematical induction***. It can be rephrased as follows. Let a statement S that depends on a natural number n be true for $n = 1$ ('base case'). If one can show that "S is true for $n = k$" implies "S is true for $n = k + 1$" ('induction step'), then S is true for all $n \in \mathbb{N}$.

Proposition 1.
$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \qquad \forall n \in \mathbb{N}\ .$$

Proof. The base case is obviously true:
$$\sum_{i=1}^{1} i \equiv 1 = \frac{1(1+1)}{2}.$$
For the induction step, let us assume that for some $k \in \mathbb{N}$,
$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2}.$$
Now, we add $k + 1$ to both sides of the equality:
$$\sum_{i=1}^{k+1} i = (k+1) + \sum_{i=1}^{k} i = (k+1) + \frac{k(k+1)}{2} = \frac{(k+1)((k+1)+1)}{2}.$$
Hence, the statement is true for all $n \in \mathbb{N}$ by the principle of mathematical induction. □

Remark 4. The principle of mathematical induction still applies if we take the base case to be some natural number $n_0 > 1$. This is true since there exists an obvious bijective mapping from $\{n_0, n_0 + 1, n_0 + 2, \ldots\}$ to $\mathbb{N}$. For an example, see **Problem 1.1.6** (*Bounds for $n!$*). For an example of pitfalls, see **Problem 1.1.7** (*All Ducks are the Same Color*).

1.5 Problems

1.1.1 **Russell's Paradox (B. Russell, 1901)**

a) Consider the set M defined as the set of all sets that do not contain themselves as an element: $M = \{x; x \notin x\}$. Discuss why this is a problematic definition.

b) A less abstract version of Russell's paradox is known as the barber's paradox: Consider a town where all men either shave themselves, or let the barber shave them and don't shave themselves. Now consider the statement

The barber is a man in town who shaves all men who do not shave themselves, and only those.

Discuss why this definition of the barber is problematic (assuming there actually is a barber in town).

hint: Ask "Does the barber shave himself?"

c) Suppose the definition of the barber is modified to read

The barber shaves all men in town who do not shave themselves, and only those.

Discuss what this modification does to the paradox.

(3 points)

1.1.2 **Distributive property of the union and intersection relations**

Show graphically that the relations $\cup$ and $\cap$ defined in ch.1, §1.1, def. 3 obey the following distributive properties: For any three sets A, B, C,

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

(2 points)

1.1.3 **Mappings**

Are the following $f : X \to Y$ true mappings? If so, are they surjective, or injective, or both?

a) $X = Y = \mathbb{Z}$, $\quad f(m) = m^2 + 1$.

b) $X = Y = \mathbb{N}$, $\quad f(n) = n + 1$.

c) $X = \mathbb{Z}$, $Y = \mathbb{R}$, $\quad f(x) = \log x$.

d) $X = Y = \mathbb{R}$, $\quad f(x) = e^x$.

(2 points)

1.1.4 **Parabolic Mapping**

Consider $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = a\,n^2 + b\,n + c$, with $a, b, c \in \mathbb{Z}$.

a) For which triplets (a, b, c) is f surjective?

b) For which (a, b, c) is f injective?

(4 points)

1.1.5 **Equivalence relations**

Consider a relation $\sim$ on a set X as in ch. 1 §1.3 def. 1, but with the properties

i) $x \sim x \quad \forall x \in X$ (reflexivity)

ii) $x \sim y \Rightarrow y \sim x \quad \forall x, y \in X$ (symmetry)

iii) $(x \sim y \wedge y \sim z) \Rightarrow x \sim z$ (transitivity)

Such a relation is called an *equivalence relation.* Which of the following are equivalence relations?

a) n divides m on $\mathbb{N}$.

b) $x \leq y$ on $\mathbb{R}$.

c) g is perpendicular to h on the set of straight lines $\{g, h, \ldots\}$ in the cartesian plane.

d) a equals b modulo n on $\mathbb{Z}$, with $n \in \mathbb{N}$ fixed.

hint: "a equals b modulo n", or $a = b \text{ mod}(n)$, with $a, b \in \mathbb{Z}$, $n \in \mathbb{N}$, is defined to be true if $a - b$ is divisible on $\mathbb{Z}$ by n; i.e., if $(a - b)/n \in \mathbb{Z}$.

(3 points)

1.1.6 **Bounds for $n!$**

Prove by mathematical induction that

$$n^n/3^n < n! < n^n/2^n \quad \forall\, n \geq 6$$

hint: $(1 + 1/n)^n$ is a monotonically increasing function of n that approaches Euler's number e for $n \to \infty$.

(4 points)

1.1.7 **All ducks are the same color**

Find the flaw in the "proof" of the following

proposition: All ducks are the same color.

proof: $n = 1$: There is only one duck, so there is only one color.

$n = m$: The set of ducks is one-to-one correspondent to $\{1, 2, \ldots, m\}$, and we assume that all m ducks are the same color.

$n = m+1$: Now we have $\{1, 2, \ldots, m, m+1\}$. Consider the subsets $\{1, 2, \ldots, m\}$ and $\{2, \ldots, m, m+1\}$. Each of these represent sets of m ducks, which are all the same color by the induction assumption. But this means that ducks #2 through m are all the same color, and ducks #1 and $m + 1$ are the same color as, e.g., duck #2, and hence all ducks are the same color.

remark: This demonstration of the pitfalls of inductive reasoning is due to George Pólya (1888 - 1985), who used horses instead of ducks.

(2 points)

2 Groups

2.1 Definition of a Group

Definition 1. Let $G \neq \varnothing$ be a set. Let $\varphi : G \times G \to G$ be a mapping that assigns to every ordered pair $(a, b) \in G \times G$ an element of G, denoted by $a \vee b$. If $\vee$ possesses the following properties: $\forall\, a, b, c \in G$,

i. $a \vee b \in G$ (closure)

ii. $(a \vee b) \vee c = a \vee (b \vee c) \equiv a \vee b \vee c$ (associativity)

iii. $\exists\, e \in G : e \vee a = a$ (existence of a neutral element)

iv. $\exists\, a^{-1} \in G : a^{-1} \vee a = e$ (existence of an inverse)

then we call G a ***group under the operation*** $\vee$ and write $(G, \vee)$. If, in addition, $\vee$ has the property: $\forall\, a, b \in G$,

v. $a \vee b = b \vee a$ (commutativity)

then we call G an ***abelian group*** and $\vee$ a ***commutative operation***.

Remark 1. "$\vee$" is used here to denote the mapping, i.e., $\varphi\Big((a, b)\Big) \equiv a \vee b$. This should not be confused with the logical operator "or".

Remark 2. For abelian groups, "$\vee$" is often written as "+" and called ***addition***. In this case, e is denoted by 0, and a^{-1} by $-a$. One usually writes $a - a = 0$ instead of $a + (-a) = 0$. With these conventions we call the group ***additive***, or a ***group under addition***.

Example 1.
(1) $(\mathbb{Z}, +)$ with + the ordinary addition is an abelian group whose neutral element is the number 0.
(2) $(\mathbb{R}, +)$ is another abelian group.

Proposition 1. $\mathbb{R} \setminus \{0\}$ *is an abelian group under ordinary multiplication. Its neutral element is the number* 1.

Proof. Check that the group satisfies the five required properties: $\forall\, a, b, c \in \mathbb{R} \setminus \{0\}$,
(i) $ab \in \mathbb{R} \setminus \{0\}$;
(ii) $(ab)c = a(bc)$;
(iii) $1a = a$;
(iv) $\exists\, a^{-1} = \frac{1}{a} \in \mathbb{R} \setminus \{0\}$: $a^{-1}a = \frac{1}{a}a = 1$;
(v) $ab = ba$. □

Remark 3. The notation $a \vee b \equiv a \cdot b \equiv ab$ and $e \equiv 1$ is used more generally, in which case the group is called ***multiplicative***, or a ***group under multiplication***. (NB: This does NOT imply that the group is abelian!)

Proposition 2. *Let* $(G, \vee)$ *be a group. Then*
(a) $a \vee a^{-1} = a^{-1} = e \;\forall a \in G$ *(left inverse = right inverse)*
(b) $a \vee e = e \vee a = a$ *(left neutral element = right neutral element)*
(c) *The neutral element* e *is unique.*

Proof.
(a) From Definition 1 iii, iv it follows that
$a^{-1} \vee a \vee a^{-1} = e \vee a^{-1} = a^{-1}$.
But a^{-1} has an inverse $(a^{-1})^{-1}$. Multiply with that from the left:
$(a^{-1})^{-1} \vee a^{-1} \vee a \vee a^{-1} = (a^{-1})^{-1} \vee a^{-1} = e$
But the lhs equals $e \vee a \vee a^{-1} = a \vee a^{-1}$.
So we have shown that the left inverse equals the right inverse AND that $a = (a^{-1})^{-1}$.
(b) $e \vee a = a \vee a^{-1} \vee a = a \vee e$.
(c) Suppose $\exists\, e_1, e_2 : e_1 \vee a = a = a \vee e_2\ \forall a$. Then
$e_1 \vee e_2 = e_2$ and $e_1 = e_1 \vee e_2$, and hence $e_2 = e_1$. □

Example 2. The set $\{a, e\}$ with an operation $\vee$ defined by $e \vee e = e$, $e \vee a = a \vee e = a$, and $a \vee a = e$ forms an abelian group.

Remark 4. For finite groups, the operation scheme can be represented by a table. For instance, for the group in **Example 2**, we have

$$\begin{array}{c|cc} & a & e \\ \hline a & e & a \\ e & a & e \end{array}.$$

For a more elaborate group table, see **Problem 1.2.1** (*Pauli Group*).

2.2 Rules of Operation

Proposition 1. *Let $(G, \vee)$ be a group. For all $a, b \in G$, $(a \vee b)^{-1} = b^{-1} \vee a^{-1}$.*

Proof. We know that $a \vee b \in G$. To complete the proof, we simply write $(a \vee b)^{-1} \vee (a \vee b) = e = b^{-1} \vee b = b^{-1} \vee (e \vee b) = b^{-1} \vee ((a^{-1} \vee a) \vee b)) = (b^{-1} \vee a^{-1}) \vee (a \vee b)$. □

Definition 1.
(a) Let $(G, \vee)$ be a multiplicative group. We write the element that is composite of $n \in \mathbb{N}$ elements in G as

$$a_1 \vee a_2 \vee \cdots \vee a_{n-1} \vee a_n \equiv a_1 a_2 \ldots a_{n-1} a_n =: \prod_{\alpha=1}^{n} a_\alpha$$

and define recursively

$$\prod_{\alpha=1}^{n+1} a_\alpha := \left(\prod_{\alpha=1}^{n} a_\alpha \right) a_{n+1}.$$

We call the element the ***product of factors*** $a_1, a_2, \ldots, a_{n-1}, a_n$.
(b) The product of n identical factors

$$\prod_{\alpha=1}^{n} a =: a^n$$

is called the ***n-th power*** of a.

Proposition 2. *Let $(G, \vee)$ be a multiplicative group. We have*

$$\left(\prod_{\alpha=1}^{m} a_\alpha\right)\left(\prod_{\beta=1}^{n} a_{m+\beta}\right) = \prod_{\rho=1}^{m+n} a_\rho.$$

Proof. We are going to complete the proof by applying mathematical induction. First, we will check that for $n = 1$, the statement is true. It is obvious that

$$\left(\prod_{\alpha=1}^{m} a_\alpha\right)\left(\prod_{\beta=1}^{1} a_{m+\beta}\right) \equiv \left(\prod_{\alpha=1}^{m} a_\alpha\right) a_{m+1} = \prod_{\rho=1}^{m+1} a_\rho.$$

Then, supposing that the statement holds for some $k \in \mathbb{N}$, we want to show that it is still valid for $n = k+1$. For $n = k$, we have

$$\left(\prod_{\alpha=1}^{m} a_\alpha\right)\left(\prod_{\beta=1}^{k} a_{m+\beta}\right) = \prod_{\rho=1}^{m+k} a_\rho.$$

Now, we multiply both sides of the equation by a_{m+k+1}. The left-hand side of the equation becomes

$$\left(\left(\prod_{\alpha=1}^{m} a_\alpha\right)\left(\prod_{\beta=1}^{k} a_{m+\beta}\right)\right) a_{m+k+1} = \left(\prod_{\alpha=1}^{m} a_\alpha\right)\left(\left(\prod_{\beta=1}^{k} a_{m+\beta}\right) a_{m+k+1}\right) = \left(\prod_{\alpha=1}^{m} a_\alpha\right)\left(\prod_{\beta=1}^{k+1} a_{m+\beta}\right).$$

The right-hand side of the equation becomes

$$\left(\prod_{\rho=1}^{m+k} a_\rho\right) a_{m+k+1} = \prod_{\rho=1}^{m+k+1} a_\rho.$$

Thus, we have shown that the statement is true for $n = k+1$:

$$\left(\prod_{\alpha=1}^{m} a_\alpha\right)\left(\prod_{\beta=1}^{k+1} a_{m+\beta}\right) = \prod_{\rho=1}^{m+k+1} a_\rho.$$

Hence, the statement is true for all $n \in \mathbb{N}$ by the principle of mathematical induction. ☐

Corollary 1. *Let $(G, \vee)$ be a multiplicative group and $a \in G$ be an arbitrary element in the group. We have*

(a) $a^m a^n = a^{m+n}$;

(b) $(a^m)^n = a^{mn}$.

Proof. See **Problem 1.2.2** (*Products*) ☐

Definition 2. The ***zeroth power*** of a is defined by $a^0 := e$, and the ***negative powers*** of a by $a^{-n} := (a^{-1})^n$.

Remark 1. The latter definition complies with **Corollary 1**, (b).

Remark 2. For additive groups, we write

$$a_1 \vee a_2 \vee \cdots \vee a_{n-1} \vee a_n \equiv a_1 + a_2 + \cdots + a_{n-1} + a_n =: \sum_{\alpha=1}^{n} a_\alpha$$

and name the composite element the ***sum*** of the a_α's. A sum of identical elements is a ***multiple*** of that element:

$$\sum_{\alpha=1}^{n} a =: na.$$

Proposition 2 and its corollaries still hold with $\prod$ replaced by $\sum$:

$$\left(\sum_{\alpha=1}^{m} a_\alpha\right) + \left(\sum_{\beta=1}^{k} a_{m+\beta}\right) = \sum_{\rho=1}^{m+k} a_\rho;$$

$$ma + na = (m+n)a,$$

and

$$mna = nma.$$

2.3 Permutations

Definition 1. Let M be a finite set and $P : M \to M$ be a bijective mapping. We call P a ***permutation*** of M.

Remark 1. If M is finite with $n \in \mathbb{N}$ elements, then M and the set $\{1, 2, \ldots, n-1, n\} \equiv \{i\}_{i=1}^{n}$ share the same cardinality. We are able to characterize every permutation P of M with its action on $\{i\}_{i=1}^{n}$:

$$E = \begin{pmatrix} 1, 2, 3, \ldots, n \\ 1, 2, 3, \ldots, n \end{pmatrix}, \quad P_1 = \begin{pmatrix} 1, 2, 3, \ldots, n \\ 2, 1, 3, \ldots, n \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1, 2, 3, \ldots, n \\ 3, 2, 1, \ldots, n \end{pmatrix}, \quad \text{etc.}$$

Definition 2. If it takes an even number of transpositions (i.e., pairwise exchanges of elements) to convert a permutation P into E, then we say that P is an ***even permutation*** and write $\operatorname{sgn} P = 1$. Otherwise, if it takes an odd number of transpositions, then we say that P is an ***odd permutation*** and write $\operatorname{sgn} P = -1$.

Remark 2. The decomposition of a permutation into transpositions is not unique, but the such defined sign is. Proof: Math books.

Example 1. For permutations listed in *Remark* 1, we have $\operatorname{sgn} E = 1$, $\operatorname{sgn} P_1 = -1$, and $\operatorname{sgn} P_2 = -1$. The permutation

$$\begin{pmatrix} 1, 2, 3, \ldots, n \\ 3, 1, 2, \ldots, n \end{pmatrix}$$

is even.

Proposition 1. *The set of permutations of a finite set M with $n \in \mathbb{N}$ elements forms a group under composition called the* **symmetric group** S_n.

Proof. Ascertain that the set of all the permutations satisfies the four group axioms: $\forall P_1, P_2 \in S_n$,
(i) $(P_1 : M \to M) \wedge (P_2 : M \to M) \Longrightarrow P_1 \circ P_2 : M \to M$;
(ii) associative laws are satisfied because of **Proposition 1** in **1.2**;
(iii)

$$E = \begin{pmatrix} 1, 2, 3, \ldots, n \\ 1, 2, 3, \ldots, n \end{pmatrix}$$

serves as the neutral element;
(iv) existence of inverses is due to the fact that bijective mappings have inverses. □

Remark 3. In general, S_n is not abelian. See **Problem 1.2.3** (*The group S_3*).

2.4 Subgroups

Definition 1. Let $(G, \vee)$ be a group and $H \neq \varnothing \subset G$. If H is also a group under $\vee$, we call it a ***subgroup*** of G.

Example 1. Let e be the neutral element of a group G. $\{e\}$ is a subgroup of G.

Theorem 1. *H is a subgroup of $(G, \vee)$ if and only if for all $a, b \in H$, $a \vee b^{-1} \in H$.*

Proof. First, it is trivial that H is a subgroup implies that for all $a, b \in H$, $a \vee b^{-1} \in H$. It is more instructive to complete the proof by contrapositive. That is, we would like to show the statement that there exists some $a, b \in H$ such that $a \vee b^{-1} \notin H$ implies that H is not a subgroup. Suppose that such a and b exist. We know that if H is a subgroup, then $a \vee b^{-1} \in H$. It directly follows that H cannot be a subgroup. Proof by contrapositive can be very useful in some cases (although the reader might find that in the current example, it seems somewhat unnecessary).

Second, we want to show that for all $a, b \in H$, $a \vee b^{-1} \in H$ implies that H is a subgroup. Suppose that for all $a, b \in H$, $a \vee b^{-1} \in H$. We need to check that H satisfies the four group axioms. Notice that if we choose two identical elements $x = y$, then $e = x \vee x^{-1} = x \vee y^{-1} \in H$. We have thus established that the neutral element is contained in H. Now, if we choose the neutral element as one of the two elements in our assumption (let $a = e$), we will have $e \vee b^{-1} = b^{-1} \in H$, for all $b \in H$; that is, existence of inverses is satisfied. What is more, combining existence of inverses and the assumption, we have $\forall b \in H, \exists!\, b^{-1} \in H : \forall a \in H, a \vee b = a \vee (b^{-1})^{-1} \in H$; in other words, we have $\forall a, b \in H, a \vee b \in H$, the closure. At last but not least, the fact that $(G, \vee)$ is a group ensures that the operation $\vee$ is associative. □

Example 2. Let us consider the following two elements of S_3,

$$E = \begin{pmatrix} 1,2,3 \\ 1,2,3 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} 1,2,3 \\ 1,3,2 \end{pmatrix}.$$

We want to apply **Theorem 1** to check whether or not $g = \{E, P\}$ is a subgroup of S_3 (under composition $\circ$). First, notice that $E^{-1} = E$, and $P^{-1} = P$, since $E \circ E = E$, and $P \circ P = E$. It is straightforward to check the following:

$$E \circ E^{-1} = E \in g;$$

$$E \circ P^{-1} = E \circ P = P \in g;$$

$$P \circ E^{-1} = P \circ E = P \in g;$$

$$P \circ P^{-1} = E \in g.$$

Hence, $g = \{E, P\}$ is a subgroup of S_3 by **Theorem 1**.

2.5 Isomorphisms and Automorphisms

Definition 1.
(a) Let $(G, \vee)$ and $(H, *)$ be groups. Let $\varphi : G \to H$ be a bijective mapping such that for all $a, b \in G$, $\varphi(a \vee b) = \varphi(a) * \varphi(b)$. Such a φ is called an ***isomorphism*** between G and H. We say that G is isomorphic to H and write $G \cong H$.
(b) Furthermore, if $G = H$, then we call φ an ***automorphism*** on G. That is, an isomorphism between a group and itself is an automorphism on the group.

Remark 1. We refer to $\varphi(a \vee b) = \varphi(a) * \varphi(b)$ by saying that "φ respects the operation".

Example 1. Consider a set G of real 2×2 g_α matrices defined by

$$G = \left\{ g_\alpha \equiv \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} ; \alpha \in [0, 2\pi[\right\},$$

and a set H of complex numbers h_β defined by

$$H = \{ \mathrm{h}_\beta \equiv e^{i\beta} ; \beta \in [0, 2\pi] \} .$$

It is easy to show that G is a group under matrix multiplication (denoted by $\cdot$), and H is a group under multiplication of complex numbers (denoted by $*$). Let us define a mapping $\varphi : G \to H$ by the relation $\varphi(g_\alpha) \coloneqq h_\alpha$. It is obvious that φ is bijective. Now we check whether φ is an isomorphism between G and H. First, notice that for all $g_\alpha, g_\beta \in G$,

$$\begin{aligned} g_\alpha \cdot g_\beta &= \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \cdot \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \\ &= \begin{pmatrix} \cos\alpha\cos\beta - \sin\alpha\sin\beta & \cos\alpha\sin\beta + \sin\alpha\cos\beta \\ -\sin\alpha\cos\beta - \cos\alpha\sin\beta & -\sin\alpha\sin\beta + \cos\alpha\cos\beta \end{pmatrix} \\ &= \begin{pmatrix} \cos(\alpha+\beta) & \sin(\alpha+\beta) \\ -\sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix} = g_{\alpha+\beta}. \end{aligned}$$

> Accordingly, we have $\varphi(g_\alpha \cdot g_\beta) = \varphi(g_{\alpha+\beta}) = h_{\alpha+\beta} = e^{i(\alpha+\beta)} = e^{i\alpha} * e^{i\beta} = h_\alpha * h_\beta = \varphi(g_\alpha) * \varphi(g_\beta)$. Hence, we have shown that $G \cong H$.

Remark 2. G is a representation of the group SO (2) (SO stands for "Special Orthogonal".) H is a reprentation of the group U (1) (U stands for "Unitary".)

Remark 3. For an example of an automorphism, and some properties of abelian groups, see **Problem 1.2.4** (*Abelian Groups*).

2.6 Problems

1.2.1. **Pauli group**

The Pauli matrices are complex 2×2 matrices defined as

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad ,$$

Now consider the set P_1 that consists of the Pauli matrices and their products with the factors -1 and $\pm i$:

$$P_1 = \{\pm\sigma_0, \pm i\sigma_0, \pm\sigma_1, \pm i\sigma_1, \pm\sigma_2, \pm i\sigma_2, \pm\sigma_3, \pm i\sigma_3\}$$

Show that this set of 16 elements forms a (nonabelian) group under matrix multiplication called the Pauli group. It plays an important role in quantum information theory.

(3 points)

1.2.2. **Products**

Prove the corollary to proposition 2 of ch.1 §2.2: If a is an element of a multiplicative group, and $n, m \in \mathbb{N}$, then

a) $a^n\, a^m = a^{n+m}$

b) $(a^n)^m = a^{nm}$

(2 points)

1.2.3 **The group S_3**

a) Compile the group table for the symmetric group S_3. Is S_3 abelian?

b) Find all subgroups of S_3. Which of these are abelian?

(6 points)

1.2.4. **A Property of Abelian Groups**

Let $(G, \vee)$ be a group. Let $a \in G$ be a fixed element, and define a mapping $\varphi : G \to G$ by $\varphi(x) = a \vee x \vee a^{-1}\ \forall x \in G$.

a) Show that φ defines an automorphism on G, called an *inner automorphism*.

b) Show that abelian groups have no inner automorphisms except for the identity mapping $\varphi(x) = x$.

