Geometric Methods in Mathematical Physics I: Multi-Linear Algebra, Tensors, a little Spinors and Special Relativity

by **Valter Moretti**
https://moretti.maths.unitn.it/home.html

Department of Mathematics,
University of Trento
Italy

Contents

Chapter 1

Introduction

The content of these lecture notes covers the first part of the lectures of a graduate course in Modern Mathematical Physics at the University of Trento. The course has two versions, one is geometric and the other is analytic. These lecture notes only concern the geometric version of the course. The analytic version regarding applications to linear functional analysis to quantum and quantum relativistic theories is covered by my books [Moretti-a], [Moretti-b] and the chapter [KhMo15].

The idea of this first part is to present a quick, but rigorous, picture of the basics of tensor calculus for the applications to mathematical and theoretical physics. In particular I try to cover the gap between the traditional presentation of these topics given by physicists, based on the practical indicial notation and the more rigorous (but not always useful in the practice) approach presented by mathematicians.

Several applications are presented as examples and exercises. Some chapters concern the geometric structure of Special Relativity theory and some theoretical issues about Lorentz group.

The reader is supposed to be familiar with standard notions of linear algebra [Lang, Sernesi], especially concerning finite dimensional vector spaces. Since the end of Chapter 8 some basic tools of Lie group theory and Lie group representation theory [KNS] are requested. The definition of Hilbert tensor product given at the end of Chapter 2 has to be seen as complementary material and requires that the reader is familiar with elementary notions on Hilbert spaces.

Some notions and results exploited several times throughout the text are listed here.
First of all, we stress that the notion of *linear independence* is always referred to *finite* linear combinations. A possibly infinite subset $A \subset V$ of vectors of the vector space V over the field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$ is said to be made of **linearly independent** elements, if for avery *finite* subset $A' \subset A$,

$$\sum_{v \in A'} c_v v = 0 \quad \text{for some } c_v \in \mathbb{K},$$

entails

$$c_v = 0 \text{ for all } v \in A'.$$

Clearly such A cannot contain the zero vector.
We remind the reader that, as a consequence of *Zorn's lemma* and as proved below, every vector space V admits a **vector basis**, that is a *possibly infinite* set $B \subset V$ made of *linearly independent* elements such that every $v \in V$ can be written as

$$v = \sum_{i=1}^{n} c_i v_i$$

for some *finite* set of (distinct) elements $v_1, \ldots, v_n \in B$ depending on v and a corresponing set of (not necessarily pairwise distinct) scalars $c_1, \ldots, c_n \in \mathbb{K} \setminus \{0\}$ depending on v. The sets of elements $v_1, \ldots, v_n \in B$ and $c_1, \ldots, c_n \in \mathbb{K} \setminus \{0\}$ are uniquely fixed by v itself in view of the linear independence of the elements of B. Indeed, if for some distinct elements $v'_1, \ldots, v'_m \in B$ and (not necessarily pairwise distinct) $c'_1, \ldots, c'_m \in \mathbb{K} \setminus \{0\}$ it again holds

$$v = \sum_{i=1}^{