(4 points)

3 Fields

3.1 Bilinear Mappings

Definition 1. Let A, B, C be additive groups with neutral elements $0_A, 0_B, 0_C$, respectively. Let $\varphi : A \times B \to C$ be a mapping defined by the relation $\varphi((a,b)) \equiv \varphi(a,b) \equiv a \cdot b \in C$. If φ satisfies distributive laws: $\forall\, a_1, a_2, a_3 \in A \wedge b_1, b_2, b_3 \in B$,

i. $(a_1 + a_2) \cdot b_1 = a_1 \cdot b_1 + a_2 \cdot b_1$;

ii. $a_1 \cdot (b_1 + b_2) = a_1 \cdot b_1 + a_1 \cdot b_2$,

then we call φ a ***bilinear mapping***.

Remark 1. In **i**, the $+$ on the left-hand side (LHS) of the equality is the addition on A. In **ii**, the $+$ on the LHS is the addition on B. In both **i** and **ii**, the $+$ on the right-hand side (RHS) is the addition on C.

Remark 2. Usually called multiplication, $\cdot$ is referred to as an exterior operation, since it connects elements from two different groups. On the other hand, $+$'s are interior operations because of the closure.

Proposition 1. *Consider A, B, C and φ in* **Definition 1***. The following statements are true: $\forall\, a \in A \wedge b \in B$,*

(1) $0_A \cdot b = a \cdot 0_B = 0_C$;
(2) $-a \cdot b = a \cdot (-b) = -(a \cdot b)$;
(3) $-a \cdot (-b) = a \cdot b$.

Proof. For this proof, we will just write symbolic expressions.

(1) $0_A = 0_A + 0_A \Longrightarrow 0_A \cdot b = (0_A + 0_A) \cdot b = 0_A \cdot b + 0_A \cdot b$
$0_C = 0_A \cdot b + (-0_A \cdot b) \equiv 0_A \cdot b - 0_A \cdot b$
$\therefore\ 0_C = 0_A \cdot b + 0_A \cdot b - 0_A \cdot b = 0_A \cdot b$

$0_B = 0_B + 0_B \Longrightarrow a \cdot 0_B = a \cdot (0_B + 0_B) = a \cdot 0_B + a \cdot 0_B$
$0_C = a \cdot 0_B - a \cdot 0_B$
$\therefore\ 0_C = a \cdot 0_B + a \cdot 0_B - a \cdot 0_B = a \cdot 0_B$

(2) $0_C = 0_A \cdot b = (a + (-a)) \cdot b = a \cdot b + (-a) \cdot b \Longrightarrow -a \cdot b = -(a \cdot b)$
$0_C = a \cdot 0_B = a \cdot (b + (-b)) = a \cdot b + a \cdot (-b) \Longrightarrow a \cdot (-b) = -(a \cdot b)$
$0_C = 0_C \Longrightarrow -a \cdot b = a \cdot (-b) = -(a \cdot b)$

(3) $-a \cdot b = a \cdot (-b) \Longrightarrow -a \cdot (-b) = a \cdot (-(-b)) = a \cdot b$ □

3.2 Fields

Definition 1. Let $(K, +)$ be an additive group with neutral element 0. Let $\cdot : K \times K \to K$ be an associative bilinear multiplication. If $K \setminus \{0\}$ is a group under $\cdot$, then we call K a ***field***.

Example 1. Under ordinary addition and multiplication, $\mathbb{R}$ is a commutative field. So is $\mathbb{Q}$, see Problem 1.3.1 (Fields). $\mathbb{Z}$ is not, since not every element has an inverse.

3.3 The Field of Complex Numbers

Theorem 1. *We can construct a commutative field* $\mathbb{C}$*, called the* **field of complex numbers***, with the following properties:*
(1) $\mathbb{R} \subset \mathbb{C}$*;*
(2) $\exists!\, i \in \mathbb{C} : i^2 = -1$*;*
(3) $\mathbb{C} = \{z = z_1 + iz_2; z_1, z_2 \in \mathbb{R}\}$*, i.e., every element* $z \in \mathbb{C}$ *can be uniquely written as* $z = z_1 + iz_2 \equiv z' + iz''$ *for some* $z_1, z_2 \in \mathbb{R}$ *(*$z', z'' \in \mathbb{R}$*).*

Remark 1. z_1 (z') and z_2 (z'') are called the ***real*** and ***imaginary*** parts of a complex number z, respectively. Note that they are both real numbers. We call $z' - iz'' =: z^* \equiv \bar{z}$ the ***complex conjugate*** of $z = z' + iz''$.

Proof. Let us consider the Cartesian product $\mathbb{R}^2 \equiv \mathbb{R} \times \mathbb{R}$. We would like to first establish that $\mathbb{R}^2$ is a field under certain addition and multiplication; thereafter, to complete the proof, we simply show that $\mathbb{C} \cong \mathbb{R}^2$. Let $a = (a_1, a_2)$, $b = (b_1, b_2)$ and $c = (c_1, c_2)$ be elements of $\mathbb{R}^2$.

First, let us turn $\mathbb{R}^2$ into an additive group by defining a proper addition: $\forall\, a, b \in \mathbb{R}^2$, $a + b := (a_1 + b_1, a_2 + b_2)$. It is easy to check that $(\mathbb{R}^2, +)$ is a group with neutral element $(0, 0)$. Second, we need to define a proper multiplication on $\mathbb{R}^2$: $\forall\, a, b \in \mathbb{R}^2$, $a \cdot b \equiv ab := (a_1 b_1 - a_2 b_2, a_1 b_2 + a_2 b_1)$. Notice that the multiplication is both commutative and distributive: $\forall\, a, b, c \in \mathbb{R}^2$,
$ab = (a_1 b_1 - a_2 b_2, a_1 b_2 + a_2 b_1) = (b_1 a_1 - b_2 a_2, b_2 a_1 + b_1 a_2) = ba$;

$$\begin{aligned}(b + c)a = a(b + c) &= a(b_1 + c_1, b_2 + c_2)\\ &= (a_1(b_1 + c_1) - a_2(b_2 + c_2), a_1(b_2 + c_2) + a_2(b_1 + c_1))\\ &= (a_1 b_1 - a_2 b_2, a_1 b_2 + a_2 b_1) + (a_1 c_1 - a_2 c_2, a_1 c_2 + a_2 c_1)\\ &= ab + ac.\end{aligned}$$

Similarly, one can also check that associative laws are satisfied: $\forall\, a, b, c \in \mathbb{R}^2$,

$$\begin{aligned}a(bc) &= a(b_1 c_1 - b_2 c_2, b_1 c_2 + b_2 c_1)\\ &= (a_1(b_1 c_1 - b_2 c_2) - a_2(b_1 c_2 + b_2 c_1), a_1(b_1 c_2 + b_2 c_1) + a_2(b_1 c_1 - b_2 c_2))\\ &= ((a_1 b_1 - a_2 b_2)c_1 - (a_1 b_2 + a_2 b_1)c_2, (a_1 b_1 - a_2 b_2)c_2 + (a_1 b_2 + a_2 b_1)c_1)\\ &= (ab)c.\end{aligned}$$

We have thus shown that $\cdot$ is an associative bilinear multiplication. Now, we want to show that $(\mathbb{R}^2 \setminus \{(0, 0)\}, \cdot)$ is a group. The fact that $\mathbb{R}$ under ordinary addition and multiplication is a field ensures that the closure is satisfied. The multiplicative identity is $(1, 0)$, because $\forall\, a \in \mathbb{R}^2$, $a \cdot (1, 0) = (1, 0) \cdot a = (a_1, a_2) = a$. We have already proven that $\cdot$ is associative. To show that existence of (multiplicative) inverses holds, let $a \neq (0, 0)$. It follows that ${a_1}^2 + {a_2}^2 \neq 0$. Notice that $\forall\, a \in \mathbb{R}^2$,

$$(a_1, a_2) \cdot \left(\frac{a_1}{{a_1}^2 + {a_2}^2}, -\frac{a_2}{{a_1}^2 + {a_2}^2} \right) = \left(\frac{{a_1}^2 + {a_2}^2}{{a_1}^2 + {a_2}^2}, \frac{-a_1 a_2 + a_2 a_1}{{a_1}^2 + {a_2}^2} \right) = (1, 0).$$

This implies that

$$a^{-1} = \left(\frac{a_1}{{a_1}^2 + {a_2}^2}, -\frac{a_2}{{a_1}^2 + {a_2}^2} \right).$$

Hence, we have shown that $\mathbb{R}^2$ is a field under designated addition and multiplication.

Notice the curious fact that $(0, 1)^2 = (0, 1) \cdot (0, 1) = (-1, 0)$ is very similar to $i^2 = -1$. Now, we can define $\mathbb{C}$ by means of an isomorphism $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ given by the relations: $\forall\, z_1, z_2 \in \mathbb{R}$,
$\varphi((0, 1)) = i \in \mathbb{C}$;
$\varphi((z_1, z_2)) = z_1 + iz_2 \in \mathbb{C}$. $\square$

Remark 2. The isomorphism can be graphically represented by the complex plane.

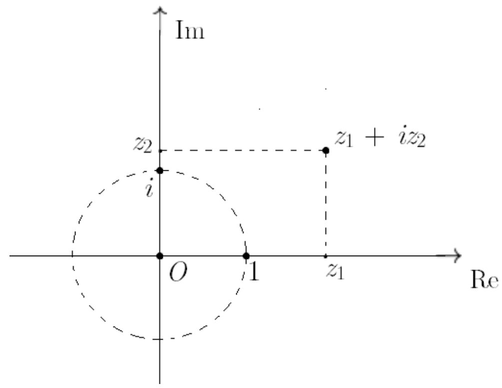


Fig. 3.3.1. The complex plane.

Proposition 1. *The set of complex numbers* $\{z = e^{i\alpha}; \alpha \in [0, 2\pi]\}$ *forms a circle centered at the origin* $0 + i0$ *with radius* 1 *in the complex plane.* **Euler's formula** *reads*

$$e^{i\alpha} = \cos\alpha + i\sin\alpha.$$

Proof. Recall the Maclaurin series of e^x: for $|x| < \infty$,

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

It directly follows that

$$\begin{aligned} e^{i\alpha} &= \sum_{n=0}^{\infty} \frac{(i\alpha)^n}{n!} \\ &= \sum_{m=0}^{\infty} \frac{(i\alpha)^{2m}}{(2m)!} + \sum_{n=0}^{\infty} \frac{(i\alpha)^{2n+1}}{(2n+1)!} \\ &= \sum_{m=0}^{\infty} (-1)^m \frac{\alpha^{2m}}{(2m)!} + i\sum_{n=0}^{\infty} (-1)^n \frac{\alpha^{2n+1}}{(2n+1)!} \\ &= \cos\alpha + i\sin\alpha. \end{aligned}$$

We also know that for each $\alpha \in [0, 2\pi]$, $e^{i\alpha} = z_1 + iz_2$ with some $z_1, z_2 \in \mathbb{R}$. Accordingly, we have ${z_1}^2 + {z_2}^2 = \cos^2\alpha + \sin^2\alpha = 1$, which describes a circle centered at the origin with radius 1 in the complex plane. $\square$

Corollary 1. *Let* $z \in \mathbb{C}$. *There exist real numbers* $r \in [0, \infty)$ *and* $\phi \in [-\pi, \pi)$ *such that* $z = re^{i\phi}$.

Proof. From **Proposition 1** it directly follows that $z = z_1 + iz_2 = r\cos\phi + ir\sin\phi = re^{i\phi}$ with $r = \sqrt{{z_1}^2 + {z_2}^2}$ and $\phi = \arctan(z_2/z_1)$. $\square$

Remark 3. r is called the ***modulus*** or ***absolute value*** of z and denoted by $r = |z|$. ϕ is called the ***argument*** of z; one writes $\phi = \arg z$.

Remark 4. $\phi \in [-\pi, \pi)$ is merely a particular choice. In general, ϕ can be defined on any interval of length 2π.

Remark 5. Let $z = re^{i\phi} \in \mathbb{C}$ for some $r \in [0, \infty)$ and $\phi \in \mathbb{R} \bmod 2\pi$. Notice that $\forall n \in \mathbb{Z}, e^{i2n\pi} = 1 \implies z = re^{i\phi} = re^{i(\phi+2n\pi)}$. That is, a complex number has multiple arguments.

Definition 1. Let $z = re^{i\phi} \in \mathbb{C}$ for some $r \in [0, \infty)$ and $\phi \in \mathbb{R} \bmod 2\pi$. Real powers of z are defined by $z^x := r^x e^{ix\phi}$ for all $x \in \mathbb{R}$.

Remark 6. The definition is consistent with **Corollary 1**, (b) in **2.2**, since $z^x = (re^{i\phi})^x = r^x e^{ix\phi}$. Note the difference that the corollary only holds for $n \in \mathbb{N}$.

Remark 7. For $x \notin \mathbb{N}$, z^x is not unique. In particular, when $x = \frac{1}{n}$ $(n \in \mathbb{N})$, z^x has n different values called ***n-th roots*** of z.

Example 1. Let us compute second roots of i. Let us first write $i = e^{i\frac{\pi}{2}} = e^{i\left(\frac{\pi}{2}+2\pi\right)}$. The second roots are $(i^{\frac{1}{2}})_0 = e^{i\frac{\pi}{4}}$ and $(i^{\frac{1}{2}})_1 = e^{i\frac{1}{2}\left(\frac{\pi}{2}+2\pi\right)} = e^{i\frac{5\pi}{4}}$.

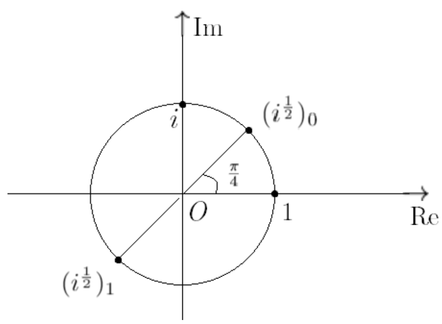


Fig. 3.3.2. The second roots of i in the complex plane.

3.4 Problems

1.3.1. **Fields**

a) Show that the set of rational numbers $\mathbb{Q}$ forms a commutative field under the ordinary addition and multiplication of numbers.

b) Consider a set F with two elements, $F = \{\theta, e\}$. On F, define an operation “plus" $(+)$, about which we assume nothing but the defining properties

$$\theta + \theta = \theta \quad , \quad \theta + e = e + \theta = e \quad , \quad e + e = \theta$$

Further, define a second operation “times" $(\cdot)$, about which we assume nothing but the defining properties

$$\theta \cdot \theta = e \cdot \theta = \theta \cdot e = \theta \quad , \quad e \cdot e = e$$

Show that with these definitions (and **no** additional assumptions), F is a field.

(7 points)

4 Vector Spaces and Tensor Spaces

4.1 Vector Spaces

Definition 1. Let $(V, +)$ be an additive group. Let K be a field with multiplicative neutral element 1_K. We define an ***exterior multiplication*** $\varphi : K \times V \to V$ that possesses the following properties:
(i) bilinearity,
(ii) associativity in the sense that $\forall\, \lambda, \mu \in K, x \in V, (\lambda\mu)x = \lambda(\mu x)$ and
(iii) $1_K x = x\ \forall\, x \in V$.
We then call V a ***vector space*** or ***linear space*** over K, or a ***K-vector space***.

Remark 1. For the sake of simplicity, we assume that K is commutative, i.e., $\forall\, \lambda, \mu \in K, \lambda\mu = \mu\lambda$.

Remark 2. Elements of V are called ***vectors***, and elements of K ***scalars***.

Example 1. Four common $\mathbb{R}$-vector spaces are shown below.

Table 4.1.1. Four common $\mathbb{R}$-vector spaces.

No.	V	The Addition on V	K	Operations on K	The Exterior Multiplication
#1	$\mathbb{R}$	the ordinary $+$	$\mathbb{R}$	the ordinary $+$ and $\cdot$	the ordinary $\cdot$
#2	$\mathbb{C}$	$\forall\, z_1, z_2 \in \mathbb{C}$, $z_1 + z_2 := ({z_1}' + {z_2}') + i({z_1}'' + {z_2}'')$.	$\mathbb{R}$	the ordinary $+$ and $\cdot$	$\forall\, \lambda \in \mathbb{R} \wedge z \in \mathbb{C}$, $\lambda \cdot z \equiv \lambda z := \lambda z' + i\lambda z''$.
#3	$\mathbb{R}^2$	the $+$ defined in **Theorem 1** of **3.3**	$\mathbb{R}$	the ordinary $+$ and $\cdot$	$\forall\, \lambda \in \mathbb{R} \wedge (x, y) \in \mathbb{R}^2$, $\lambda(x, y) := (\lambda x, \lambda y)$.
#4	$\mathbb{R}^n$	a componentwise $+$ similar to the one on $\mathbb{R}^2$	$\mathbb{R}$	the ordinary $+$ and $\cdot$	$\forall\, \lambda \in \mathbb{R} \wedge \boldsymbol{x} \equiv (x_1, \ldots, x_n) \in \mathbb{R}^n$, $\lambda\boldsymbol{x} := (\lambda x_1, \ldots, \lambda x_n)$.

More generally, given an arbitrary field K, we can make K^n a K-vector space by the following definitions. We first define an addition on K^n to turn it into an additive group: $\forall\, \boldsymbol{k} \equiv (k_1, \ldots, k_n), \boldsymbol{l} \equiv (l_1, \ldots, l_n) \in K^n, \boldsymbol{k} + \boldsymbol{l} := (k_1 + l_1, \ldots, k_n + l_n)$. It is easy to check that $(K^n, +)$ is a group with neutral element $\underbrace{(0_K, \ldots, 0_K)}_{n\ 0_K\text{'s}}$, where 0_K is the additive identity in K. K^n will be further promoted to a K-vector space if we define the exterior multiplication by $\forall\, k \in K \wedge \boldsymbol{l} \in K^n, k\boldsymbol{l} := (kl_1, \ldots, kl_n)$.

Remark 3. For another example, see Problem 1.4.1 (Function space).

4.2 Basis Sets

Definition 1. Let V be a K-vector space. If there exist a finite number of vectors $p_1, p_2, \ldots, p_n \in V$ such that $\forall\, x \in V, \exists\, \lambda_1, \lambda_2, \ldots, \lambda_n \in K : x = \sum_{i=1}^n \lambda_i p_i$, then we say that the set $\{p_i\}_{i=1}^n$ ***spans*** V, and we call V a ***finite-dimensional*** vector space.

Example 1. Let us consider $\mathbb{R}^2$ as an $\mathbb{R}$-vector space. The set $\{(1, 0), (0, 2)\}$ spans $\mathbb{R}^2$; so does $\{(1, 0), (0, 1), (1, 1)\}$.

Definition 2. If any of the n vectors $p_1, p_2, \ldots, p_n$ can be expressed as a linear combination of the remaining $n-1$ vectors, the we call the set $\{p_i\}_{i=1}^n$ ***linearly dependent***. Otherwise, we say the n vectors are ***linearly independent***.

Example 2. In $\mathbb{R}^2$, $(1,0)$, $(0,1)$ and $(1,1)$ are linearly dependent.

Definition 3. A ***basis*** is a set of linearly independent vectors that spans a vector space. We call such vectors ***basis vectors*** and denote them by $e_1, e_2, \ldots, e_n$. If there are n basis vectors in a basis, then we say that the corresponding vector space is n***-dimensional***.

Example 3. $\{(1,0),(0,2)\}$, $\{(1,0),(0,1)\}$, $\{(1,0),(1,1)\}$ and $\{(0,1),(1,1)\}$ are all bases of $\mathbb{R}^2$.

Proposition 1. *Let V be a K-vector space with neutral element ϑ. Let $p_1, p_2, \ldots, p_n$ be linearly independent vectors. Then*

$$\sum_{i=1}^{n} \lambda_i p_i = \vartheta \quad \Longrightarrow \quad \lambda_i = 0_K \ \forall\, i \in \{i\}_{i=1}^n$$

where 0_K is the additive identity in K.

Proof. Suppose there exists a $\lambda_\mathrm{j} \neq 0_K$ such that $\sum_{i=1}^n \lambda_i p_i = \vartheta$. Then λ_j has a multiplicative inverse, and hence

$$p_\mathrm{j} = -\lambda_\mathrm{j}^{-1} \sum_{i \neq \mathrm{j}} \lambda_i p_i$$

This would make $p_1, p_2, \ldots, p_n$ linearly dependent, which contradicts our premise. Hence no such λ_j can exist. □

Proposition 2. *Let V be a K-vector space with neutral element ϑ. Let $\{e_i\}_{i=1}^n$ be a basis of V. Any arbitrary vector $x \in V$ can be written as*

$$x = \sum_{i=1}^{n} \lambda_i e_i,$$

where $\{\lambda_i\}_{i=1}^n \subseteq K$ is a unique set of scalars that is characteristic of x.

Remark 1. We refer to the formula

$$x = \sum_{i=1}^{n} \lambda_i e_i$$

as "expanding x in the basis $\{e_i\}_{i=1}^n$". We say that the set of scalars $\{\lambda_i\}_{i=1}^n$ is a ***representation*** of x.

Proof. The fact that $\{e_i\}_{i=1}^n$ spans V implies that there exist $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$ such that for all $x \in V$,

$$x = \sum_{i=1}^{n} \lambda_i e_i.$$

Let us now show that $\{\lambda_i\}_{i=1}^n$ is unique. Suppose that x can also be written as

$$x = \sum_{i=1}^{n} \alpha_i e_i,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n \in K$. It directly follows that

$$\vartheta = x - x = \sum_{i=1}^{n} \lambda_i e_i - \sum_{j=1}^{n} \alpha_j e_j = \sum_{i=1}^{n} (\lambda_i - \alpha_i) e_i.$$

Proposition 1 further implies that $\forall\, i, \lambda_i = \alpha_i$. □

Remark 2. We often use the notation $\lambda_i \equiv x^i$ and call the x^i's the ***components*** or ***coordinates*** of the vector x in the basis $\{e_i\}_{i=1}^n$. We write

$$x = \sum_{i=1}^{n} \lambda_i e_i \equiv \sum_{i=1}^{n} x^i e_i \equiv x^i e_i.$$

The implied summation over pairs of upper and lower indices is called the ***Einstein summation convention***.

Example 4. $\{e_1 = (1, 0, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, 0, \ldots, 1)\}$ is called the ***standard basis*** of $\mathbb{R}^n$.

Remark 3. **Proposition 2** indicates that there is a one-to-one correspondence between any vector $x \in V$ and the n-tuple of its components. One can further show that all n-dimensional K-vector spaces are isomorphic to K^n.

4.3 Tensor Spaces

Definition 1. Let V be a K-vector space.
(a) A mapping $f : V \to K$ is called a ***linear form*** if $\forall\, x, y \in V \wedge \lambda \in K$,
  (i) $f(x + y) = f(x) + f(y)$;
  (ii) $f(\lambda x) = \lambda f(x)$.
(b) A mapping $g : V \times V \to K$ is called a ***bilinear form*** if $\forall\, x, y, z \in V \wedge \lambda \in K$,
  (i) $g(x + y, z) = g(x, z) + g(y, z)$;
  (ii) $g(x, y + z) = g(x, y) + g(x, z)$;
  (iii) $g(\lambda x, y) = \lambda g(x, y) = g(x, \lambda y)$.

Remark 1. Note the relation between bilinear forms and bilinear mappings (See **3.1**).

Definition 2. Let V be a K-vector space. Let $\{e_i\}_{i=1}^n$ be a basis of V. Let $f : V \times V \to K$ be a bilinear form. The scalars $t_{ij} := f(e_i, e_j)$ are called the ***components*** or ***coordinates*** of f in the basis $\{e_i\}_{i=1}^n$.

Proposition 1. *A bilinear form can be completely characterized by its components.*

Proof. Let V be a K-vector space. Let $\{e_i\}_{i=1}^n$ be a basis of V. Let $f : V \times V \to K$ be a bilinear form. Let $x, y \in V$. We have

$$x = \sum_{i=1}^{n} x^i e_i \equiv x^i e_i, \quad \text{and} \quad y = \sum_{j=1}^{n} y^j e_j \equiv y^j e_j.$$

It follows that

$$\begin{aligned} f(x,y) &= f\left(\sum_{i=1}^{n} x^i e_i, y\right) \equiv f(x^i e_i, y) \\ &= \sum_{i=1}^{n} x^i f(e_i, y) \equiv x^i f(e_i, y) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} x^i y^j f(e_i, e_j) \equiv x^i y^j f(e_i, e_j) \equiv x^i y^j t_{ij}. \end{aligned}$$

Hence, after we obtain all the components t_{ij}, the bilinear form is completely determined, because we are able to compute $f(x, y)$ for any arbitrary x and y.

The reader ought to appreciate the Einstein summation convention from now on, unless he or she really relishes $\sum$. For his convenience, the slothful typist will employ the Einstein summation convention in the rest of this note. □

Definition 3. Let $\{e_i\}_{i=1}^n$ be a basis of a K-vector space V. Let $f : V \times V \to K$ be a bilinear form. The scalars $t_{ij} = f(e_i, e_j)$ are also called the ***components*** or ***coordinates*** of a ***rank-2 tensor*** t in the basis $\{e_i\}_{i=1}^n$. A rank-2 tensor is equivalent to a bilinear form. In general, a high-rank tensor is corresponding to a multilinear form.

Theorem 1. *Let K be a field. The set of all rank-2 tensors is an n^2-dimensional K-vector space.*

Proof. See Problem 1.4.2 (The space of rank-2 tensors). □

Example 1. Let us consider $\mathbb{R}^3$ with the standard basis $\{e_1 = (1, 0, 0), e_2 = (0, 1, 0), e_3 = (0, 0, 1)\}$. Recall that $\mathbb{R}^3$ is an $\mathbb{R}$-vector space. The well-known ***Levi-Civita tensor*** or ***completely antisymmetric tensor of rank 3*** is the tensor corresponding to the trilinear form $\varepsilon : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ with components $\varepsilon(e_i, e_j, e_k) = \varepsilon_{ijk}$, where the ***Levi-Civita symbol*** ε_{ijk} is given by

$$\varepsilon_{ijk} = \operatorname{sgn}\begin{pmatrix} i, j, k \\ 1, 2, 3 \end{pmatrix}.$$

Remark 2. The cross product of 3-vectors is conveniently written in terms of ϵ_{ijk}, see Problem 1.4.3 (Cross product of 3-vectors).

Remark 3. Notice that in **Definition 3**, we singled out the phrase "in the basis $\{e_i\}_{i=1}^n$". We would like to accentuate the fact that components of a tensor depend on the basis chosen. For instance, components of the Levi-Civita tensor in an arbitrary basis $\{\tilde{e}_i\}_{i=1}^n$ are generally not given by the Levi-Civita symbol, i.e., $\varepsilon(\tilde{e}_i, \tilde{e}_j, \tilde{e}_k) \neq \varepsilon(e_i, e_j, e_k) = \varepsilon_{ijk}$. Also, for now, note that a symbol is merely a token, i.e., the Levi-Civita symbol is conveniently introduced to express components of the Levi-Civita tensor in the standard basis.

Example 2. Now, let us consider $\mathbb{R}^n$ with the standard basis $\{e_i\}_{i=1}^n$. The rank-2 tensor corresponding to the bilinear form $\delta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ with components

$$\delta(e_i, e_j) = \delta_{ij} := \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \tag{4.3.1}$$

is called (***Euclidean***) ***Kronecker delta***.

Remark 4. δ_{ij} is an example of a symmetric rank-2 tensor, see Problem 1.4.4 (Symmetric tensors).

4.4 Dual Spaces

Let V be an n-dimensional K-vector space. Let $\{e_i\}_{i=1}^n$ be a basis of V. Let $f : V \to K$ be a linear form. For all $x \in V$, we have $x = x^i e_i$, where $x^i \in K$. It follows that

$$f(x) = f(x^i e_i) = x^i f(e_i) \equiv x^i u_i,$$

where $u_i := f(e_i) \in K$.

Remark 1. Every linear form on V can be written in this form, i.e., the scalars u_i uniquely determine the form.

Remark 2. The set of u_i, and hence the linear forms on V, form a K-vector space, denoted V^*, that is isomorphic to K^n and hence to V (see **Theorem 1** of **4.3** and *Remark* 3 of **4.2**).

Definition 1. Let V be an n-dimensional K-vector space. The space V^* of linear forms on $V$is called ***dual*** to V. Elements of V^* are called ***covectors***. There exists a one-to-one correspondence between covectors (elements of V^*) and vectors (elements of V).

Remark 3. Covectors are defined via linear forms, in analogy to rank-n tensors being defined via n-linear forms, hence covectors can be regarded as rank-1 tensors.

Definition 2. The scalar $f(x)$ is called the ***scalar product*** of the vector x and covector u that corresponds to the linear form f. We write $x \cdot u := x^i u_i$.

Remark 4. Covectors are also called ***covariant vectors***, in which case vectors are referred to as ***contravariant vectors***.

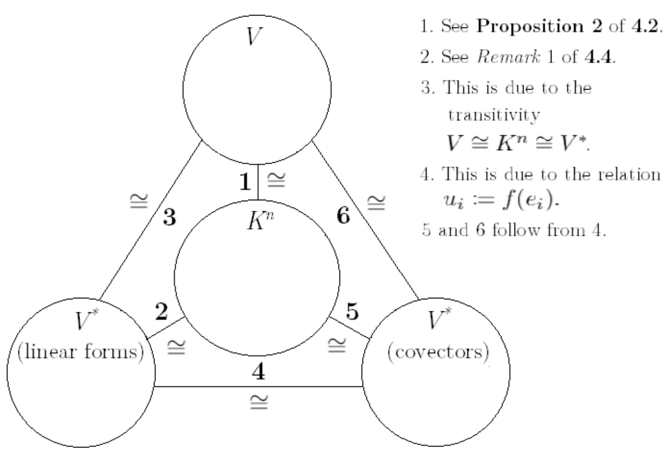


Fig. 4.4.1. A summary of various isomorphisms.

Remark 5. Because $V^* \cong V$, there is no need to distinguish between them. We can define the ***covariant components*** of a vector y to be the components of the covector u that corresponds to y under the isomorphism, i.e., $y_i := u_i$. The ***contravariant components*** of the same vector are the y^i. As a result, we can write the scalar product as $x \cdot u \equiv x \cdot y = x^i y_i$. We will revisit these concepts in **4.8**.

Remark 6. In physics we often denote vectors by $|x\rangle$ and covectors by $\langle y|$ and write the scalar product as $\langle y|x\rangle = y_i x^i$.

Remark 7. The covectors $e^1 = (1, 0, \ldots, 0)$, $e^2 = (0, 1, 0, \ldots, 0)$, etc. form a cartesian basis of V^* called the ***cartesian cobasis***. (Here 0 and 1 are the additive and multiplicative neutral elements, respectively, of K.)

Remark 8. The scalar product

$$e^i \cdot e_j =: \delta^i{}_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{4.4.1}$$

is identical to the Euclidian Kronecker delta defined in §4.3 Example 2.

Remark 9. $\delta^i{}_j$ is *always* defined by (4.4.1). However, $\delta^i{}_j = \delta_{ij}$ only of $\mathbb{R}_n$ is considered a Euclidian space. More generally, $\delta^i{}_j \neq \delta_{ij}$, see §4.8.

Definition 3.
(a) A bilinear form $f : V^* \times V^* \to K$ acting on the co-basis defines a ***contravariant tensor of rank 2*** whose components are given by $t^{ij} := f(e^i, e^j)$. More generally, n-linear forms that map an n-fold Cartesian product of V^* to K define rank-n contravariant tensors. To further distinguish them from the tensors defined in §4.3 we call the latter ***covariant tensors***.
(b) As its name suggests, a ***mixed tensor of rank 2*** is corresponding to a bilinear form $f : V \times V^* \to K$ (or $f : V^* \times V \to K$). A high-rank mixed tensor can be defined in a similar way, e.g., $f : V^* \times V \times V^* \to K$ is equivalent to a rank-3 mixed tensor with components $t^i{}_j{}^k := f(e^i, e_j, e^k)$.

Example 1. $\delta^i{}_j$ in *Remark* 6 is a mixed tensor of rank 2.

Remark 10. Vectors can be regarded as contravariant tensors of rank 1.

Remark 11. $\delta^i{}_j$ takes the cartesian co-basis vector $\#i$ and computes its j^{th} coordinate with respect to the cartesian basis: $(e^i)_j = \delta^i{}_j$. Similarly, $(e_i)^j = \delta_i{}^j$. We see that $\delta^i{}_j = \delta_j{}^i$, which is why some authors (e.g., Landau & Lifshitz) write δ^i_j.

Definition 4. Let x and y be two contravariant vectors. The ***tensor product*** of x and y yields a contravariant tensor whose components equal the product of the two vectors' components, i.e., $t^{ij} := x^i y^j$. We write $t = x \otimes y$.

Remark 12. Any $x \in V \cong V^*$ can be considered either
∘ a vector (i.e., an element of V), or
∘ a co-vector (i.e., the elements of V^* that corresponds to $x \in V$ under the isomorphism.
⇒ We can expand x in either
∘ a basis: $x = x^i e_i$, or
∘ the corresponding co-basis: $x = x_i e^i$
with x^i and x_i the contravariant and covariant coordinates, respectively, of the same vector x.

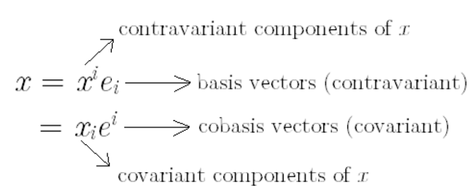


Fig. 4.4.2. Contravariant and covariant components of a vector x.

Remark 13. Although $V \cong V^*$, we do not yet know the isomorphism explicitly. That is, we do not know how basis vectors are related to the corresponding co-basis vector, how covariant vectors are related to contravariant ones, or how the covariant coordinates of a given vector are related to the contravariant coordinates of the same vector. In what we call Euclidian space, $e_i = e^i$ and $x^i = x_i$, and we do not have to distinguish between covariant and contravariant indices (e.g., ${\delta^i}_j = \delta_{ij}$, see Remark 9 above). This property represents an additional postulate that characterizes Euclidian space. There are other possibilities, see §4.8.

4.5 Metric Spaces

Definition 1. Let M be a set. Let $\rho : M \times M \to \mathbb{R}$ be a mapping. If ρ possesses the following properties: $\forall\, x, y, z \in M$,

1. $\rho(x, y) \geqslant 0 \;\wedge\; (\rho(x, y) = 0 \iff x = y)$; (positive semidefiniteness)
2. $\rho(x, y) = \rho(y, x)$; (symmetry)
3. $\rho(x, z) \leqslant \rho(x, y) + \rho(y, z)$, (triangle inequality)

we will call M a ***metric space*** with the ***metric*** ρ.

Remark 1. A set with a designated metric defines a metric space.

Example 1. Let $M = \mathbb{R}$. We can define a metric on M: $\forall\, x, y \in M$,

$$\rho(x, y) = |x - y| \coloneqq \begin{cases} x - y, & \text{if } x \geqslant y, \\ y - x, & \text{otherwise.} \end{cases}$$

The reader ought to verify that such a ρ satisfies the three properties in **Definition 1** (See **Problem 1.4.5**).

Definition 2. Let M be a metric space with metric ρ. Let $(a_n)_{n\in\mathbb{N}} \subseteq M$ be an infinite sequence. We say that $L \in M$ is the ***limit*** of the sequence (or the sequence ***converges*** to L), if $\forall\, \epsilon > 0, \exists\, N \in \mathbb{N} : \forall\, n > N, \rho(a_n, L) < \epsilon$. We write $\lim\limits_{n\to\infty} a_n = L$, $a_n \to L$ or $\lim\limits_{n\to\infty} \rho(a_n, L) = 0$.

Proposition 1. *A sequence has at most one limit.*

Proof. See **Problem 1.4.6**. □

Definition 3. Let M be a metric space with metric ρ. An infinite sequence $(a_n)_{n\in\mathbb{N}} \subseteq M$ is called a ***Cauchy sequence*** if it satisfies the ***Cauchy condition***:

$$\forall\, \epsilon > 0, \exists\, N \in \mathbb{N} : \forall\, m, n > N, \rho(a_m, a_n) < \epsilon.$$

Remark 2. For Cauchy sequences, we write $\lim\limits_{m,n\to\infty} \rho(a_m, a_n) = 0$, or simply $\rho(a_m, a_n) \to 0$.

Proposition 2. *Every convergent sequence is a Cauchy sequence.*

Proof. See **Problem 1.4.6**. □

Remark 3. The converse of **Proposition 2** is not true in a general metric space.

Example 2. Let $M = \mathbb{Q}$ with the metric defined in **Example 1** and $a_n = \left(1 + \frac{1}{n}\right)^n$. a_n is divergent, since $\lim_{n\to\infty} a_n = e \notin \mathbb{Q}$.

Proposition 3. *Let $M = \mathbb{R}$ with the metric defined in **Example 1**. Every Cauchy sequence in M is convergent.*

Sketch of Proof.

(i) Firstly, we need to show that every Cauchy sequence is bounded. This follows from the Cauchy condition.

(ii) Secondly, we establish that a Cauchy sequence is convergent if and only if it has a convergent subsequence.

(iii) Finally, we employ ***Bolzano-Weierstrass theorem*** that every bounded sequence in $\mathbb{R}$ has a convergent subsequence to complete the proof. □

Definition 4. A metric space M is called ***complete*** if every Cauchy sequence in M converges.

Remark 4. In **Example 2**, we saw that $\mathbb{Q}$ is not complete.

Proposition 4. *An incomplete metric space M can be completed by adding an appropriate set. The completion of M is unique up to isomorphism.*

We omit the proof, because it is beyond the level of this course.

Example 3. $\mathbb{R}$ is a completion of $\mathbb{Q}$.

4.6 Banach Spaces

Definition 1. Let B be an $\mathbb{R}$-vector space (or a $\mathbb{C}$-vector space) with null vector ϑ. Let $\|\cdot\| : B \to \mathbb{R}$ be a mapping. If $\|\cdot\|$ possesses the following properties: $\forall\, x, y \in B \wedge a \in \mathbb{R}$ (or $\mathbb{C}$),

1. $\|x\| \geqslant 0 \;\wedge\; (\|x\| = 0 \iff x = \vartheta)$; (positive semidefiniteness)
2. $\|x + y\| \leqslant \|x\| + \|y\|$; (triangle inequality)
3. $\|ax\| = |a|\, \|x\|$, (linearity)

we will call $\|\cdot\|$ a ***norm*** on B.

The image of a vector $x \in B$, $\|x\|$ is called its ***norm***. The norm of the difference of two vectors $x, y \in B$, $\|x - y\| =: d(x, y)$ is called the ***distance*** between x and y.

Remark 1. $d : B \times B \to \mathbb{R}$ defines a metric on B. The reader ought to verify that d satisfies the three properties in **Definition 1** of **4.5**. There exist other metrics.

Remark 2. Let $x \in B$. Notice that $\|x\| = \|x - \vartheta\| = d(x, \vartheta)$.

Remark 3. As we saw, a normed vector space is a metric space. However, for an arbitrary set, the existence of a metric does not imply that a norm also exists.

Remark 4. Every linear space over $\mathbb{R}$ (or $\mathbb{C}$) with a norm is a metric space.

Definition 2. A linear space over $\mathbb{R}$ (or $\mathbb{C}$) with a norm that is complete is called a ***Banach space***, or simply ***B-space***.

Example 1. As a vector space, $\mathbb{R}$ with the norm given by $\|x\| \coloneqq |x|$ ($x \in \mathbb{R}$) is a Banach space. Similarly, $\mathbb{C}$ with $\|z\| \coloneqq |z| = \sqrt{{z_1}^2 + {z_2}^2}$ ($z = z_1 + iz_2 \in \mathbb{C}$) is also a B-space.

Definition 3. Let B be a Banach space over $\mathbb{C}$. Let $\ell : B \to \mathbb{C}$ be a linear form (See **Definition 1**, (a) of **4.3**). Let $x \in B$. The norm of ℓ is defined as

$$\|\ell\| \coloneqq \sup_{\|x\|=1} \Big\{|\ell(x)|\Big\}.$$

Remark 5. The vector space of linear forms on B, B^* is the dual vector space to B (See **Definition 1** of **4.4**).

Proposition 1. *The norm of linear forms defines a norm on* B^*.

Proof. See **Problem 1.4.7**. □

Theorem 1. B^* *is complete and hence a B-space.*

We omit the proof, because it is beyond the level of this course. The interested reader might want to consult *A Course of Higher Mathematics, Vol. 5* by V. I. Smirnov.

4.7 Hilbert Spaces

Definition 1. Let H be a linear space over $\mathbb{C}$ with null vector ϑ. Let $(\cdot,\cdot) : H \times H \to \mathbb{C}$ be a mapping that possesses the following properties: $\forall\, x, y, z \in H \land \lambda \in \mathbb{C}$,

(i) $(x, y) = (y, x)^*$; (symmetry)

(ii) $(x, x) \geqslant 0 \;\land\; ((x, x) = 0 \iff x = \vartheta)$; (positive semidefiniteness)

(iii) $(x + y, z) = (x, z) + (y, z)$;

(iv) $(\lambda x, y) = \lambda^*(x, y)$.

The norm of a vector $x \in H$ is defined by $\|x\| \coloneqq \sqrt{(x, x)}$.

Remark 1. The mapping $(\cdot,\cdot)$ is usually called a ***scalar product*** on H.

Remark 2. Note the subtle difference between the scalar product and bilinear forms (See **Definition 1**, (b) of **4.3**).

Lemma 1. *For $x, y \in H$,*

$$|(x,y)|^2 \leqslant (x,x)(y,y).$$

The inequality is named **Cauchy-Schwarz inequality** *(also known as* **Bunyakovsky inequality***).*

Proof. Obviously, the inequality holds when either x or y is the null vector. Let $x, y \neq \vartheta$. As a result, $(x,x), (y,y) > 0$. Let us define

$$z := x - \frac{(y,x)}{(y,y)} y.$$

Notice that

$$(z,y) = \left(x - \frac{(y,x)}{(y,y)} y, y \right) = (x,y) - \frac{(y,x)^*}{(y,y)^*}(y,y) = (x,y) - (y,x)^* = 0.$$

Now, let us compute (x,x):

$$\begin{aligned}
(x,x) &= \left(z + \frac{(y,x)}{(y,y)} y, z + \frac{(y,x)}{(y,y)} y \right) \\
&= (z,z) + \left(z, \frac{(y,x)}{(y,y)} y \right) + \left(\frac{(y,x)}{(y,y)} y, z \right) + \left(\frac{(y,x)}{(y,y)} y, \frac{(y,x)}{(y,y)} y \right) \\
&= (z,z) + \frac{(y,x)}{(y,y)}(z,y) + \frac{(y,x)^*}{(y,y)^*}(z,y)^* + \frac{|(y,x)|^2}{(y,y)^2}(y,y) \\
&= (z,z) + \frac{|(y,x)|^2}{(y,y)}.
\end{aligned}$$

Because of positive semidefiniteness, we have

$$(x,x) \geqslant \frac{|(y,x)|^2}{(y,y)},$$

which implies the Cauchy-Schwarz inequality. □

Remark 3. A similar inequality exists for any scalar product defined on an arbitrary linear space. $(\cdot,\cdot)$ on H is a particular instance.

Proposition 1. *The norm in* ***Definition 1*** *is indeed a norm (See* ***Definition 1*** *of* ***4.6****).*

Proof. See **Problem 1.4.8**. □

Definition 2. Let $x, y \in H$. A metric on H can be defined by $\rho(x,y) := \|x - y\| = \sqrt{(x-y, x-y)}$.

Proposition 2. *The metric in* ***Definition 2*** *satisfies the three properties in* ***Definition 1*** *of* ***4.5****.*

The proof is straightforward and thus left as an exercise to the reader.

Definition 3. A complete H is called a ***Hilbert space***, or simply an ***H-space***.

Remark 4. Every H-space is a B-space.

Definition 4. Let $y \in H$ be given. We can define a linear form $\ell : H \to \mathbb{C}$ by $\ell(x) \coloneqq (y, x)$ for all $x \in H$.

Proposition 3. *The linear form in **Definition 4** is indeed a linear form (See **Definition 1** of 4.3).*

Proof. See **Problem 1.4.8**. □

Proposition 4. *Every linear form on H can be expressed in the form of $\ell(x)$, i.e., $\forall\, \ell : H \to \mathbb{C}, \exists!\, y \in H : \forall\, x \in H, \ell(x) = (y, x)$.*

We omit the proof, because it is beyond the level of this course.

Corollary 1. *Like B-spaces, the vector space of linear forms on H, H^* is the dual vector space to H (See **Definition 1** of 4.4). H^* is isomorphic to H. What is more, H^* itself is an H-space.*

Like before, we omit the proof, because it is beyond the level of this course.

Definition 5. Let $\ell \in H^*$ with the corresponding vector $y \in H$. We can define a mapping $\langle\cdot|\cdot\rangle : H^* \times H \to \mathbb{C}$ by $\langle \ell | x\rangle \coloneqq \ell(x) = (y, x)$ for all $x \in H$.

Remark 5. For each $\ell \in H^*$, there exists a unique $y \in H$ such that $\langle \ell | x\rangle = \ell(x) = (y, x)$.

Remark 6. Because $H^* \cong H$, there is no need to distinguish between them. We sloppily write $\langle y|x\rangle \coloneqq \langle \ell | x\rangle = (y, x)$. Note that $\langle y|$ is a linear functional, which takes a vector and returns a complex number.

Remark 7. In quantum mechanics, states of a system are represented by elements of a Hilbert space.

4.8 Generalized Metrics and Minkowski Spaces

4.8.1 Scalar Products

Definition 1. Let V be an n-dimensional $\mathbb{R}$-vector space. Let $\{e_i\}_{i=1}^n$ be a basis. Let $g : V \times V \to \mathbb{R}$ be a symmetric bilinear form, i.e., $\forall\, x, y \in V, g(x, y) = g(y, x)$. g corresponds to a symmetric rank-2 tensor whose components are given by $g_{ij} = g(e_i, e_j) = g(e_j, e_i) = g_{ji}$. Let g have an inverse g^{-1} with components $(g^{-1})_{ij} = g^{ij}$, where $g_{ij} g^{jk} = \delta_i^k$.

Let $x, y \in V$. We call the real number $g(x, y) \equiv x \cdot y \equiv xy = x^i g_{ij} y^j$ the (***generalized***) ***scalar product*** of x and y. g is called the (***generalized***) ***metric***, or equivalently the ***metric tensor***.

Remark 1. The metric in **Definition 1** is not the same as the one defined in **4.5**. For instance, positive semidefiniteness can be violated, i.e., $\exists\, x, y \in V : g(x, y) < 0$; it is even possible that $g(x, x) < 0$.

Remark 2. Recall that the $\mathbb{R}$-vector space V is isomorphic to $\mathbb{R}^n$ (See *Remark* 3 of **4.2**). Therefore, in the rest of this section, we will just consider $\mathbb{R}^n$ with a metric g instead of a general $\mathbb{R}$-vector space.

Definition 2. An ***adjoint basis*** (or a ***cobasis***) $\{e^i\}_{i=1}^n$ is a set of ***cobasis vectors*** $e^i := g^{ij} e_j$.

Remark 3. Such defined e^i's are vectors in V, while cobasis vectors in **4.4** are elements of V^*. However, because $V \cong V^*$, we can obscure the difference here by defining cobasis vectors in V.

Remark 4. The relation between e_i and e^j is given by

$$e_i = \delta_i^k e_k = (g_{ij} g^{jk}) e_k = g_{ij}(g^{jk} e_k) = g_{ij} e^j.$$

Definition 3. Let $x \in V$ be given. Coordinates of x in a basis $\{e_i\}_{i=1}^n$, x^i are called ***contravariant***. Coordinates of x in the cobasis $\{e^i\}_{i=1}^n$, x_i are called ***covariant***.

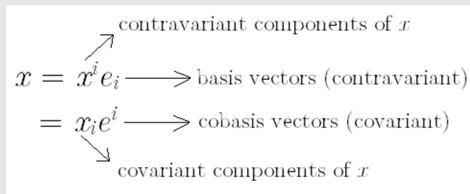


Fig. 4.4.2. Contravariant and covariant components of a vector x.

Remark 5. So far, all the definitions in this section are consistent with the ones in **4.4**. However, we have now specified a relation between bases and cobases (See *Remark* 11 of **4.4**).

Proposition 1. *Let $x \in V$. The contravariant and covariant components of x are related by*

$$x_i = g_{ij} x^j, \quad \textit{and} \quad x^i = g^{ij} x_j.$$

Proof. For the first equality, we have

$$x_i e^i = x = x^j e_j = x^j (g_{ji} e^i) = (x^j g_{ji}) e^i = (g_{ij} x^j) e^i,$$

which implies $x_i = g_{ij} x^j$. For the second,

$$x^i = \delta_k^i x^k = (g^{ij} g_{jk}) x^k = g^{ij}(g_{jk} x^k) = g^{ij} x_j.$$

□

Corollary 1. *Let $x, y \in V$. The scalar product of x and y can now be written as*

$$g(x,y) = x^i g_{ij} y^j = \begin{cases} (x^i g_{ij}) y^j = x_j y^j \\ x^i (g_{ij} y^j) = x^i y_i. \end{cases}$$

Remark 6. The form of the scalar product in **Corollary 1** is consistent with *Remark* 4 of **4.4**.

Remark 7. According to Eq. (4.4.1),

$$g(e^i, e_j) = g^i{}_j = e^i \cdot e_j = \delta_j^i \neq \delta_{ij} = g_{ik} \delta_j^k = g_{ij}.$$

In Euclidean space, $\delta_j^i = \delta_{ij}$. We will revisit this concept in **4.8.3**.

4.8.2 Basis Transformations

Definition 4. A ***real*** $m \times n$ ***matrix*** D is a rectangular array of real numbers in m rows and n columns:

$$D = \begin{pmatrix} D^1{}_1 & D^1{}_2 & D^1{}_3 & \cdots & D^1{}_n \\ D^2{}_1 & D^2{}_2 & D^2{}_3 & \cdots & D^2{}_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D^m{}_1 & D^m{}_2 & D^m{}_3 & \cdots & D^m{}_n \end{pmatrix}.$$

The $D^i{}_j$'s are called ***matrix elements***. In this course, we mainly consider ***real square matrices***, i.e., real matrices with an equal number of rows and columns.

(i) A square matrix D is ***invertible*** if there exists another square matrix D^{-1} such that $D^i{}_j (D^{-1})^j{}_k = (D^{-1})^i{}_j D^j{}_k = \delta^i_k$. We also write $DD^{-1} = D^{-1}D = \mathbb{1}_n$, where $\mathbb{1}_n$ is the $n \times n$ ***identity matrix***

$$\text{n rows}\left\{ \overbrace{\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}}^{\text{n columns}} \right. \equiv \operatorname{diag}\{\underbrace{1, 1, \ldots, 1}_{n\ 1\text{'s}}\}.$$

(ii) The ***transpose*** of an $m \times n$ matrix D, D^T is an $n \times m$ matrix with matrix elements $(D^T)^i{}_j = D_j{}^i$.

(iii) The ***product*** of an $l \times m$ matrix A and an $m \times n$ matrix B is an $l \times n$ matrix with matrix elements $(AB)^i{}_j := A^i{}_k B^k{}_j$.

(iv) The ***determinant*** of an $n \times n$ square matrix D is a number defined by

$$\det D \equiv \begin{vmatrix} D^1{}_1 & D^1{}_2 & D^1{}_3 & \cdots & D^1{}_n \\ D^2{}_1 & D^2{}_2 & D^2{}_3 & \cdots & D^2{}_n \\ D^3{}_1 & D^3{}_2 & D^3{}_3 & \cdots & D^3{}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D^n{}_1 & D^n{}_2 & D^n{}_3 & \cdots & D^n{}_n \end{vmatrix} =: \sum_{\pi \in S_n} \left(\operatorname{sgn} \pi \prod_{i=1}^{n} D^i{}_{\pi(i)} \right),$$

where S_n is the symmetric group (See **Proposition 1** of **2.3**).

Remark 8. Matrix elements are not like components of a tensor: we usually do not distinguish between upper and lower indices. For matrices, the distinction is only important when Einstein summation convention is involved. Therefore, for a matrix D, we can usually write $D^i{}_j = D_{ij} = D_i{}^j = D^{ij}$. It is crucial to tell the difference between row and column indices though.

Example 1. Let us consider a general 2×2 matrix D and compute its determinant:

$$\begin{aligned} \det D = \begin{vmatrix} D^1{}_1 & D^1{}_2 \\ D^2{}_1 & D^2{}_2 \end{vmatrix} &= \sum_{\pi \in S_2} \left(\operatorname{sgn} \pi \prod_{i=1}^{2} D^i{}_{\pi(i)} \right) \\ &= \operatorname{sgn} \begin{pmatrix} 1,2 \\ 1,2 \end{pmatrix} \prod_{i=1}^{2} D^i{}_{\binom{1,2}{1,2}(i)} + \operatorname{sgn} \begin{pmatrix} 1,2 \\ 2,1 \end{pmatrix} \prod_{j=1}^{2} D^j{}_{\binom{1,2}{2,1}(j)} \\ &= 1 \cdot D^1{}_1 \cdot D^2{}_2 + (-1) \cdot D^1{}_2 \cdot D^2{}_1 = D^1{}_1 D^2{}_2 - D^1{}_2 D^2{}_1. \end{aligned}$$

Proposition 2. *Let A, B and D be matrices. We have*
(i) $(AB)^T = B^T A^T$*;*
(ii) $(D^{-1})^T = (D^T)^{-1}$*;*
(iii) $\det(AB) = \det A \cdot \det B$*;*
(iv) $\det\left(D^{-1}\right) = \dfrac{1}{\det D}$*;*
(v) $\det\left(D^T\right) = \det D$*.*

Proof. For (i) and (ii), we will just write symbolic expressions.
(i) $\left((AB)^T\right)^i{}_j = (AB)_j{}^i = A_j{}^k B_k{}^i = (A^T)^k{}_j (B^T)^i{}_k = (B^T)^i{}_k (A^T)^k{}_j = (B^T A^T)^i{}_j$
(ii) $D^T (D^{-1})^T = (D^{-1}D)^T = \mathbb{1}_n{}^T = \mathbb{1}_n$
$(D^{-1})^T D^T = (DD^{-1})^T = \mathbb{1}_n{}^T = \mathbb{1}_n$
$\therefore\ (D^{-1})^T = (D^T)^{-1}$
For (iii), (iv) and (v), please consult *A Course of Higher Mathematics, Vol. 3* by V. I. Smirnov. □

Definition 5. Let us consider $\mathbb{R}^n$. Let $\{e_i\}_{i=1}^n$ be a basis. Let D be an invertible $n \times n$ real matrix. We can define a second basis $\{\tilde{e}_i\}_{i=1}^n$ by the ***basis transformation*** $\tilde{e}_i \coloneqq e_j (D^{-1})^j{}_i$.

Remark 9. The ***inverse basis transformation*** is given by

$$e_i = e_k \delta_i^k = e_k \left((D^{-1})^k{}_j D^j{}_i\right) = \left(e_k (D^{-1})^k{}_j\right) D^j{}_i = \tilde{e}_j D^j{}_i.$$

Proposition 3. $\{\tilde{e}_i\}_{i=1}^n$ *is indeed a basis (See* ***Definition 3*** *of* ***4.2****).*

Proof. To show that $\{\tilde{e}_i\}_{i=1}^n$ is a basis, we need to establish that $\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n$ are linearly independent and span the $\mathbb{R}$-vector space $\mathbb{R}^n$. Let $x \in \mathbb{R}^n$. From *Remark* 9 it follows that

$$x = x^j e_j = x^j (\tilde{e}_i D^i{}_j) = (D^i{}_j x^j)\tilde{e}_i \coloneqq \tilde{x}^i \tilde{e}_i, \tag{4.8.1}$$

i.e., $\{\tilde{e}_i\}_{i=1}^n$ spans $\mathbb{R}^n$. Now, let $\{\tilde{\lambda}^i\}_{i=1}^n \subset \mathbb{R}$. Let us consider the sum $S = \tilde{\lambda}^i \tilde{e}_i$. According to **Definition 5**, we have

$$S = \tilde{\lambda}^i \tilde{e}_i = \tilde{\lambda}^i (e_j (D^{-1})^j{}_i) = ((D^{-1})^j{}_i \tilde{\lambda}^i) e_j \coloneqq \lambda^j e_j.$$

The fact that $\{e_i\}_{i=1}^n$ is a basis implies that $S = 0 \Rightarrow \lambda^i = 0$ ($i \in \{i\}_{i=1}^n$). Furthermore, because D^{-1} is invertible, $\tilde{\lambda}^i = 0$ ($i \in \{i\}_{i=1}^n$). We have thus shown that $\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n$ are linearly independent. □

Proposition 4. *Let $x \in \mathbb{R}^n$ be a vector with components x^i in the basis $\{e_i\}_{i=1}^n$. In the basis $\{\tilde{e}_i\}_{i=1}^n$, components of x are given by $\tilde{x}^i = D^i{}_j x^j$.*

Proof. See Eq. (4.8.1). □

Remark 10. The inverse relation is $x^i = (D^{-1})^i{}_j \tilde{x}^j$.

Remark 11. Such a D applied on components of a vector is called a ***coordinate transformation***.

Proposition 5. *Let $g_{ij} = e_i \cdot e_j$ be the metric associated with the basis $\{e_i\}_{i=1}^n$. Let D^{-1} be a basis transformation $\tilde{e}_i = e_j {(D^{-1})^j}_i$. The metric that corresponds to the new basis $\{\tilde{e}_i\}_{i=1}^n$ is given by*

$$\tilde{g}_{ij} = {\left((D^{-1})^T\right)_i}^k g_{k\ell} {(D^{-1})^\ell}_j \quad (\text{or } \tilde{g} = (D^{-1})^T g D^{-1}).$$

The inverse relation is $g = D^T \tilde{g} D$.

Proof. For this proof, we will just write symbolic expressions.

(1)

$$\tilde{g}_{ij} = \tilde{e}_i \cdot \tilde{e}_j = e_k {(D^{-1})^k}_i \cdot e_\ell {(D^{-1})^\ell}_j = {\left((D^{-1})^T\right)_i}^k (e_k \cdot e_\ell) {(D^{-1})^\ell}_j = {\left((D^{-1})^T\right)_i}^k g_{k\ell} {(D^{-1})^\ell}_j$$

(2)

$$\begin{aligned} \tilde{g} &= (D^{-1})^T g D^{-1} \\ D^T \tilde{g} D &= D^T (D^{-1})^T g D^{-1} D \\ g &= D^T \tilde{g} D \end{aligned}$$

□

4.8.3 Normal Coordinate Systems

Lemma 1. *For all invertible $n \times n$ symmetric matrices $M = M^T$ with complex matrix elements, there exists a matrix D such that $M = D^T M D = \mathrm{diag}\{m_1, m_2, \ldots, m_n\}$, where $m_1, m_2, \ldots, m_n$ are nonzero complex numbers.*

This is called (***finite-dimensional***) ***spectral theorem***. We omit the proof, because it is well-established; the reader will be able to find the proof in general textbooks on linear algebra.

Corollary 2. *Let g_{ij} be a metric on $\mathbb{R}^n$. Recall that g can be represented by an invertible real symmetric matrix. According to **Lemma 1**, there exists a transformation such that $\tilde{g}_{ij} = \lambda_i \cdot \delta_{ij}$ (the $\cdot$ indicates that this is not a summation), where δ_{ij} is the Euclidean Kronecker delta, and λ_i's are nonzero complex numbers.*

Theorem 1. *Let g_{ij} be a metric on $\mathbb{R}^n$. There exists a transformation such that*

$$g^* = \mathrm{diag}\{\underbrace{1, \ldots, 1}_{m \text{ 1's}}, \underbrace{-1, \ldots, -1}_{n-m \ -1\text{'s}}\} \quad (0 \leqslant m \leqslant n). \tag{4.8.2}$$

Proof. **Corollary 2** ensures the existence of a transformation such that $\tilde{g}_{ij} = \lambda_i \cdot \delta_{ij}$, where δ_{ij} is the Euclidean Kronecker delta, and $\lambda_i \neq 0$ ($i \in \{i\}_{i=1}^n$). To complete the proof, we would like to first permute the order of basis vectors so that $\lambda_1, \lambda_2, \ldots, \lambda_m > 0$, and $\lambda_{m+1}, \lambda_{m+2}, \ldots, \lambda_n < 0$. Now, let us define a second transformation by the symmetric matrix

$$(D^{-1})^i{}_j \coloneqq \frac{1}{\sqrt{|\lambda_i|}} \cdot \delta^i_j.$$

According to **Proposition 5**, the new metric is given by

$$\begin{aligned}
\tilde{\tilde{g}}_{ij} &= \left((D^{-1})^T\right)_i{}^k \tilde{g}_{k\ell} (D^{-1})^\ell{}_j \\
&= \left(\frac{1}{\sqrt{|\lambda_i|}} \cdot \delta^k_i\right) (\lambda_k \cdot \delta_{k\ell}) \left(\frac{1}{\sqrt{|\lambda_\ell|}} \cdot \delta^\ell_j\right) \\
&= \frac{\lambda_k}{\sqrt{|\lambda_i \lambda_\ell|}} \cdot \delta^k_i \delta_{k\ell} \delta^\ell_j = \frac{\lambda_i}{\sqrt{|\lambda_i \lambda_\ell|}} \cdot \delta_{i\ell} \delta^\ell_j = \frac{\lambda_i}{\sqrt{|\lambda_i \lambda_j|}} \cdot \delta_{ij} \\
&= \frac{\lambda_i}{|\lambda_i|} \cdot \delta_{ij} = \begin{cases} \delta_{ij}, & \text{if } i \leqslant m, \\ -\delta_{ij}, & \text{if } m < i \leqslant n, \end{cases}
\end{aligned}$$

as desired. □

Definition 6. If the metric corresponding to a basis is of the form in Eq. (4.8.2), we will call the basis a ***normal coordinate system***.

Remark 12. The integer m in Eq. (4.8.2) is characteristic of the vector space and remains invariant under basis transformations. This is an implication of ***Sylvester's rigidity theorem***.

Example 2. Let $m = n$ in Eq. (4.8.2). We have $g = \mathbb{1}_n$. One can further check that $g_{ij} = \delta_{ij}$ is compatible with **Definition 1** of **4.5**. The $\mathbb{R}$-vector space, $\mathbb{R}^n$ with the metric $\mathbb{1}_n$ is named the ***n-dimensional Euclidean space***, denoted E^n. In E^n, normal coordinate systems are called ***Cartesian coordinate systems***. Notice that for any $x \in E^n$,

$$x^i = \delta_{ij} x^j = g_{ij} x^j = x_i,$$

i.e., there is no need to distinguish between being contravariant and covariant in Euclidean space (See *Remark* 11 of **4.4**).

Example 3. Now, let $m = 1$ and $n \geqslant 2$. In this case, we have $g = \mathrm{diag}\{1, \underbrace{-1, \ldots, -1}_{n-1\ -1\text{'s}}\}$ that is a generalized metric. $\mathbb{R}^n$ with the metric g is called the ***n-dimensional Minkowski space*** M^n. In this space, normal coordinate systems are called ***inertial coordinate frames***. It is straightforward to show that for any $x \in M^n$, its contravariant and covariant components are related by

$$x^i = \begin{cases} x_i, & \text{if } i = 1, \\ -x_i, & \text{if } 1 < i \leqslant n. \end{cases}$$

Remark 13. One formulation of the special relativity is based on the postulate that classical mechanical systems can be described as collections of particles moving in the space M^4. Let $x \in M^4$. As physicists, we often use the notation $x = (x^0, x^1, x^2, x^3) \coloneqq (ct, \mathbf{x})$, where ct is the temporal component of the ***four-vector*** x, and $\boldsymbol{x}$ the spatial component. Thereinto, c is a characteristic velocity, namely the speed of light in vacuum.

4.8.4 Normal Coordinate Transformations

Definition 7. A coordinate transformation D is ***normal*** if it transforms one normal coordinate system into another. In other words, the metric of the form in Eq. (4.8.2) is invariant under a normal coordinate transformation, i.e.,

$$g_{ij} = \tilde{g}_{ij} = \left((D^{-1})^T\right)_i{}^k g_{k\ell} (D^{-1})^\ell{}_j \quad (\text{or } g = \tilde{g} = (D^{-1})^T g D^{-1}),$$

which implies that $g = D^T g D$.

Example 4. Let D be a normal coordinate transformation in the n-dimensional Euclidean space. Recall that in E^n, $g = \mathbb{1}_n$. According to **Definition 7**, D must satisfy the relation

$$\mathbb{1}_n = D^T \mathbb{1}_n D = D^T D.$$

We call such a D ***orthogonal***.

Example 5. In M^n, normal coordinate transformations are called ***Lorentz transformations***.

Lemma 2.
(i) The inverse of a normal coordinate transformation is also normal.
(ii) The product of two normal coordinate transformations is normal as well.

Proof. Let g be a metric of the form in Eq. (4.8.2).
(i) See **Definition 7**.
(ii) Let D_1 and D_2 be normal coordinate transformations. By definition, we have

$$g = D_1{}^T g D_1, \quad \text{and} \quad g = D_2{}^T g D_2.$$

Combining the two equalities, we obtain

$$g = D_1{}^T \left(D_2{}^T g D_2\right) D_1 = \left(D_1{}^T D_2{}^T\right) g (D_2 D_1) = (D_2 D_1)^T g (D_2 D_1).$$

□

Theorem 2. *All the normal coordinate transformations for a specific metric form a non-abelian group under matrix multiplication.*

Proof. To complete the proof, we need to check that the set of all the normal coordinate transformations satisfies the four group axioms:
(i) closure is satisfied because of **Lemma 2**, (ii);
(ii) matrix multiplication is associative;
(iii) the identity matrix $\mathbb{1}_n$ always serves as the multiplicative identity;
(iv) existence of inverses is due to **Lemma 2**, (i). □

Remark 14. The group of all the normal coordinate transformations in E^n is called the ***orthogonal group***, denoted $O(n)$. In M^n, the group of all the Lorentz transformations is called the ***pseudo-orthogonal group***, denoted $O(1, n-1)$.

Proposition 6. *Let g be a metric of the form in Eq.* (4.8.2)*. Let D be a normal coordinate transformation. We have* $\det D = \pm 1$.

Proof. According to **Definition 7**, we have

$$\begin{aligned} g &= D^T g D \\ \det g &= \det\left(D^T g D\right) = \det\left(D^T\right) \cdot \det g \cdot \det D \\ 1 &= (\det D)^2 \\ \det D &= \pm 1. \end{aligned}$$

□

4.9 Problems

1.4.1 **Function space**

Consider the set C of continuous functions $f : [0,1] \to \mathbb{R}$. Show that by suitably defining an addition on C, and a multiplication with real numbers, one can make C an additive vector space over $\mathbb{R}$.

(2 points)

1.4.2. **The space of rank-2 tensors**

a) Prove the theorem of ch.1 §4.3: Let V be a vector space V of dimension n over K. Then the space of rank-2 tensors, defined via bilinear forms $f : V \times V \to K$, forms a vector space of dimension n^2.

b) Consider the space of bilinear forms f on V that is equivalent to the space of rank-2 tensors, and construct a basis of that space.

hint: On the space of tensors, define a suitable addition and multiplication with scalars, and construct a basis of the resulting vector space.

(5 points)

1.4.3. **Cross product of 3-vectors**

Let $x, y \in \mathbb{R}_3$ be vectors, and let ϵ_{ijk} be the Levi-Civita symbol. Show that the (covariant) components of the cross product $x \times y$ are given by

$$(x \times y)_i = \epsilon_{ijk} x^j y^k$$

(1 point)

1.4.4. **Symmetric tensors**

Let V be an n-dimensional vector space over K with some basis, let $f : V \times V \to K$ be a bilinear form, and let t be the rank-2 tensor defined by f. Show that f is symmetric, i.e. $f(x, y) = f(y, x)\ \forall x, y \in V$, if and only if the components of the tensor with respect to the given basis are symmetric, i.e., $t_{ij} = t_{ji}$.

(2 points)

1.4.5. $\mathbb{R}$ **as a metric space**

Consider the reals $\mathbb{R}$ with $\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by $\rho(x, y) = |x - y|$. Show that this definition makes $\mathbb{R}$ a metric space.

(3 points)

1.4.6. **Limits of sequences**

a) Show that a sequence in a metric space has at most one limit.

   hint: Assume there are two limits, and use the triangle inequality to show that they must be the same.

b) Show that every sequency with a limit is a Cauchy sequence.

(3 points)

1.4.7. **Banach space**

Let B be a K-vector space (k = $\mathbb{R}$ or $\mathbb{C}$) with null vector θ. Let $|| \ldots || \ :\ \text{B} \to \mathbb{R}$ be a mapping such that

(i) $||x|| \geq 0\ \forall\ x \in \text{B}$, and $||x|| = 0$ iff $x = \theta$.

(ii) $||x + y|| \leq ||x|| + ||y||\ \forall\ x, y \in \text{B}$.

(iii) $||\lambda x|| = |\lambda| \cdot ||x||\ \forall\ x \in \text{B}, \lambda \in \text{K}$.

Then we call $|| \ldots ||$ a **norm** on B, and $||x||$ the **norm** of x.

Further define a mapping $d\ :\ \text{B} \times \text{B} \to \mathbb{R}$ by

$d(x, y) := ||x - y||\ \forall\ x, y \in \text{B}$

Then we call $d(x, y)$ the **distance** between x and y.

a) Show that d is a metric in the sense of §4.5, i.e., that every linear space with a norm is in particular a metric space.

If the normed linear space B with distance/metric d is complete, then we call B a **Banach space** or **B-space**.

b) Show that $\mathbb{R}$ and $\mathbb{C}$, with suitably defined norms, are B-spaces. (For the completeness of $\mathbb{R}$ you can refer to §4.5 example (3), and you don't have to prove the completeness of $\mathbb{C}$ unless you insist.)

Now let B^* be the dual space of B, i.e., the space of linear functionals ℓ on B, and define a norm of ℓ by

$||\ell|| := \sup_{||x||=1} \{|\ell(x)|\}$

c) Show that the such defined norm on B^* is a norm in the sense of the norm defined on B above.

(In case you're wondering: B^* is complete, and hence a B-space, but the proof of completeness is difficult.)

(5 points)

1.4.8. **Hilbert space**

a) Show that the norm on a Hilbert space defined by §4.7 def. 1 is a norm in the sense of the definition in Problem 19.

   hint: Use the Cauchy-Schwarz inequality (§4.7 lemma).

b) Show that the mappings ℓ defined in §4.7 def. 4 are linear forms in the sense of §4.3 def. 1(a).

(3 points)

1.4.9. **Lorentz transformations in** M_2

Consider the 2-dimensional Minkowski space M_2 with metric $g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and 2×2 matrix representations of the pseudo-orthogonal group $O(1,1)$ that leaves g invariant.

a) Let $\sigma, \tau = \pm 1$, and $\phi \in \mathbb{R}$. Show that any element of $O(1,1)$ can be written in the form

$$D_{\sigma,\tau}(\phi) = \begin{pmatrix} 1 & 0 \\ 0 & \tau \end{pmatrix} \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} \begin{pmatrix} \sigma & 0 \\ 0 & 1 \end{pmatrix}$$

To study $O(1,1)$ it thus suffices to study the matrices $D(\phi) := D_{+1,+1} = \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix}$.

b) Show explicitly that the set $\{D(\phi)\}$ forms a group under matrix multiplication (which is a subgroup of $O(1,1)$ that is sometimes denoted by $SO^+(1,1)$), and that the mapping $\phi \to D(\phi)$ defines an isomorphism between this group and the group of real numbers under addition.

c) Show that there exists a matrix J (called the *generator* of the subgroup) such that every $D(\phi)$ can be written in the form

$$D(\phi) = e^{J\phi}$$

and determine J explicitly.

(6 points)

1.4.10. **Time-like and space-like intervals**

Consider two points (ct_x, x^1, x^2, x^3) and (ct_y, y^1, y^2, y^3) in Minkowski space. The interval between the two points is called *time-like* if

$$c^2(t_x - t_y)^2 > (x^1 - y^1)^2 + (x^2 - y^2)^2 + (x^3 - y^3)^2 \quad ,$$

and *space-like* if

$$c^2(t_x - t_y)^2 < (x^1 - y^1)^2 + (x^2 - y^2)^2 + (x^3 - y^3)^2 \quad .$$

Show that in interval that is time-like or space-like in some inertial frame is also time-like or space-like in any other inertial frame. (This reflects the invariance of the speed of light.)

(2 points)

1.4.11. **Special Lorentz transformations in** M_4

Consider the Minkowski space M_4.

a) Show that the following transformations are Lorentz transformations:

i) $D^\mu{}_\nu = \begin{pmatrix} 1 & 0 \\ 0 & R^i{}_j \end{pmatrix} \equiv R^\mu{}_\nu$ (rotations)

where $R^i{}_j$ is any Euclidian orthogonal transformation.

ii) $D^\mu{}_\nu = \begin{pmatrix} \cosh\alpha & \sinh\alpha & 0 & 0 \\ \sinh\alpha & \cosh\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \equiv B^\mu{}_\nu$ (Lorentz boost along the x-direction)

with $\alpha \in \mathbb{R}$.

iii) $D^\mu{}_\nu = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \equiv P^\mu{}_\nu \quad \text{(parity)}$

iv) $D^\mu{}_\nu = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \equiv T^\mu{}_\nu \quad \text{(time reversal)}$

b) Let L be the group of all Lorentz transformations. Show that the rotations defined in part a) i) are a subgroup of L, and so are the Lorentz boosts defined in part a) ii).

c) Let $I^\mu{}_\nu = \delta^\mu{}_\nu$ be the identity transformation. Show that the sets $\{I, P\}$, $\{I, T\}$, and $\{I, P, T, PT\}$ are subgroups of L.

(4 points)

1.4.12. **General Lorentz transformations in** M_4

Let D be a general Lorentz transformation in M_4.

a) Show that $|D^0{}_0| \geq 1$, and that $(D^0{}_1)^2 + (D^0{}_2)^2 + (D^0{}_3)^2 = (D^1{}_0)^2 + (D^2{}_0)^2 + (D^3{}_0)^2$.

b) Let $L_{++} = \{D \in L; \det D > 0, D^0{}_0 > 0\}$. (This is called the set of proper orthochronous Lorentz transformations, and one can show that it is a subgroup of L.) Show that any Lorentz transformation can be written as an element of L_{++} followed by either P, or T, or PT. It thus suffices to study L_{++}.

c) Show that any element of L_{++} can be written as a spatial rotation $R(\Phi, \Theta, \Psi)$ followed by a Lorentz boost $B(\alpha)$ followed by a rotation about the 3-axes followed by a rotation about the 2-axis. In a symbolic notation:

$$D = \begin{pmatrix} 1 & 0 \\ 0 & R_2(\phi)R_3(\theta) \end{pmatrix} B(\alpha) \begin{pmatrix} 1 & 0 \\ 0 & R(\Phi, \Theta, \Psi) \end{pmatrix}$$

L_{++} is thus characterized by six parameters: 3 Euler angles Φ, Θ, Ψ, the boost parameter α, and two additional rotation angles ϕ and θ.

(7 points)

5 Tensor Fields

5.1 Tensor Fields

Definition 1. Let us consider the $\mathbb{R}$-vector space $\mathbb{R}^n$ with a generalized metric. Let D be a normal coordinate transformation. A ***tensor field*** is a mapping that assigns each $x \in \mathbb{R}^n$ a rank-N tensor $t^{i_1,i_2,\ldots,i_N}(x)$, which transforms under D in the following way: $\tilde{x} = Dx$, and

$$\tilde{t}^{i_1,i_2,\ldots,i_N}(\tilde{x}) = \left(D^{i_1}{}_{j_1} D^{i_2}{}_{j_2} \cdots D^{i_N}{}_{j_N}\right) t^{j_1,j_2,\ldots,j_N}(x) = \left(\prod_{k=1}^{N} D^{i_k}{}_{j_k}\right) t^{j_1,j_2,\ldots,j_N}(x).$$

Remark 1. The field in **Definition 1** is not the same as the one defined in **3.2**.

Proposition 1. *Homogeneous tensor fields, i.e.,* $\forall\, x \in \mathbb{R}^n, t^{i_1,\ldots,i_N}(x) = t^{i_1,\ldots,i_N}$*, are consistent with* ***Definition 3*** *of* ***4.3****.*

Proof. Let $f : \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{N\ \mathbb{R}^n\text{'s}} \to \mathbb{R}$ be a multilinear form. Let $\{e_i\}_{i=1}^n$ be a normal coordinate system. Let D^{-1} be a normal coordinate transformation $\tilde{e}_i = e_j (D^{-1})^j{}_i$. For any $x_1, \ldots, x_N \in \mathbb{R}^n$, we have

$$\begin{aligned} f(x_1, \ldots, x_N) &= f\Big((x_1)_{j_1} e^{j_1}, \ldots, (x_N)_{j_N} e^{j_N}\Big) = f\Big((\tilde{x}_1)_{i_1} \tilde{e}^{i_1}, \ldots, (\tilde{x}_N)_{i_N} \tilde{e}^{i_N}\Big) \\ &= (x_1)_{j_1} \cdots (x_N)_{j_N}\, f\left(e^{j_1}, \ldots, e^{j_N}\right) = (\tilde{x}_1)_{i_1} \cdots (\tilde{x}_N)_{i_N}\, f\left(\tilde{e}^{i_1}, \ldots, \tilde{e}^{i_N}\right) \\ &= (x_1)_{j_1} \cdots (x_N)_{j_N}\, t^{j_1,\ldots,j_N} = (\tilde{x}_1)_{i_1} \cdots (\tilde{x}_N)_{i_N}\, \tilde{t}^{i_1,\ldots,i_N}. \end{aligned}$$

Let g_{ij} be the metric corresponding to $\{e_i\}_{i=1}^n$ and $\{\tilde{e}_i\}_{i=1}^n$. Now, notice that

$$x_j = g_{ji} x^i = g_{ji} (D^{-1})^i{}_k \tilde{x}^k = \left(g D^{-1}\right)_{jk} \tilde{x}^k = \left(D^T g\right)_{jk} \tilde{x}^k = (D^T)_j{}^i g_{ik} \tilde{x}^k = D^i{}_j \tilde{x}_i,$$

which further implies that

$$(\tilde{x}_1)_{i_1} \cdots (\tilde{x}_N)_{i_N}\, \tilde{t}^{i_1,\ldots,i_N} = (x_1)_{j_1} \cdots (x_N)_{j_N}\, t^{j_1,\ldots,j_N} = D^{i_1}{}_{j_1} \cdots D^{i_N}{}_{j_N} (\tilde{x}_1)_{i_1} \cdots (\tilde{x}_N)_{i_N}\, t^{j_1,\ldots,j_N}.$$

By comparison, we conclude that

$$\tilde{t}^{i_1,\ldots,i_N} = \left(D^{i_1}{}_{j_1} \cdots D^{i_N}{}_{j_N}\right) t^{j_1,\ldots,j_N},$$

as desired. □

Remark 2. **Proposition 1** implies that all tensors must transform in the same way as homogeneous tensor fields under a normal coordinate transformation. As physicists, we often define tensors by means of this transformation property without referring to multilinear forms.

Remark 3. In an n-dimensional vector space, a rank-N tensor can be regarded as a set of n^N scalars $t^{i_1,\ldots,i_N}$ that are associated with a basis and possess the transformation property.

Example 1. A vector x is a rank-1 tensor, because for any arbitrary coordinate transformation D, $\tilde{x}^i = D^i{}_j x^j$.

Example 2. Metric tensors are indeed tensors, since for any coordinate transformation D,

$$\tilde{g}^{ij} = \left(\tilde{g}^{-1}\right)_{ij} = \left(Dg^{-1}D^T\right)_{ij} = D_i{}^k(g^{-1})_{k\ell}(D^T)^\ell{}_j = D_i{}^k D_j{}^\ell (g^{-1})_{k\ell} = D^i{}_k D^j{}_\ell g^{k\ell}.$$

Remark 4. Metric tensors of the form in Eq. (4.8.2) are special, since for any normal coordinate transformation, $\tilde{g} = g$; nevertheless, they still possess the transformation property.

Example 3. Let us apply the criterion to check whether or not the Levi-Civita tensor is a tensor. Let $\{e_i\}_{i=1}^n$ and $\{e^i\}_{i=1}^n$ be a basis and its corresponding cobasis, respectively. Let D^{-1} be a coordinate transformation $\tilde{e}_i = e_j(D^{-1})^j{}_i$. The relation between $\tilde{e}^i$ and e^j is given by $\tilde{e}^i = D^i{}_j e^j$, because for any $x \in \mathbb{R}^n$,

$$\tilde{x}_i \tilde{e}^i = x = x_j e^j = (D^i{}_j \tilde{x}_i)e^j = \tilde{x}_i(D^i{}_j e^j).$$

Now, we are able to compute components of the Levi-Civita tensor in the new cobasis $\{\tilde{e}^i\}_{i=1}^n$:

$$(\tilde{\varepsilon}_L)^{ijk} = \varepsilon(\tilde{e}^i, \tilde{e}^j, \tilde{e}^k) = \varepsilon(D^i{}_\ell e^\ell, D^j{}_m e^m, D^k{}_n e^n) = D^i{}_\ell D^j{}_m D^k{}_n \varepsilon(e^\ell, e^m, e^n) = D^i{}_\ell D^j{}_m D^k{}_n (\varepsilon_L)^{\ell mn},$$

which indicates that the Levi-Civita tensor is indeed a tensor.

Remark 5. The Levi-Civita tensor is undoubtedly a tensor, since it corresponds to a trilinear form. On the other hand, we will later show that the Levi-Civita symbol is not a tensor.

Definition 2. Recall that the ***Levi-Civita symbol*** ε^{ijk} [a] is given by

$$\varepsilon^{ijk} = \operatorname{sgn}\begin{pmatrix} i, j, k \\ 1, 2, 3 \end{pmatrix}.$$

We assign ε^{ijk} to each normal coordinate system in $\mathbb{R}^3$ so that ε^{ijk} is promoted to an entity that is invariant under normal coordinate transformations, i.e., $\tilde{\varepsilon}^{ijk} = \varepsilon^{ijk}$.

[a] In fact, $\varepsilon^{ijk} = -\varepsilon_{ijk} = -\operatorname{sgn}\binom{i,j,k}{1,2,3}$ (See **Example 1** of **4.3**). However, we omit the difference here, because we are more interested in the transformation property of the Levi-Civita symbol.

Remark 6. We would like to once again accentuate the fact that components of the Levi-Civita tensor in an arbitrary cobasis $\{e^i\}_{i=1}^n$ are generally not given by the Levi-Civita symbol, i.e., $\varepsilon(e^i, e^j, e^k) = (\varepsilon_L)^{ijk} \neq \varepsilon^{ijk}$.

Definition 3. A ***rank-N pseudo-tensor*** $t^{i_1,i_2,\dots,i_N}$ transforms under a normal coordinate transformation D in the following way:

$$\tilde{t}^{i_1,i_2,\dots,i_N} = \det D\left(D^{i_1}{}_{j_1} D^{i_2}{}_{j_2} \cdots D^{i_N}{}_{j_N}\right) t^{j_1,j_2,\dots,j_N} = \det D\left(\prod_{k=1}^N D^{i_k}{}_{j_k}\right) t^{j_1,j_2,\dots,j_N}.$$

Lemma 1. *Let $x_1, \dots, x_N \in \mathbb{R}^n$. Let D be a normal coordinate transformation. For any antisymmetric multilinear form f on $\mathbb{R}^n$, we have*

$$f\Big(D^{i_1}{}_{j_1}(x_1)^{j_1}, \dots, D^{i_N}{}_{j_N}(x_N)^{j_N}\Big) = \det D \cdot f\Big((x_1)^{i_1}, \dots, (x_N)^{i_N}\Big).$$

Proof. Please consult **Chapter 4.7** of *Algebra, Vol. I* by B. L. van der Waerden. $\square$

Example 4. Now, we are about to show that the Levi-Civita symbol is a rank-3 pseudo-tensor. Let $\{e^i\}_{i=1}^n$ be the standard cobasis. In $\{e^i\}_{i=1}^n$, we have $(\varepsilon_L)^{ijk} = \varepsilon(e^i, e^j, e^k) = \varepsilon^{ijk}$. Let D be a normal coordinate transformation $\tilde{e}^i = D^i{}_j e^j$. According to **Lemma 1**, components of the Levi-Civita tensor in the new cobasis $\{\tilde{e}^i\}_{i=1}^n$ are given by

$$\begin{aligned}(\tilde{\varepsilon}_L)^{ijk} = \varepsilon(\tilde{e}^i, \tilde{e}^j, \tilde{e}^k) = \varepsilon(D^i{}_\ell e^\ell, D^j{}_m e^m, D^k{}_n e^n) = D^i{}_\ell D^j{}_m D^k{}_n \varepsilon(e^\ell, e^m, e^n) = D^i{}_\ell D^j{}_m D^k{}_n \varepsilon^{\ell mn} \\ = \det D \cdot \varepsilon(e^i, e^j, e^k) = \det D \cdot \varepsilon^{ijk}.\end{aligned}$$

We also have $(\tilde{\varepsilon}_L)^{ijk} = \det D \cdot \varepsilon^{ijk} = \det D \cdot \tilde{\varepsilon}^{ijk}$, because ε^{ijk} is invariant under the normal coordinate transformation D. Recall that $\det D = \pm 1 = \frac{1}{\det D}$. Therefore,

$$\tilde{\varepsilon}^{ijk} = \frac{1}{\det D} D^i{}_\ell D^j{}_m D^k{}_n \varepsilon^{\ell mn} = \det D (D^i{}_\ell D^j{}_m D^k{}_n) \varepsilon^{\ell mn},$$

i.e., ε^{ijk} is a pseudo-tensor.

Remark 7. It is entertaining to show that $D^i{}_\ell D^j{}_m D^k{}_n \varepsilon^{\ell mn} = \det D \cdot \varepsilon^{ijk}$ in the following way:

$$\begin{aligned}D^i{}_\ell D^j{}_m D^k{}_n \varepsilon^{\ell mn} &= \sum_{\ell=1}^3 \sum_{m=1}^3 \sum_{n=1}^3 D^i{}_\ell D^j{}_m D^k{}_n \operatorname{sgn}\begin{pmatrix}\ell, m, n \\ 1, 2, 3\end{pmatrix} \\ &= \sum_{\binom{\ell,m,n}{1,2,3} \in S_3} \left[\operatorname{sgn}\begin{pmatrix}\ell, m, n \\ 1, 2, 3\end{pmatrix} D^i{}_{\binom{\ell,m,n}{1,2,3}(1)} D^j{}_{\binom{\ell,m,n}{1,2,3}(2)} D^k{}_{\binom{\ell,m,n}{1,2,3}(3)}\right] \\ &= \operatorname{sgn}\begin{pmatrix}i, j, k \\ 1, 2, 3\end{pmatrix} \sum_{\binom{\ell,m,n}{1,2,3} \in S_3} \left[\operatorname{sgn}\begin{pmatrix}\ell, m, n \\ 1, 2, 3\end{pmatrix} \prod_{s=1}^3 D^s{}_{\binom{\ell,m,n}{1,2,3}(s)}\right] \\ &= \det D \cdot \varepsilon^{ijk}.\end{aligned}$$

5.2 Gradient, Curl and Divergence

The pedantic typist decides to employ upright boldface letters to represent vectors and matrices in the rest of this section. Components of a vector are denoted in the usual way. For example, $x^i \equiv (\mathbf{x})^i$ are the contravariant components of a vector $\mathbf{x}$.

Definition 1. Let us consider the $\mathbb{R}$-vector space $\mathbb{R}^n$ with a generalized metric. A ***scalar field*** is a mapping $f : \mathbb{R}^n \to \mathbb{R}$ that assigns each $\mathbf{x} \in \mathbb{R}^n$ a scalar $f(\mathbf{x}) \in \mathbb{R}$. Likewise, a ***vector field*** $\mathbf{v} : \mathbb{R}^n \to \mathbb{R}^n$ assigns each $\mathbf{x}$ a vector $\mathbf{v}(\mathbf{x}) \in \mathbb{R}^n$. Let $\mathbf{x} \in \mathbb{R}^n$ be an arbitrary vector.

(i) The ***gradient*** of a scalar field f, $\boldsymbol{\nabla} f$ is a vector field defined by

$$(\boldsymbol{\nabla} f)_i(\mathbf{x}) := \frac{\partial f}{\partial x^i}(\mathbf{x}) \equiv \partial_i f(\mathbf{x}) \quad (i \in \{i\}_{i=1}^n).$$

(ii) The ***curl*** of a vector field $\mathbf{v}$, $\boldsymbol{\nabla} \times \mathbf{v}$ is a vector field defined by

$$(\boldsymbol{\nabla} \times \mathbf{v})^i(\mathbf{x}) := \varepsilon^{ijk} \partial_j v_k(\mathbf{x}) \quad (i \in \{i\}_{i=1}^n).$$

(iii) The ***divergence*** of the vector field $\mathbf{v}$, $\boldsymbol{\nabla} \cdot \mathbf{v}$ is a scalar field defined by

$$(\boldsymbol{\nabla} \cdot \mathbf{v})(\mathbf{x}) = \partial_i v^i(\mathbf{x}).$$

Proposition 1.
For any $\mathbf{x} \in \mathbb{R}^n$*,*

(i) the gradient of a scalar field at $\mathbf{x}$ *transforms in the same way as a covariant vector under a normal coordinate transformation;*

(ii) the curl of a vector field at $\mathbf{x}$ *transforms as a pseudo-vector.*

What is more,

(iii) the divergence of a vector field is indeed a scalar field and transforms as a scalar at each $\mathbf{x} \in \mathbb{R}^n$*.*

Proof. See **Problem 1.5.2**.
Hint. Let $\{\mathbf{e}_i\}_{i=1}^n$ be a basis. Let $\mathbf{D}^{-1}$ be a normal coordinate transformation $\tilde{\mathbf{e}}_i = \mathbf{e}_j(\mathbf{D}^{-1})^j{}_i$. Recall that $x^i = (\mathbf{D}^{-1})^i{}_j\tilde{x}^j$, which implies that

$$(\mathbf{D}^{-1})^i{}_j = \frac{\partial x^i}{\partial \tilde{x}^j}.$$

For (i), let $f(\mathbf{x})$ be a scalar field. Applying the chain rule, one can easily show that

$$(\tilde{\boldsymbol{\nabla}}\tilde{f})_i(\tilde{\mathbf{x}}) = \frac{\partial f}{\partial \tilde{x}^i}(\mathbf{x}) = (\mathbf{D}^{-1})^j{}_i\frac{\partial f}{\partial x^j}(\mathbf{x}).$$

To complete the proof, one just needs to establish that any covariant vector y_i transforms under $\mathbf{D}^{-1}$ in the following way:

$$\tilde{y}_i = (\mathbf{D}^{-1})^j{}_i y_j.$$

□

Remark 1. Let $\mathbf{x} \in \mathbb{R}^n$ be given. The contravariant components of the gradient of a scalar field f at $\mathbf{x}$ can be defined by

$$(\boldsymbol{\nabla} f)^i(\mathbf{x}) \coloneqq \frac{\partial f}{\partial x_i}(\mathbf{x}) \equiv \partial^i f(\mathbf{x}) \quad (i \in \{i\}_{i=1}^n).$$

The reader ought to verify that it does transform as a contravariant vector.

5.3 Tensor Products and Tensor Traces

Definition 1. Let s and t be tensors of rank M and rank N, respectively. The ***tensor product*** of s and t yields a rank-$(M+N)$ tensor $u = s \otimes t$ whose components are given by

$$u^{i_1,\ldots,i_{M+N}} = s^{i_1,\ldots,i_M}t^{i_{M+1},\ldots,i_{M+N}}.$$

Proposition 1. *The tensor product of two tensors or two pseudo-tensors is a tensor. The tensor product of one tensor and one pseudo-tensor is a pseudo-tensor.*

Proof. See **Problem 1.5.3**. □

Definition 2. Let $t^{i_1,\ldots,i_{N+2}}$ be a rank-$(N+2)$ tensor (or pseudo-tensor). The $(1,2)$-***trace*** or $(1,2)$-***contraction*** [a] of t is defined as the rank-N tensor (or pseudo-tensor) u with components

$$u^{i_1,\ldots,i_N} := g_{jk}t^{j,k,i_1,\ldots,i_N} = t_k{}^{k,i_1,\ldots,i_N}.$$

[a] As the name suggests, components of t are summed over the first two indices.

Proposition 2. *Such defined u is indeed a tensor (or pseudo-tensor).*

Proof. See **Problem 1.5.3**. □

Example 1. Let $\mathbf{x} \in \mathbb{R}^n$ be given. The curl of a vector field $\mathbf{v}$ at $\mathbf{x}$, $(\boldsymbol{\nabla} \times \mathbf{v})(\mathbf{x})$ can be regarded as successive contractions of a rank-5 pseudo-tensor:

$$(\boldsymbol{\nabla} \times \mathbf{v})^i(\mathbf{x}) = \varepsilon^{ijk}\partial_j v_k(\mathbf{x}) = g_{j\ell}\varepsilon^{ijk}\partial^\ell v_k(\mathbf{x}) = g_{km}g_{j\ell}\varepsilon^{ijk}\partial^\ell v^m(\mathbf{x}).$$

According to **Proposition 2**, the curl is a pseudo-vector. This is consistent with **Proposition 1**, (ii) of **5.2**.

5.4 Minkowski Tensors

Let us consider M^4, i.e., $\mathbb{R}^4$ with the metric $\mathbf{g} = \operatorname{diag}\{1,-1,-1,-1\}$. Let $\{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be a basis. Let $A \in M^4$ be a four-vector with contravariant components $A^\mu = (A^0, A^1, A^2, A^3) \equiv (A^0, \mathbf{A})$ and covariant components $(A_0, A_1, A_2, A_3) = A_\mu = g_{\mu\nu}A^\nu = (A^0, -A^1, -A^2, -A^3) \equiv (A^0, -\mathbf{A})$. Thereinto, $\mathbf{A}$ can be treated as a ***three-vector*** in the 3-dimensional Euclidean space $E^3 \subset M^4$, which is spanned by the basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Let F be the rank-2 tensor

$$F^{\mu\nu} = \left(\begin{array}{c:ccc} F^{00} & F^{01} & F^{02} & F^{03} \\ \hdashline F^{10} & F^{11} & F^{12} & F^{13} \\ F^{20} & F^{21} & F^{22} & F^{23} \\ F^{30} & F^{31} & F^{32} & F^{33} \end{array}\right) = \left(\begin{array}{c:c} F^{00} & \mathbf{F}_{\text{hor}} \\ \hdashline \mathbf{F}_{\text{ver}} & F^{ij} \end{array}\right).$$

Like $\mathbf{A}$, $\mathbf{F}_{\text{hor}}$ and $\mathbf{F}_{\text{ver}}$ can also be regarded as three-vectors; F^{ij} can be considered as a rank-2 tensor in E^3.

Example 1. In electromagnetism, the ***field-strength tensor*** is given by $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, where

$$\partial^\mu := \frac{\partial}{\partial x_\mu} = \left(\frac{\partial}{\partial x_0}, \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right) = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\boldsymbol{\nabla}\right).$$

Remark 1. Greek indices (running from 0 to 3) are employed to label both temporal and spatial components of four-vectors. Latin indices (running from 1 to 3) are used to label only the spatial component.

Remark 2. If F is symmetric, i.e., $F^{\mu\nu} = F^{\nu\mu}$, we have $\mathbf{F}_{\text{hor}} = \mathbf{F}_{\text{ver}}$ and $F^{ij} = F^{ji}$.

Remark 3. If F is antisymmetric, i.e., $F^{\mu\nu} = -F^{\nu\mu}$, then $\mathbf{F}_{\text{hor}} = -\mathbf{F}_{\text{ver}}$, $F^{ij} = -F^{ji}$, and $F^{\mu\mu} = 0$.

Lemma 1. *In E^3, the set of antisymmetric rank-2 tensors is isomorphic to the set of pseudo-vectors.*

Proof. Let t be an arbitrary antisymmetric rank-2 tensor in E^3. t can be written as

$$t^{ij} = \begin{pmatrix} 0 & v_3 & -v_2 \\ -v_3 & 0 & v_1 \\ v_2 & -v_1 & 0 \end{pmatrix} = \varepsilon^{ijk} v_k,$$

for some $\mathbf{v} \in E^3$. According to **Proposition 1** of **5.3**, $\mathbf{v}$ is a pseudo-vector. We have thus shown that in E^3, there exists a one-to-one correspondence between antisymmetric rank-2 tensors and pseudo-vectors. ☐

Corollary 1. *In M^4, any antisymmetric rank-2 tensor is of the form*

$$\left(\begin{array}{c:ccc} 0 & a_1 & a_2 & a_3 \\ \hdashline -a_1 & 0 & v_3 & -v_2 \\ -a_2 & -v_3 & 0 & v_1 \\ -a_3 & v_2 & -v_1 & 0 \end{array}\right) = \left(\begin{array}{c:c} 0 & \mathbf{a} \\ \hdashline -\mathbf{a}^T & t^{ij} \end{array}\right),$$

for some three-vector $\mathbf{a}$ and pseudo-three-vector $\mathbf{v}$.

Remark 4. Let

$$F^{\mu\nu} = \left(\begin{array}{c:c} 0 & \mathbf{a} \\ \hdashline -\mathbf{a}^T & t^{ij} \end{array}\right).$$

Let us first derive the mixed tensors ${F_\mu}^\nu$ and ${F^\mu}_\nu$:

$${F_\mu}^\nu = g_{\mu\alpha} F^{\alpha\nu} = \left(\begin{array}{c:c} 1 & \mathbf{0} \\ \hdashline \mathbf{0}^T & -\mathbb{1}_3 \end{array}\right) \left(\begin{array}{c:c} 0 & \mathbf{a} \\ \hdashline -\mathbf{a}^T & t^{ij} \end{array}\right) = \left(\begin{array}{c:c} 0 & \mathbf{a} \\ \hdashline \mathbf{a}^T & -t^{ij} \end{array}\right),$$

and

$${F^\mu}_\nu = F^{\mu\alpha} g_{\alpha\nu} = \left(\begin{array}{c:c} 0 & \mathbf{a} \\ \hdashline -\mathbf{a}^T & t^{ij} \end{array}\right) \left(\begin{array}{c:c} 1 & \mathbf{0} \\ \hdashline \mathbf{0}^T & -\mathbb{1}_3 \end{array}\right) = \left(\begin{array}{c:c} 0 & -\mathbf{a} \\ \hdashline -\mathbf{a}^T & -t^{ij} \end{array}\right).$$

We can further obtain $F_{\mu\nu}$:

$$F_{\mu\nu} = g_{\mu\alpha} g_{\nu\beta} F^{\alpha\beta} = {F_\mu}^\beta g_{\beta\nu} = \left(\begin{array}{c:c} 0 & \mathbf{a} \\ \hdashline \mathbf{a}^T & -t^{ij} \end{array}\right) \left(\begin{array}{c:c} 1 & \mathbf{0} \\ \hdashline \mathbf{0}^T & -\mathbb{1}_3 \end{array}\right) = \left(\begin{array}{c:c} 0 & -\mathbf{a} \\ \hdashline \mathbf{a}^T & t^{ij} \end{array}\right) = \left(\begin{array}{c:c} 0 & -\mathbf{a} \\ \hdashline \mathbf{a}^T & t_{ij} \end{array}\right).$$

Notice that

$$\begin{aligned} F^{\mu\nu} F_{\mu\nu} &= -F^{\nu\mu} F_{\mu\nu} = -(FF)^\nu{}_\nu = -\operatorname{Tr}\{FF\} \\ &= -\operatorname{Tr}\left\{\left(\begin{array}{c:c} 0 & \mathbf{a} \\ \hdashline -\mathbf{a}^T & \mathbf{t} \end{array}\right) \left(\begin{array}{c:c} 0 & -\mathbf{a} \\ \hdashline \mathbf{a}^T & \mathbf{t} \end{array}\right)\right\} = -\operatorname{Tr}\left\{\left(\begin{array}{c:c} \mathbf{a}\mathbf{a}^T & \mathbf{a}\mathbf{t} \\ \hdashline \mathbf{t}\mathbf{a}^T & \mathbf{a}^T\mathbf{a} + \mathbf{t}\mathbf{t} \end{array}\right)\right\} \\ &= -\|\mathbf{a}\|^2 - \operatorname{Tr}\left\{\mathbf{a}^T\mathbf{a} + \mathbf{t}\mathbf{t}\right\} = 2\left(\|\mathbf{v}\|^2 - \|\mathbf{a}\|^2\right), \end{aligned}$$

i.e., $F^{\mu\nu} F_{\mu\nu}$ is a scalar.

5.5 Problems

1.5.1. **Transformations of tensor fields**

a) Consider a covariant rank-n tensor field $t_{i_1 \ldots i_n}(x)$ and find its transformation law under normal coordinate transformations that is analogous to §5.1 def.1; i.e., find how $\tilde{t}_{i_1 \ldots i_n}(\tilde{x})$ is related to $t_{i_1 \ldots i_n}(x)$.

b) Convince yourself that your result is consistent with the transformation properties of (i) a covector x_i (the case $n = 1$), and (ii) the covariant components of the metric tensor g_{ij}.

(4 points)

1.5.2. **Curl and divergence**

Show that the curl and the divergence of a vector field transform as a pseudovector field and a scalar field, respectively.

(3 points)

1.5.3. **Tensor products, and tensor traces**

Prove Propositions 1 and 2 from ch. 1 §5.3.

(4 points)

Chapter 2

Topics in Analysis

1 Reminder: Real Analysis

Note: This paragraph summarizes some material that is covered in typical courses on calculus (including multivariate calculus). At the UO that would be MATH 251-3 plus MATH 281,2.

1.1 Differentiation and integration

Consider mappings (called "functions" in this context) $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$. We say that $\vec{f}$ is an m-vector-valued function of n real variables and write

$$\vec{f}(\vec{x}) \equiv \vec{f}(x_1, \ldots, x_n) = \vec{y} \qquad , \qquad \vec{x} = (x^1, \ldots, x^n) \in \mathbb{R}^n$$

$$\vec{y} = (y^1, \ldots, y^n) \in \mathbb{R}^m$$

For $m = 1$ we write f instead of $\vec{f}$.

Definition 1.
(a) For $n = m = 1$ we define the ***derivative*** of f, $f' \equiv df/dx : \mathbb{R} \to \mathbb{R}$ by

$$f'(x) := \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[f(x + \epsilon) - f(x) \right] \qquad (*)$$

and higher derivatives by $d^2 f/dx^2 := \frac{d}{dx} f'$, etc.

(b) For $n > 1$, $m = 1$ we define ***partial derivatives*** $\partial f/\partial x^i \equiv \partial_i f : \mathbb{R}^n \to \mathbb{R}$ by $(*)$ applied to the argument x^i, and the ***gradient*** of f, $\partial f/\partial \vec{x} \equiv \vec{\nabla} f : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\vec{\nabla} f(\vec{x}) := (\partial_1 f(\vec{x}), \ldots, \partial_n f(\vec{x}))$$

(c) For $n = 1$, $m > 1$ we define $d\vec{f}/dx : \mathbb{R} \to \mathbb{R}^m$ by

$$\frac{d\vec{f}}{dx} := \left(\frac{df_1}{dx}, \ldots, \frac{df_m}{dx} \right)$$

(d) For $n = m$ we define the ***divergence*** $\text{div}\vec{f} \equiv \vec{\nabla} \cdot \vec{f} : \mathbb{R}^n \to \mathbb{R}$ by

$$\vec{\nabla} \cdot \vec{f}(\vec{x}) := \partial_i f^i(\vec{x})$$

(e) For $n = m = 3$ we define the ***curl*** $\text{curl}\vec{f} \equiv \vec{\nabla} \times \vec{f} : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$\left(\vec{\nabla} \times \vec{f}(\vec{x})\right)^i := \epsilon^{ij}{}_k \partial_j f^k(\vec{x})$$

Remark 1. If the space is Euclidian, then $\partial_i = \partial^i$, and $\epsilon^{ij}{}_k = \epsilon_{ijk}$.

Definition 2. Let $I = [t_0, t_1] \subset \mathbb{R}$ and $\vec{x} : I \to \mathbb{R}^n$ a function of t. Let $f : \mathbb{R}^n \times I \to \mathbb{R}$ by a real-valued function of $\vec{x}$ and t. Then we define the ***total derivative*** of f with respect to t, $df/dt : I \to \mathbb{R}$ by

$$\frac{df}{dt}(t^*) \equiv \left.\frac{df}{dt}\right|_{t=t^*} := \partial_t f(\vec{x}(t^*), t^*) + \partial_i f(\vec{x}(t^*), t^*) \frac{dx^i}{dt}(t^*)$$

Proposition 1. *Taylor expansion*
Let $f : \mathbb{R}^n \to \mathbb{R}$ be m times differentiable at $\vec{x}$. Then there exists a neighborhood of $\vec{x}$ where f can be represented by a power series

$$f(x_1 + \epsilon, x_2, \ldots, x_n) = f(x_1, \ldots, x_n) + \epsilon \frac{\partial f}{\partial x_1}(x_1, \ldots, x_n) + \ldots + \frac{1}{m!}\,\epsilon^m \frac{\partial^m f}{\partial x_1^m}(x_1, \ldots x_n) + r_m$$

and analogously for other variables.

Proof. Analysis course. □

Remark 2. Taylor's theorem gives an explicit upper bound for the remainder r_m.

Definition 3. Let $f : I \to \mathbb{R}$ be a real-valued function of $t \in I = [t_-, t_+] \subset \mathbb{R}$. Then the ***Riemann integral***

$$F = \int_{t_-}^{t_+} dt\, f(t) := \lim_{N\to\infty} \sum_{i=1}^{N-1} f(t_i)(t_{i+1} - t_i)$$

with $t_1 = t_-$, $t_N = t_+$ is defined as the limit of a sum as indicated, provided the limit exists.

Remark 3. The generalization fo $f : I_1 \times i_1 \to \mathbb{R}$, $F = \int_{t_-}^{t_+} dt \int_{u_-}^{u_+} du\, f(t, u)$ is straightforward.

Remark 4. F is a special case of a ***functional***, i.e., a mapping that maps functions onto numbers.

Remark 5. A geometric interpretation of F is the area under the function f, see Fig. 2.1.1.

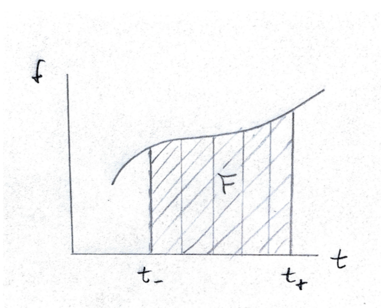


Fig. 2.1.1. Geometric interpretation of the Riemann integral.

1.2 Paths, and line integrals

Definition 1. (a) Let $I = [t_-, t_+] \subset \mathbb{R}$ and let $\vec{q} : I \to \mathbb{R}^n$ be continuously differentiable. Then the set $\mathcal{C} := \{\vec{q}(t), t \in I\} \subset \mathbb{R}^n$ is called a **path** or **curve** in $\mathbb{R}^n$, and $\vec{q}(t)$ is called a **parametrization** of $\mathcal{C}$ with **parameter** t.
(b) $\mathcal{C}$ inherits an order from the greater/lesser order defined on I : $\vec{q}(t_1) < \vec{q}(t_2)$ by definition if and only if $t_1 < t_2$
(c) The **tangent vector** $\vec{\tau}(t)$ in the point $\vec{q}(t)$ is defined as

$$\vec{\tau}(t) := \frac{d}{dt}\vec{q}(t) \equiv \dot{\vec{q}}(t)$$

Definition 2. Let $L : \mathbb{R}^n \times \mathbb{R}^n \times I \to \mathbb{R}$ be a function of $\vec{q}, \dot{\vec{q}}$, and t that is sufficiently well behaved with respect to all arguments. Let $\vec{q}(t)$ be a parametrization of the path $\mathcal{C}$ and define a functional

$$\mathcal{S}_L(\mathcal{C}) := \int_{t_-}^{t_+} dt\, L(\vec{(q)}(t), \dot{\vec{q}}(t), t)$$

Then for a given L, $\mathcal{S}_L(\mathcal{C})$ is characteristic of $\mathcal{C}$

Example 1. $n = 2$. Length of path $\mathcal{C}$ with parameterization $\vec{q}(t)$:

$$\begin{aligned} \ell_{\mathcal{C}} &= \lim_{N\to\infty} \sum_{i=1}^{N-1} \sqrt{\left(\vec{q}(t_{i+1}) - \vec{q}(t_i)\right)^2} \\ &= \lim_{N\to\infty} \sum_{i=1}^{N-1} (t_{i+1} - t_i) \sqrt{\frac{\left(\vec{q}(t_{i+1} - \vec{q}(t_i)\right)^2}{(t_{i+1} - t_i)^2}} \\ &= \int_{t_-}^{t_+} dt\, \sqrt{\vec{\tau}^{\,2}(t)} \end{aligned}$$

by 1.1. defn. 3.
$\Rightarrow$ The choice $L(\vec{q}, \dot{\vec{q}}, t) = \sqrt{\dot{\vec{q}}^2}$ yields $\mathcal{S}_L(\mathcal{C}) = \ell_{\mathcal{C}}$.

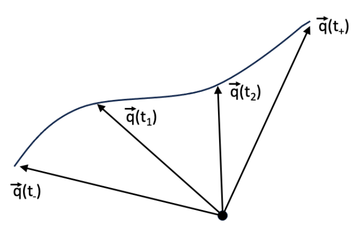


Fig. 2.1.2. A path in $\mathbb{R}^2$.

Example 2. With L the Lagrangian of a mechanical system , $\mathcal{S}_L$ is the action (see PHYS 611).

Definition 3. Let $\vec{f} : \mathbb{R}^n \to \mathbb{R}^n$ be a function, and $\mathcal{C}$ a path in $\mathbb{R}^n$ with parameterization $\vec{q}(t)$. Then the ***line integral*** of $\vec{f}$ over $\mathcal{C}$ is defined as

$$\int_{\mathcal{C}} d\vec{l} \cdot \vec{f} := \int_{t_-}^{t_+} dr\, \vec{\tau}(t) \cdot \vec{f}(\vec{q}(t))$$

with $d\vec{l} := \vec{\tau}(t)dt$ the ***integration measure***.

Remark 1. Integration along a closed curve is denoted by $\oint_{\mathcal{C}} d\vec{\ell}$

1.3 Surfaces, and surface integrals

Definition 1. (a) Let $I_t = [t_-, t_+] \in \mathbb{R}$ and $I_u = [u_-, u_+] \in \mathbb{R}$ be intervals, and let $\vec{r} : I_t \times I_u \to \mathbb{R}^3$ be a continuously differentiable function of t and u. Then

$$S = \{\vec{r}(t,u); t, u \in I_t \times I_u\}$$

is called a **surface** in $\mathbb{R}^3$ with **parameterization** $\vec{r}(t,u)$.
(b) The **standard normal vector** of S in the point $\vec{r}(t,u)$ is defined as

$$\vec{n}(t,u) := \partial_t \vec{r}(t,u) \times \partial_u \vec{r}(t,u)$$

Example 1. Let $I_\varphi = [0, 2\pi]$, $I_\theta = [0, \pi]$. Then

$$\vec{r}(\varphi, \theta) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$$

parameterizes a spherical surface in $\mathbb{R}^3$. The standard normal vector is

$$\vec{n}(\varphi, \theta) = (-\sin^2\theta\cos\varphi, -\sin^2\theta\sin\varphi, -\sin\theta\cos\theta)$$

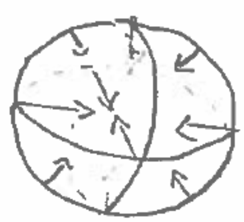


Fig. 2.1.3. Spherical surface with standard normal vectors.

Definition 2. (a) Let $\vec{f} : \mathbb{R}^3 \to \mathbb{R}^3$ be a function and S a surface in $\mathbb{R}^3$ with parameterization $\vec{r}(t,u)$ and standard normal vector $\vec{n}(t,u)$. Then the **surface integral** of $\vec{f}$ over S is defined as

$$\int_S d\vec{\sigma} \cdot \vec{f} := \int_{t_-}^{t_+} dt \int_{u_-}^{u_+} du\, \vec{n}(t,u) \cdot \vec{f}(t,u)$$

with $d\vec{\sigma} := \vec{n}(t,u)\, dt\, du$ the integration measure.

(b) The **area** of S is defined as

$$A(S) := \int_{t_-}^{t_+} dt \int_{u_-}^{u_+} du\, |\vec{n}(t,u)|$$

Example 2. Calculate the surface area of the sphere from **Example 1**: $|\vec{n}(\theta,\varphi)| = |\sin\theta|$
$\Rightarrow \quad A = \int_0^{2\pi} d\varphi \int_0^{\pi} d\theta\, |\sin\theta| = 2\pi \int_{-1}^{1} d\cos\theta = 4\pi$

Example 3. A flat surface parameterized by the cartesian coordinates of its points: $\vec{r}(x,y,z) = (x,y,0)$.
The standard normal vector is

$$\vec{n}(x,y) = \begin{pmatrix}1\\0\\0\end{pmatrix} \times \begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}0\\0\\1\end{pmatrix}$$

and hence the area is

$$A(S) = \int_S dx\, dy$$

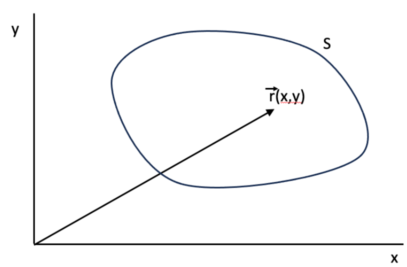


Fig. 2.1.4. A flat surface.

Theorem 1. *(Gauss) Let $V \subset \mathbb{R}^3$ be a volume with surface (V) and let $\vec{f} : \mathbb{R}^3 \to \mathbb{R}^3$ be a function. Then*

$$\int_V dV\, \vec{\nabla} \cdot \vec{f} = \int_{(V)} d\vec{\sigma} \cdot \vec{f}$$

with $dV = dx\, dy\, dz$ the measure of the volume integral.

Proof. See MATH 281,2 or equivalent. □

Theorem 2. *(Stokes) Let S be a surface in $\mathbb{R}^3$ bounded by a curve (S). Then*

$$\int_S d\vec{\sigma} \cdot (\vec{\nabla} \times \vec{f}) = \oint_{(S)} d\vec{\ell} \cdot \vec{f}$$

Proof. See MATH 281,2 or equivalent. □

2 Complex-valued functions of complex arguments

Consider the field $\mathbb{C}$ of complex numbers $z = z_1 + iz_2 \equiv z' + iz''$ (z_1, z_2, z', $z'' \in \mathbb{R}$) as constructed in Ch. 1 §3.3.

2.1 Complex functions

Definition 1.
(a) Let $f : \mathbb{C} \to \mathbb{C}$ be a mapping in the sense of ch.1 §1.2. Then we call f a (single-valued) **complex valued function** of a complex argument.
(b) Generalize the concept of a mapping such that each pre-image can have $n \in \mathbb{N}$ images. Then we call f a (n-valued) **function**.

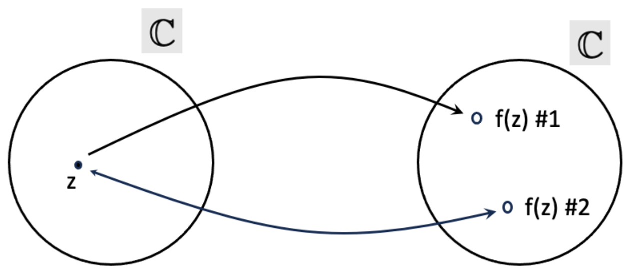


Fig. 2.2.1. A 2-valued function.

Example 1. $f(z) = z^2$ is a single-valued function.

Example 2. $f(z) = z^{1/2}$ is a two-valued function.

Example 3. $f(z) = e^z := \sum_{n=0}^{\infty} \frac{1}{n!} z^n$ is a single valued function.

Example 4. $z = re^{i\varphi} \equiv re^{i(\varphi+2\pi n)}, n \in \mathbb{Z}$
$\ln z = \ln(re^{i(\varphi+2\pi n)}) = \ln r + \ln e^{i(\varphi+2\pi n)} = \ln r + i(\varphi + 2\pi n)$
is a $\mathbb{Z}$-valued function.

Definition 2.
A multi-valued function $f(z)$ can be made single-valued by cutting the complex plane along a **branch cut** chosen such that the function remains single-valued if the place is not crossed.

Example 5. Make $f(z) = z^{1/2}$ single-valued by choosing the branch cut along the negative real axis. Then $i^{1/2} = e^{i\pi/4}$ uniquely, etc.

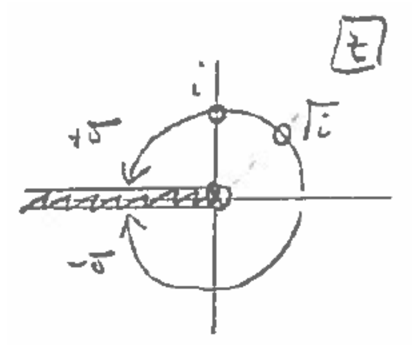


Fig. 2.2.2. The square root with the branch cut along the negative real axis.

Example 6. $f(z) = \ln z$ can be made single-valued by choosing the same branch cut.

Remark 1. The branch cut is property of the individual function, **not** of the complex plane in general.

Remark 2. For a given function, the choice of the branch cut is not unique. For instance, choosing the cut in **Example 5** and **Example 6** along the positive real axis corresponds to $\varphi \in [0, 2\pi[$.

Remark 3. For functions of the form $f(g(z))$, the branch cut will start at a **"branch point"** z_0 determined by $g(z_0) = 0$, rather than at the origin.

Example 7. $f(z) = \ln(z - 1)$

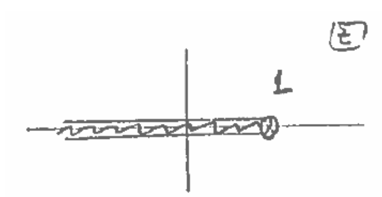


Fig. 2.2.3. The branch cut for $\ln(1 - z)$ chosen along the negative real axis.

Example 8.

$$f(z) = \ln \frac{z-1}{z+1} = \ln(z-1) - \ln(z+1)$$

The branch cuts cancel each other for $[-\infty, -1[$ (See **Problem 2.2.1**.)

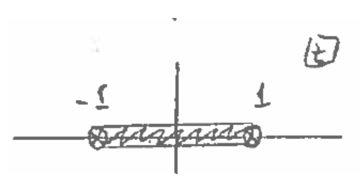


Fig. 2.2.4. The branch cut for $\ln(z - 1) - \ln(z + 1)$ chosen along the real axis.

Definition 3. For a two-valued function, we can continue the function across the cut onto a second **sheet**, on which the function takes on the other possible value. The two sheets will cover the entire complex plane 2-fold and form the **Riemann surface** for the function. An analogous construction works for n-valued functions.

Example 9. $f(z) = z^{1/2}$ Continuation on either sheet past the cut brings one back to the other sheet:

```
In[1]:= Plot3D[Im[Sqrt[x + I y]], {x, -2, 2}, {y, -2, 2}, PlotPoints → 30]
```

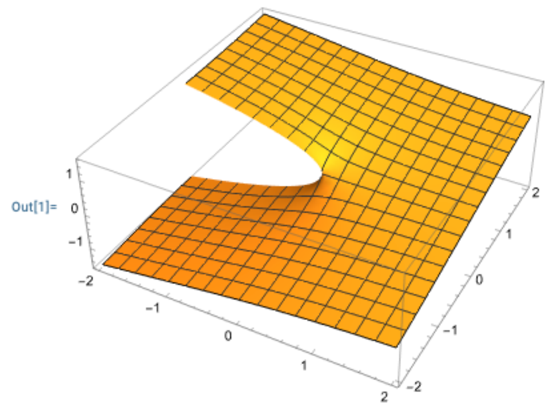


Fig. 2.2.5. The complex plane with a cut for the square root.

```
In[2]:= ParametricPlot3D[{r Cos[phi], r Sin[phi], Sqrt[r] Sin[phi / 2]}, {r, 0, 1},
          {phi, 0, 4 Pi}, PlotPoints → {20, 60}]
```

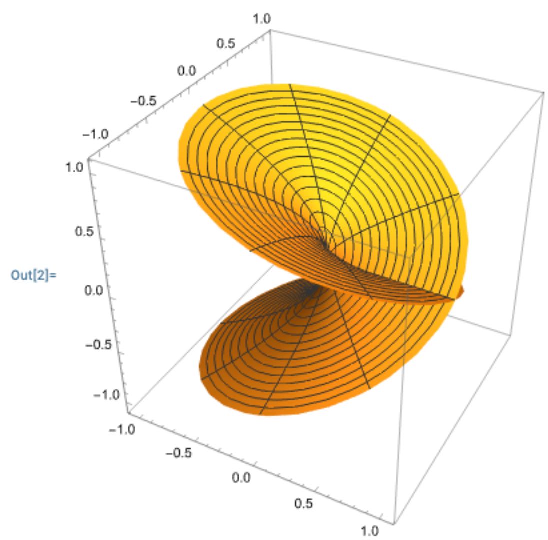


Fig. 2.2.6. The Riemann surface for the square root.

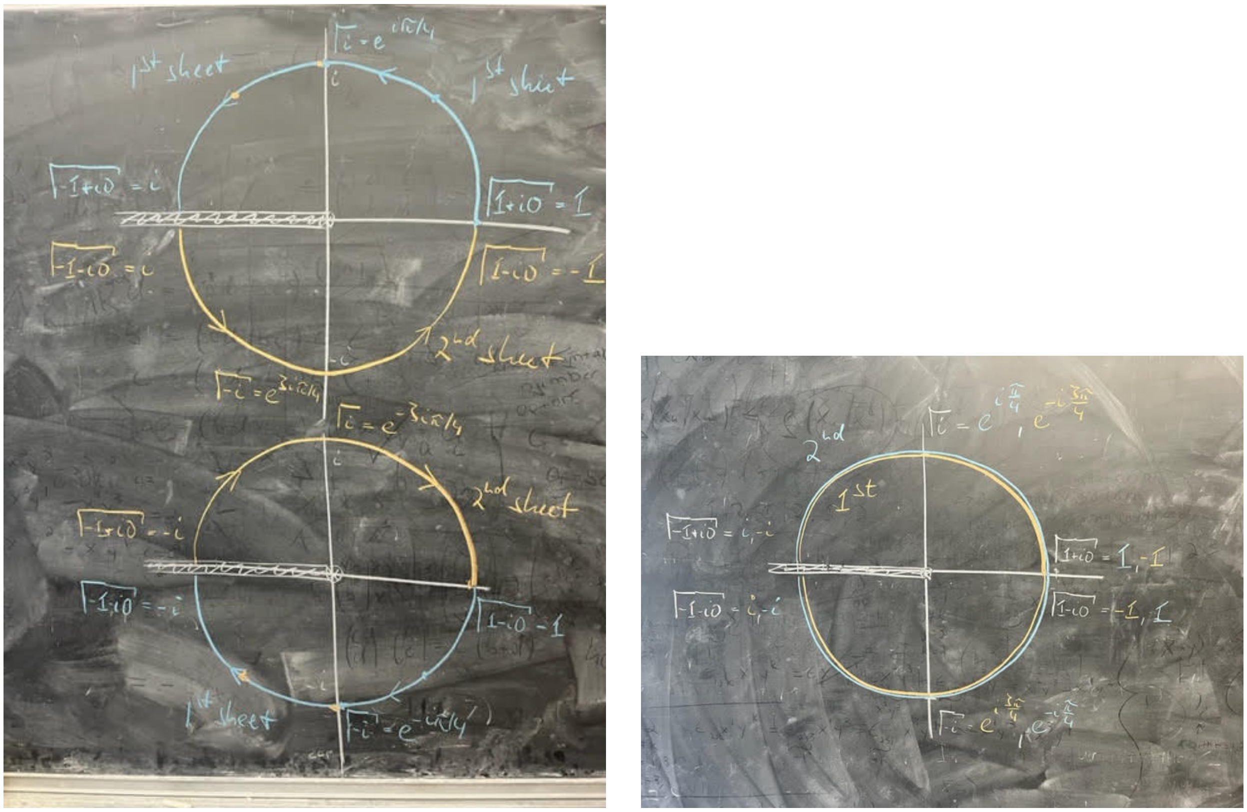


Figure 2.11: Fig. 2.2.7. A low-tech version of Fig. 2.2.6. It is very instructive to construct such a sketch yourself.

2.2 Holomorphy (aka Analyticity)

Definition 1. $f(z)$ is called **continuous** in $z_0 \in \mathbb{C}$ if $f(z_0)$ exists and $\lim_{z\to z_0} f(z) = f(z_0)$.

Definition 2. $f(z)$ is called **differentiable** in z_0 with derivative $\left.\frac{df}{dz}\right|_{z_0} := \lim_{z\to z_0} \frac{f(z)-f(z_0)}{z-z_0}$ provided the limit exists.

Remark 1. This is an obvious generalization of the corresponding concept for real function.

Remark 2. The limit must exist no matter how z approaches z_0 in the complex plane!

Remark 3. Continuity and differentiability are much stronger requirements than the corresponding ones in real analysis!

Definition 3. Let $\Omega \subset \mathbb{C}$ be a region in $\mathbb{C}$ and let $f : \Omega \to \mathbb{C}$ be a function. f is called **holomorphic** (or **analytic**) on Ω if it is differentiable in all points $z_0 \in \Omega$

Theorem 1. ***(Cauchy-Riemann):***
$f(z) = f'(z', z'') + if''(z', z''), z = z' + iz''$ is holomorphic on Ω iff

$$\frac{\partial f'}{\partial z'} = \frac{\partial f''}{\partial z''} \quad \textit{and} \quad \frac{\partial f'}{\partial z''} = -\frac{\partial f''}{\partial z'} \tag{2.1}$$

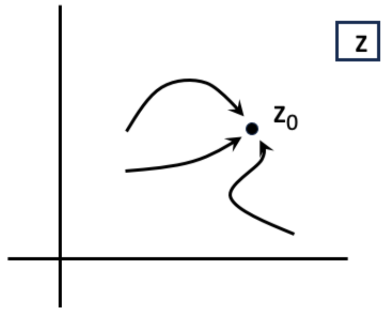


Fig. 2.2.8. Differentiability in $\mathbb{C}$ is a strong requirement.

for all $z \in \Omega$

Proof. See **Problem 2.2.3**. □

Example 1. $f(z) = z^2 = (z'^2 - z''^2) + i2z'z''$
Note that $f' = z'^2 - z''^2$ and $f'' = 2z'z''$

$$\frac{\partial f'}{\partial z'} = 2z' = \frac{\partial f''}{\partial z''}$$

$$\frac{\partial f'}{\partial z''} = -2z'' = -\frac{\partial f''}{\partial z'}$$

$\Rightarrow f$ is holomorphic on $\mathbb{C}$

Example 2.

$$f(z) = \frac{1}{z} = \frac{z'}{z'^2 + z''^2} - i\frac{z''}{z'^2 + z''^2} \Rightarrow f' = \frac{z'}{|z|^2}\ ,\ f'' = \frac{z''}{|z|^2}$$

$$\frac{\partial f'}{\partial z'} = \boxed{\frac{|z|^2 - 2z'^2}{|z|^4}}\quad ,\quad \frac{\partial f''}{\partial z''} = -\frac{|z|^2 - 2z''^2}{|z|^4} = \frac{-z'^2 + z''^2}{|z|^4} = \boxed{\frac{|z|^2 - 2z'^2}{|z|^4}}$$

$$\frac{\partial f'}{\partial z''} = \boxed{\frac{-2z'z''}{|z|^4}}\quad ,\quad \frac{\partial f''}{\partial z'} = -\frac{-2z'z''}{|z|^4} = \boxed{\frac{2z'z''}{|z|^4}}$$

$\Rightarrow f$ is holomorphic on $\mathbb{C} \setminus \{0\}$

Example 3. $f(z) = 1/(z - z_0)^n$ with $n \in \mathbb{N}$ is holomorphic on $\mathbb{C} \setminus \{z_0\}$

Corollary 1. *If* $f : \Omega \to \mathbb{C}$ *is holomorphic, then* f' *and* f'' *satisfy* ***Laplace's differential equation***

$$\frac{\partial^2 \varphi}{\partial z'^2} + \frac{\partial^2 \varphi}{\partial z''^2} = 0, \quad \varphi = f', f''$$

everywhere in Ω

Proof.

$$\begin{aligned}\text{Cauchy-Riemann} \quad &\Rightarrow \quad \frac{\partial^2 f'}{\partial z'^2} = \frac{\partial^2 f''}{\partial z''\,\partial z'} = -\frac{\partial^2 f'}{\partial z''^2} \\ \text{and} \quad & \quad \frac{\partial^2 f''}{\partial z'^2} = -\frac{\partial^2 f'}{\partial z''\,\partial z'} = -\frac{\partial^2 f''}{\partial z''^2}\end{aligned}$$

□

Remark 4. Here we assumed that the second derivatives exist. One can show that they do, and so do all higher derivatives! This is an example of how strong a condition holomorphism is.

2.3 Problems

2.2.1. **Lindhard function**

Consider the function $f : \mathbb{C} \to \mathbb{C}$ (which plays an important role in the theory of many-electron systems) defined by

$$f(z) = \log\left(\frac{z-1}{z+1}\right)$$

The *spectrum* $f'' : \mathbb{R} \to \mathbb{R}$ and the *reactive part* $f' : \mathbb{R} \to \mathbb{R}$ of f are defined by

$$f''(\omega) := \frac{1}{2i}\left[f(\omega + i0) - f(\omega - i0)\right] \qquad , \qquad f'(\omega) := \frac{1}{2}\left[f(\omega + i0) + f(\omega - i0)\right]$$

where $f(\omega \pm i0) := \lim_{\epsilon\to 0} f(\omega \pm i\epsilon)$.

a) Show that f' and f'' are indeed real-valued functions.

b) Determine f'' and f' explicitly, and plot them for $-3 < \omega < 3$.

c) Show that

$$\int_{-\infty}^{\infty} \frac{d\omega}{\pi}\, \frac{f''(\omega)}{\omega - z} = f(z)$$

(5 points)

2.2.2. **Another causal function**

The function considered in Problem 2.2.1 is an example of a class of complex functions called *causal functions* that are important in the theory of many-particle systems. Another member of this class is

$$g(z) = \sqrt{z^2 - 1} - z$$

Determine the spectrum and the reactive part of $g(z)$, and plot them for $-3 < \omega < 3$.

(3 points)

2.2.3. **Proof of the Cauchy-Riemann Theorem**

Prove the Cauchy-Riemann theorem from ch.2 §2.2:

a) Let $f(z) = f'(z', z'') + i f''(z', z'')$ be analytic everywhere in $\Omega \subseteq \mathbb{C}$. Show that the Cauchy-Riemann equations

$$\frac{\partial f'}{\partial z'} = \frac{\partial f''}{\partial z''} \qquad \text{and} \qquad \frac{\partial f'}{\partial z''} = -\frac{\partial f''}{\partial z'}$$

hold $\forall z \in \Omega$.

hint: Start with the difference quotient $(f(z) - f(z_0))/(z - z_0)$ and require that it's limit for $z \to z_0$ exists if z_0 is approached on paths either parallel to the real axis, or parallel to the imaginary axis.

b) Let the Cauchy-Riemann equations hold in a point $z_0 \in \Omega$. Show that this implies that f is analytic in the point z_0.

hint: Consider $f(z) - f(z_0)$ and expand $f'(z', z'')$ and $f''(z', z'')$ in Taylor series about z_0.

(8 points)

2.2.4. **Exponentials**

Consider the exponential function

$$f(z) = e^z = e^{z' + iz''}$$

a) Show that $f(z)$ is analytic everywhere in $\mathbb{C}$.

b) Convince yourself explicitly that the real and imaginary parts of f obey Laplace's differential equation.

c) Show that $df/dz|_z = f(z)$.

d) Show that $\cos z$ and $\sin z$, defined by

$$\cos z = \frac{1}{2}\left(e^{iz} + e^{-iz}\right) \quad , \quad \sin z = \frac{1}{2i}\left(e^{iz} - e^{-iz}\right)$$

are analytic everywhere in $\mathbb{C}$, and that

$$\frac{d}{dz}\cos z = -\sin z \quad , \quad \frac{d}{dz}\sin z = \cos z \ .$$

(4 points)

3 Integration in the complex plane

3.1 Path integrals

Definition 1. Let $I := [t_1, t_2] \subset \mathbb{R}$ be a real interval and let $z : I \to \mathbb{C}$ be continuously differentiable. The we define a **path** $\mathcal{C}$ as the set $\mathcal{C} := \{z(t), t \in I\}$ and we will often write $\mathcal{C}(t)$ instead of $z(t)$. If $\mathcal{C}(t_1) = \mathcal{C}(t_2)$ we say the path is **closed**.

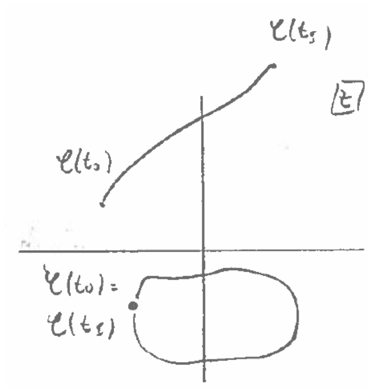


Fig. 2.2.9. An open path and a closed path.

Remark 1. This definition is the same as in §1.2 if we take into account the isomorphism between $\mathbb{R}$ and $\mathbb{C}$

Definition 2. Let $\Omega \subset \mathbb{C}$. Let $\mathcal{C}$ be a path with $\mathcal{C}(t) \in \Omega \quad \forall t \in I$. Let $f : \Omega \to \mathbb{C}$ be a complex function. We define the **path integral** of f along $\mathcal{C}$ by

$$\int_{\mathcal{C}} dz\, f(z) := \int_{t_0}^{t_1} dt\, f(\mathcal{C}(t)) \frac{d\mathcal{C}}{dt}$$

with the r.h.s being an ordinary Riemann Integral over I.

Remark 2. Note that this is similar to, but different from §1.2 **Definition 3**.

Definition 3. Let $f : \Omega \to \mathbb{C}$ be a complex function. $F : \Omega \to \mathbb{C}$ is called an **indefinite integral** of f if F is holomorphic on Ω and $\left.\frac{dF}{dz}\right|_{z_0} = f(z_0) \quad \forall z_0 \in \Omega$.

Definition 4. A region $\Omega \subset \mathbb{C}$ is called **simply connected** if any path in Ω can be continuously deformed to a point.

Remark 3. "Simply connected" means "no holes".

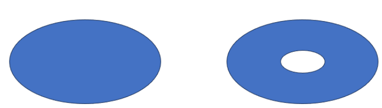


Fig. 2.2.10. A region that's simply connected, and a region that isn't.

Theorem 1. ***Cauchy's Integral Theorem***
Let $\Omega \subset \mathbb{C}$ be simply connected and let $f : \Omega \to \mathbb{C}$ be holomorphic on Ω. Then f has an indefinite integral F and

$$\int_{\mathcal{C}} dz\, f(z) = F(z_1) - F(z_0)$$

for all paths $\mathcal{C} : [t_0, t_1] \to \mathbb{C}$ with $\mathcal{C}(t_0) = z_0, \mathcal{C}(t_1) = z_1$.

Proof. See, e.g., Smirnov III/2 Sec. 4, or Whittaker & Watson (Not difficult but lengthy) □

Remark 4. This says in particular that the integral depends only on the starting and ending points of $\mathcal{C}$. For fixed z_0 and z_1 the integral is independent of the path.

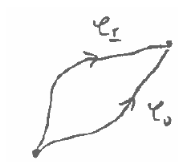


Fig. 2.2.11. Two paths with the same starting and end points.

Remark 5. This implies that one can deform the integration contour (path) without changing the value of the integral, provided one stays within the region where the function is holomorphic.

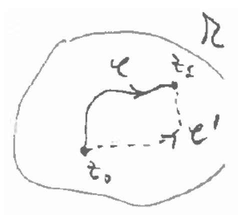


Fig. 2.2.12. Deformation of the integration contour.

Corollary 1. *Let $\Omega \subset \mathbb{C}$ be simply connected and let $f : \Omega \to \mathbb{C}$ be holomorphic. Let $\mathcal{C}$ be a closed path in Ω. Then*

$$\oint_{\mathcal{C}} dz\, f(z) = 0$$

Proposition 1. *Let $\Omega \subset \mathbb{C}$ be a region (not necessarily simply connected). Let $f(z)$ be holomorphic in Ω. Let $F(z)$ be an indefinite integral of $f(z)$ that is single-valued $\forall z \in \Omega$. Then*

$$\int_{\mathcal{C}} dz\, f(z) = F(z_1) - F(z_0)$$

for all paths that start at z_0 and end in z_1.

Proof. Let $\mathcal{C}(t_0) = z_0$, $\mathcal{C}(t) = z$, and define $F_1(z) := \int_{\mathcal{C}} dz'\, f(z')$. Then

$$\begin{aligned}\frac{dF_1}{dz}(z) &= \frac{d}{dz}\int_{t_0}^{t} dt'\, f(\mathcal{C}(t'))\,\frac{d\mathcal{C}}{dt'}\\ &= \lim_{\delta z\to 0}\frac{1}{\delta z}\int_{t}^{t+\delta t} dt'\, f(\mathcal{C}(t'))\,\frac{d\mathcal{C}}{dt'}\\ &= \lim_{\delta z\to 0}\frac{1}{\delta z}\,\delta t\,\frac{d\mathcal{C}}{dt'}\Big|_t\, f(\mathcal{C}(t))\\ &= \lim_{\delta z\to 0}\frac{1}{\delta z}\,\delta t\,\frac{d\mathcal{C}}{dt'}\, f(\mathcal{C}(t))\\ &= f(z)\end{aligned}$$

$\Rightarrow F_1(t) = F(z)+$ constant
Choose $z = z_0 \Rightarrow F_1(z_0) = 0 \Rightarrow$ constant $= -F(z_0)$ □

Remark 6. If Ω is not simply connected, then we can always cut it such that $\mathcal{C}$ does not cross any cuts. $\Rightarrow$ We can assume $f(z)$ to be single-valued.

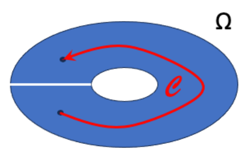


Fig. 2.2.13. Cutting the region to make the function single valued.

3.2 Laurent series

Lemma 1. *Let $f : \Omega \to \mathbb{C}$ be holomorphic. Let $z_0 \in \Omega$. Then $\exists R > 0$ such that*

$$f(z) = \sum_{n=0}^{\infty} f_n (z - z_0)^n \;\forall z, |z - z_0| < R \tag{2.2}$$

This representation of f in the vicinity of z_0 is unique and

$$f_n = \frac{1}{n!}\frac{d^n f}{dz^n}\Big|_{z_0}$$

Proof. See Smirnov III/2 Sec. 11 □

Remark 1. Equation 2.2 is called the **Taylor series** for f about z_0.

Remark 2. If a Taylor series exists, then the function is called analytic in z_0. Lemma $\Rightarrow$ Every holomorphic function is analytic. The converse is obviously true $\Rightarrow$ Holomorphism and analyticity are equivalent. From now on we will say "analytic" instead of "holomorphic".

Corollary 1. *Let* $\tilde{\tilde{g}}(z) = \sum_{n=0}^{\infty} \tilde{g}_n (z-z_0)^n$ *for* $|z-z_0| < R_1$ *be the Taylor series for the function* $\tilde{g}$ *in the vicinity of* z_0*. Then* $\exists\ R_2 > 0$ *such that the function* $g(z)$ *defined by* $g(z) = \sum_{n=0}^{\infty} \tilde{g}_n \frac{1}{(z-z_0)^n}$ *is analytic for all* z *with* $|z-z_0| > R_2$

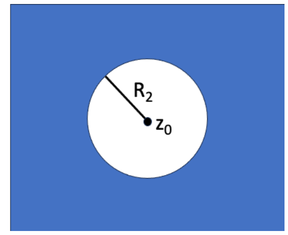


Fig. 2.2.14. $g(z)$ is analytic for $|z-z_0| > R_2$.

Proof. Define $\tilde{z} := \frac{1}{z-z_0} \Rightarrow g(z) = \sum_{n=0}^{\infty} \tilde{g}_n \tilde{z}^n =: \tilde{g}(z)$
Lemma $\Rightarrow \exists\ 1/R_2 > 0 : \tilde{g}(\tilde{z}$ exists and is analytic for $|\tilde{z}| < 1/R_2$
$\Rightarrow g(z)$ exists and is analytic for $\frac{1}{|z_0 - z|} < R_2 \iff |z-z_0| > R_2$. □

Corollary 2. *The function*

$$h(z) := f(z) + g(z) = \sum_{n=0}^{\infty} f_n (z-z_0)^n + \sum_{n=0}^{\infty} \tilde{g}(z-z_0)^{-n} = \sum_{n=-\infty}^{\infty} h_n (z-z_n)^n$$

with

$$h_n = \begin{cases} f_n & \text{for } n > 0 \\ \tilde{g}_{-n} & \text{for } n < 0 \\ f_n + \tilde{g}_n & \text{for } n = 0 \end{cases}$$

is analytic on the annulus $R_2 < |z-z_0| < R_1$.

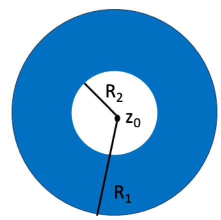


Fig. 2.2.15

Remark 3. This obviously is a useful statement only if $R_2 < R_1$ (as shown in Fig. 2.2.15). Otherwise the Laurent series that defines $h(z)$ does not converge anywhere.

Definition 1. $\sum_{n=-\infty}^{\infty} h_n (z-z_0)^n$ is called the **Laurent series** for the function $h(z)$ on the annulus $R_2 < |z-z_0| < R_1$.

Theorem 1. *Let* $f(z)$ *be analytic or an annulus* A *around* z_0*. Then on* A $f(z)$ *can be uniquely expanded in a Laurent series, i.e., there exist unique coefficients* f_n *such that* $f(z) = \sum_{n=-\infty}^{\infty} f_n (z-z_0)^n\ \forall z \in A$.

Proof. E.g. Smirnov III/2 Sec.15. □

3.2.1 Special Case

Consider the case where $R_2 = 0$. $f(z)$ is analytic on all of Ω except in $z = z_0$.
We then have **3 possibilities:**

(i) $f_n = 0 \; \forall \, n < 0$
Then $f(z)$ is analytic in z_0 as well

(ii) $\exists \, n < 0 : f_n \neq 0 \wedge \exists \, m > 0 : f_n = 0 \; \forall \, n < -m$

Remark 4. This means that the strongest singularity in the series is a term $\dfrac{1}{(z - z_0)^m}$.

Definition 2. We say that $f(z)$ has a **pole of multiplicity m** at the point z_0. For $m = 1$ we call the pole a **simple pole**.

(iii) $\nexists \, m > 0 : f_n = 0 \; \forall \, n < -m$
Remark 5. The regularity is stronger than any power. This is true in particular if $f_n \neq 0 \; \forall \, n < 0$.

Definition 3. We say that $f(z)$ has an **essential singularity** at z_0.

Definition 4. f_{-1} is called the residue of f in z_0, $f_{-1} =: \mathrm{Res} f(z_0)$.

Example 1. $f(z) = \dfrac{1}{z}$ has a simple pole at $z = 0$ with residue $\mathrm{Res} f(0) = 1$, and no other singularity.

Example 2. $f(z) = \dfrac{1}{(z - i)^2}$ has a pole of multiplicity 2 in $z = i$ and residue $\mathrm{Res} f(i) = 0$.

Example 3. $f(z) = e^{-1/z}$ has an essential singularity at $z = 0$.

Example 4. $f : \mathbb{C} \setminus \{0\} \to \mathbb{C}$, $f(z) = e^z = \sum_{n=0}^{\infty} \frac{1}{n!} z^n$ has $f_n = 0 \; \forall \, n < 0$ and can be analytically continued to all of $\mathbb{C}$: Define $\tilde{f} : \mathbb{C} \to \mathbb{C}$ such that $\tilde{f}(z) = f(z)$ for $z \neq 0$ and $f(0) = 1$. Then $\tilde{f}$ exists and is analytic on all of $\mathbb{C}$.

3.3 The residue theorem

Lemma 1. *Let $f(z) = (z - z_0)^n$, with $n \neq -1$. Let $\mathcal{C}$ be a closed path around z_0. Then*

$$\oint_{\mathcal{C}} dz \, f(z) = 0$$

Fig. 2.2.16

Proof. Let $n \geq 0$. Then (2.3) follows from Cauchy's Integral Theorem (§2.1).
Let $n \leq -2$. Then we have an indefinite integral of f given by

$$F(z) = \frac{1}{n+1}(z - z_0)^{n+1}$$

$F(z)$ is single-valued and analytic everywhere except at $z = z_0$
(2.3) follows from Proposition 1 in §3.1. □

Remark 1. For $n = -1$ the argument breaks down since $F(z) \propto \ln(z - z_0)$ is not single-valued, i.e. there is no single valued function $F(z)$ that $\frac{dF}{dz} = f(z)\ \forall z \neq z_0$.

Lemma 2. *Let $f(z) = \frac{1}{z - z_0}$, and let $\mathcal{C}$ be a closed path that goes around z_0 counterclockwise. Then*

$$\oint_{\mathcal{C}} dz\, f(z) = 2\pi i$$

Proof. Consider the path $\mathcal{C} + l_1 + \mathcal{C} + l_2$ (see Fig. 2.2.17), which does not contain z_0.

$$\Rightarrow \oint_{\mathcal{C}+l_1+l_2+\mathcal{C}} dz\, f(z) = 0$$

$$\Rightarrow \oint_{\mathcal{C}} dz\, f(z) + \oint_{\mathcal{C}} dz\, f(z) + \underbrace{\int_{l_1} dz\, f(z) + \int_{l_2} dz\, f(z)}_{=0} = 0$$

$$\Rightarrow \oint_{\mathcal{C}} dz\, f(z) + \oint_{\mathcal{C}} dz\, f(z) = 0$$

$$\Rightarrow \oint_{\mathcal{C}} dz\, f(z) = -\oint_{\mathcal{C}} dz\, f(z)$$

But $\mathcal{C}$ can be parameterized by $z - z_0 = Re^{i\varphi}$ $(0 \leq \varphi \leq 2\pi)$
$\Rightarrow dz = Re^{i\varphi}\, d\varphi$. Therefore,

$$\oint_{\mathcal{C}} dz\, f(z) = iR \int_0^{-2\pi} d\varphi\, e^{i\varphi} \frac{1}{Re^{i\varphi}} = -2\pi i \Rightarrow \oint_{\mathcal{C}} dz\, f(z) = 2\pi i$$

□

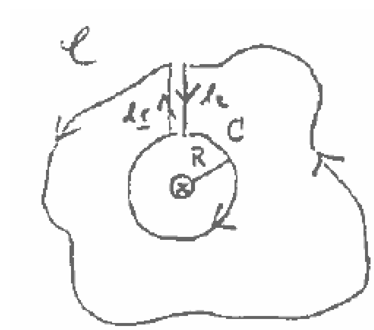


Fig. 2.2.17. A modification of the path in Fig. 2.2.16.

Theorem 1. *Residue Theorem*
Let $f : \Omega \to \mathbb{C}$ be analytic except for isolated points $z_1, ..., z_n \in \Omega$. Let $\mathcal{C}$ be a closed path that encloses $z_1, ..., z_n$. Then

$$\oint_{\mathcal{C}} dz\, f(z) = 2\pi i \sum_{j=1}^{n} Resf(z_j)$$

where the integration along $\mathcal{C}$ is counterclockwise.

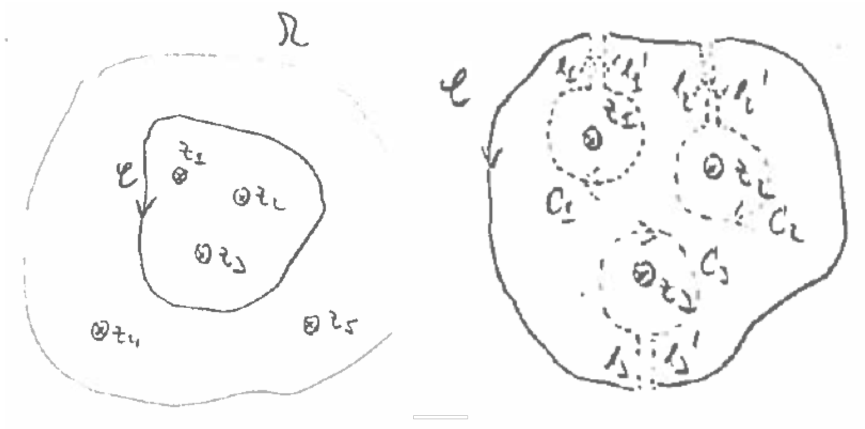


Fig. 2.2.18. A path around singular points and its modification.

Proof. Consider the contour $\mathcal{C} + \sum_{j=1}^{n}(l_j + l'_j) + \sum_{j=1}^{n} \mathcal{C}_j$ (see Fig. 2.2.18)
Using the arguments used to prove Lemma 1 we can write

$$\oint_{\mathcal{C}} dz\, f(z) = -\sum_{j=1}^{n} \oint_{\mathcal{C}_j} dz\, f(z)$$

Using theorem §3.2, the circles can be closed such that $f(z)$ can be uniquely expanded in a Laurent series along the circles.

$$\begin{aligned}
\oint_{\mathcal{C}} dz\, f(z) &= -\sum_{j=1}^{n} \oint_{\mathcal{C}_j} dz \sum_{m=-\infty}^{\infty} f_m^{(j)} (z - z_j)^m \\
&= -\sum_{j=1}^{n} \oint_{\mathcal{C}_j} dz\, f_{-1}^{(j)} \frac{1}{z - z_j} && \text{By Lemma 1} \\
&= 2\pi i \sum_{j=1}^{n} f_{-1}^{(j)} && \text{By Lemma 2} \\
&= 2\pi i \sum_{j=1}^{n} \mathrm{Res} f(z_j)
\end{aligned}$$

□

Remark 2. This theorem is invaluable for many different physical applications.

Remark 3. We have reduced a class of integrals to determining the residues of the integrand!

Proposition 1. *Let $f(z)$ have a pole of multiplicity m in the point z_0. Then*

$$Resf(z_0) = \frac{1}{(m-1)!} \frac{d^{m-1}}{z^{m-1}}\Big|_{z=z_0} [f(z)(z - z_0)^m]$$

Proof. Define $g(z) = (z - z_0)^m f(z)$ which is analytic, so we can use §3.2 to Taylor expand it as $g(z) = g_0 + g_1(z - z_0) + g_2(z - z_0)^2 + ...$ with $g_n = \frac{1}{n!}\frac{d^n}{z^n}\Big|_{z_0} g(z)$

$$\Rightarrow f(z) = \frac{g_0}{(z - z_0)^n} + \frac{g_1}{(z - z_0)^{n-1}} + ... + \frac{g_{m-1}}{z - z_0} + g_m + ...$$

$$\Rightarrow \text{Res} f(z_0) = g_{m-1} = g_{m-1} = \frac{1}{(m-1)!}\frac{d^{m-1}}{z^{m-1}}\Big|_{z_0} [(z - z_0)^m f(z)]$$

□

Example 1. $m = 1 : \text{Res} f(z_0) = (z - z_0) f(z)\Big|_{z_0} = f_{-1}$ §2.2 Definition 4 ✓

Example 2.

$$f(z) = \frac{1}{(z - z_0)^n}, n \neq 1 \text{ Then } \text{Res} f(z_0) = \frac{1}{(n-1)!}\frac{d^{m-1}}{z^{m-1}}\Big|_{z_0}\frac{(z - z_0)^n}{(z - z_0)^n} = 0 \text{ §3.3 Lemma 1✓}$$

Remark 4. Strategy for solving integrals:
(i) Determine the analytic structure of the integrand.
(ii) Deform the contour to simplify the integration while staying away from singularities.
(iii) Employ the residue theorem and do any additional integrals that arise from deforming the contour.

3.4 Simple applications of the residue theorem

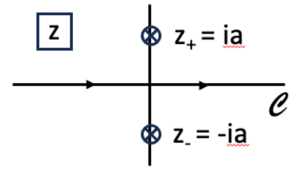
Consider some simple examples

Example 1. $f : \mathbb{R} \to \mathbb{R},\ f(x) = \dfrac{1}{x^2 + a^2} \qquad (a \in \mathbb{R}).$

Find $g(k) := \displaystyle\int_{-\infty}^{\infty} dx\, e^{-ikx} f(x)$ ("Fourier Transform").

Continue f into the complex plane:

$$f(z) = \frac{1}{z^2 + a^2} = \frac{1}{(z - ia)(z + ia)} =: \frac{1}{(z - z_-)(z - z_+)}$$

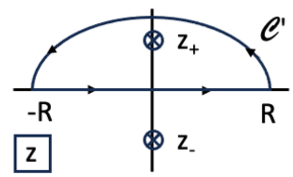


$f(z)$ is analytic everywhere except at $z_\pm = \pm ia$

1st Case: $k < 0$

Consider the integral over the path $\mathcal{C}'$

$$\begin{aligned}\mathcal{I}(k) :&= \int_{\mathcal{C}'} dz\, e^{-ikz} f(z) = 2\pi i \operatorname{Res}|_{z_+} f(z) e^{-ikz} \\ &= 2\pi i\, \frac{e^{-ikz_+}}{z_+ - z_-} = 2\pi i\, \frac{e^{-ikia}}{2ia} = \frac{\pi}{a}\, e^{ka} \\ &= \frac{\pi}{a}\, e^{-|k|a}\end{aligned}$$



and the integral over the semicircle with radius R:

$$\mathcal{J}_R(k) := \int_{\frown} dz\, e^{-ikz} f(z)$$

Then $g(k) = \mathcal{I}(k) - \lim\limits_{R\to\infty} \mathcal{J}_R(k)$

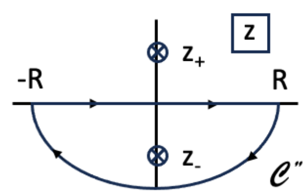
But $\left|\lim\limits_{R\to\infty} \mathcal{J}_R(k)\right| = \left|\lim\limits_{R\to\infty} \displaystyle\int_{\frown} dz\, e^{-ikz} f(z)\right| \le \lim\limits_{R\to\infty} \displaystyle\int_{\frown} dz\, |e^{-ikz}||f(z)| \le \lim\limits_{R\to\infty} \dfrac{\pi R}{R^2 + a^2} = 0$

$\Rightarrow g(k) = \mathcal{I}(k)$

2nd Case: $k > 0$

Integrate over $\mathcal{C}''$. Same argument as above, but with $z_+ \leftrightarrow z_-$:

$$\begin{aligned}\mathcal{I}(k) &= -2\pi i \operatorname{Res}\Big|_{z_-} f(z) e^{-ikz} \\ &= -2\pi i\, \frac{e^{-ikz_-}}{z_- - z_+} = -2\pi i\, \frac{e^{ikia}}{-2ia} \\ &= \frac{\pi}{a}\, e^{-ka}\end{aligned}$$



Combining the two cases we have

$$g(k) = \frac{\pi}{a}\, e^{-|k|a}\ \forall\, k \in \mathbb{R}$$

3.5 Another application of complex analysis: The Airy function Ai(x)

3.6 Problems

2.3.1. **Laurent series**

Find the Laurent series for the function

$$f(z) = 1/(z^2 + 1)$$

in the point $z = i$. That is, find the coefficients f_n that enter the theorem in ch. 2 §3.2.

(3 points)

2.3.2. **Applications of the residue theorem**

Use complex analysis to evaluate the real integrals

a)

$$\int_{-\infty}^{\infty} dx \, \frac{1}{x^4+1}$$

b)

$$\int_{-\infty}^{\infty} dx \, \frac{\sin x}{x}$$

hint: Write $\sin x = (e^{ix} - e^{-ix})/2i$ and consider the resulting two integrals with complex integrands. Why is this a good strategy?

c)

$$\int_{-\infty}^{\infty} dx \, \frac{\sin x}{x} \frac{1}{1+x^2}$$

and check your results by means of Wolfram Alpha.

Let $a \in \mathbb{C}$ with $\mathrm{Re}\, a > 0$. Use the residue theorem to show that

d)

$$\int_{-\infty}^{\infty} dx \, e^{-ax^2} = \sqrt{\pi/a}$$

Now let $a \in \mathbb{R}$ and consider the integral

e)

$$\int_{-\infty}^{\infty} \frac{dx}{x} \frac{1}{x^2+a^2}$$

and define its Cauchy principal value by

$$\lim_{R \to 0} [\int_{-\infty}^{-R} dx \, f(x) + \int_{R}^{\infty} dx \, f(x)]$$

with $f(x) = 1/x(x^2+a^2)$. Determine the Cauchy principal value using the residue theorem. Is the result consistent with the expectation for a real symmetric integral over an antisymmetric integrand?

hint: Go around the pole on a semicircle of radius R and let $R \to 0$.

(17 points)

2.3.3. **Matsubara frequency sum**

Let $f(z)$ have simple poles at z_j ($j = 1, 2, \ldots$), and no other singularities. Let $f(|z| \to \infty)$ go to zero faster then $1/z$. Consider the infinite sum

$$S = -T \sum_{n=-\infty}^{\infty} f(i\Omega_n)$$

with $\Omega_n = 2\pi T n$ and $T > 0$. Show that

$$S = \sum_j n(z_j) \, \mathrm{Res}\, f(z_j)$$

where $n(z) = 1/(e^{z/T} - 1)$ is the Bose distribution function.

hint: Show that $n(z)$ has simple poles at $z = i\Omega_n$, and integrate $n(z)\, f(z)$ over an infinite circle centered on the origin.

note: Sums of this form are important in finite-temperature quantum field theory. In this context, T is the temperature and Ω_n is called a “bosonic Matsubara frequency".

(3 points)

4 Fourier transforms and generalized functions

4.1 The Fourier transform in classical analysis

Let $f : \mathbb{R}^n \to \mathbb{C}$ be a complex-valued function that is absolutely integrable, i.e., $\int d\vec{x}\,|f(\vec{x})| < \infty$.

Remark 1. The space $\gamma^{(1)}$ of such functions is a vector space over $\mathbb{C}$ under addition of functions (proof easy).

Notation: $\vec{x} = (x_1, x_2, \ldots, x_n)) \in \mathbb{R}^n$, $\int d\vec{x} = \int_{\mathbb{R}^n} dx_1\, dx_2 \ldots dx_n$, $\vec{k} \cdot \vec{x} = k_1 x_1 + k_2 x_2 + \ldots + k_n x_n$ for $\vec{k} \in \mathbb{R}^n$

Definition 1. The function

$$\hat{f}(\vec{k}) := \int d\vec{x}\; e^{-i\vec{k}\cdot\vec{x}}\, f(\vec{x}) \equiv FT(f)(\vec{k})$$

is called **Fourier transform** (FT) of f.

Remark 2. $\hat{f} : \mathbb{R}^n \to \mathbb{C}$ is another complex-valued function on $\mathbb{R}^n$.

Remark 3. The FT is a linear integral transformation.

Remark 4. $FT(\lambda_1 f_1 + \lambda_2 + f_2) = \lambda_1\, FT(f_1) + \lambda_2\, FT(\lambda_2) \quad \forall \lambda_1, \lambda_2 \in \mathbb{C}$ due to linearity.

Proposition 1. *$\hat{f}(\vec{k})$ is bounded and continuous.*

Proof.

$$\left|\hat{f}(\vec{k})\right| = \left|\int d\vec{x}\, e^{-i\vec{k}\cdot\vec{x}}\, f(\vec{x})\right| \leq \int d\vec{x}\, \left|e^{-i\vec{k}\cdot\vec{x}}\, f(\vec{x})\right| = \int d\vec{x}\, |f(\vec{x})| < \infty$$

$\Rightarrow \hat{f}$ is bounded.

$$|\hat{f}(\vec{k}_1) - \hat{f}(\vec{k}_2)| = \left|\int d\vec{x}\, \left(e^{-i\vec{k}_1\cdot\vec{x}} - e^{-i\vec{k}_2\cdot\vec{x}}\right) f(\vec{x})\right| \leq \int d\vec{x}\, \left|e^{-i\vec{k}_1\cdot\vec{x}} - e^{-i\vec{k}_2\cdot\vec{x}}\right| |f(\vec{x})| \to 0 \quad \text{for} \quad \vec{k}_1 \to \vec{k}_2$$

$\Rightarrow \hat{f}$ is continuous. □

Proposition 2. *Let $x_\ell f(\vec{x})$ be absolutely integrable. Then $\hat{f}(\vec{k})$ is differentiable with respect to k_ℓ and*

$$\frac{\partial}{\partial k_\ell}\, \hat{f}(\vec{k}) = FT(-ix_\ell f)(\vec{k})$$

Proof.

$$\frac{\partial}{\partial k_\ell} \int d\vec{x}\, e^{-i\vec{k}\cdot\vec{x}}\, f(\vec{x}) = -i \int d\vec{x}\, e^{-i\vec{k}\cdot\vec{x}}\, x_\ell\, f(\vec{x}) = FT(-ix_\ell f)(\vec{k})$$

□

4.2 Inverse Fourier transforms

4.3 Test functions

4.4 Generalized functions

4.5 Dirac's δ-function

4.6 Problems

2.4.1. **1-d Fourier transforms**

Consider a function f of one real variable x. Calculate the Fourier transforms $\hat{f}(k) = \int dx\, e^{-ikx}\, f(x)$ of the following functions:

a) $f(x) = \begin{cases} 1 & \text{for } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad .$

b) $f(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad .$

c) $f(x) = e^{-(x/x_0)^2} \quad .$

(3 points)

2.4.2. **3-d Fourier transforms**

Consider a function f of one vector variable $\boldsymbol{x} \in \mathbb{R}^3$. The Fourier transform $\hat{f}$ of f is defined as

$$\hat{f}(\boldsymbol{k}) = \int d\boldsymbol{x}\; e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\, f(\boldsymbol{x}) \quad .$$

Calculate the Fourier transforms of the following functions:

a)

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{for } r < r_0 \qquad (r = |\boldsymbol{x}|) \\ 0 & \text{otherwise} \end{cases} \quad .$$

b)

$$f(\boldsymbol{x}) = 1/r \quad .$$

hint: Consider $g(\boldsymbol{x}) = \frac{1}{r}\, e^{-r/r_0}$ and let $r_0 \to \infty$.

(3 points)

2.4.3. **More 1-d Fourier transforms**

Consider a function of time $f(t)$ and define its Fourier transform

$$\hat{f}(\omega) := \int dt\; e^{i\omega t}\, f(t)$$

and its Laplace transform $F(z)$ as

$$F(z) = \pm i \int dt\, e^{izt}\, f_{\pm}(t) \qquad (\pm \text{ for } \mathrm{sgn}(\mathrm{Im}\, z) = \pm 1)$$

with z a complex frequency and $f_{\pm}(t) = \Theta(\pm t)\, f(t)$. Further define

$$F''(\omega) = \frac{1}{2i}\left[F(\omega + i0) - F(\omega - i0)\right] \qquad , \qquad F'(\omega) = \frac{1}{2}\left[F(\omega + i0) + F(\omega - i0)\right]$$

Calculate $F''(\omega)$ and $F'(\omega)$ for

a) $f(t) = e^{-|t|/\tau}$

b) $f(t) = e^{i\omega_0 t}$

hint: $\lim_{\epsilon \to 0} \epsilon/(x^2 + \epsilon^2) = \pi\, \delta(x)$, with $\delta(x)$ the familiar Dirac delta-function, which we will study in detail in Week 10.

Show that in both cases $\int \frac{d\omega}{\pi} \frac{F''(\omega)}{\omega} = F'(\omega = 0)$.

note: These concepts are important for the theory of response functions.

(4 points)

2.4.4. **Regularizations of the constant function, and of the sign function**

Prove the following statements from ch.2 §3.4:

a) The sequence $f_n(x) = e^{-x^2/n^2}$ is a regular sequence of test functions that is a regularization of the generalized function $f(x) \equiv 1$.

b) The sequence $f_n(x) = \tanh(nx)$ is a regularization of the generalized function $f(x) = \operatorname{sgn} x$.

(4 points)

2.4.5. **Distribution limits**

a) Show that the sequences
$$f_n(x) = \frac{1}{\pi x} \sin(nx) \qquad (n = 1, 2, \ldots)$$
and
$$g_n(x) = \frac{1}{\pi n} \frac{1}{x^2 + 1/n^2} \qquad (n = 1, 2, \ldots)$$
yield the δ-function as $n \to \infty$ in a distribution-limit sense:
$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} g_n(x) = \delta(x)$$

b) Show that
$$\frac{d}{dx} \operatorname{sgn} x = 2\,\delta(x)$$
and
$$\frac{d}{dx}\,\Theta(x) = \delta(x)$$
with $\Theta(x)$ the step function.

c) Show that
$$\lim_{\epsilon\to 0} \epsilon\,|x|^{\epsilon-1} = 2\,\delta(x)$$
hint: Start with ch.3 §2.4 example 6 and differentiate.

(6 points)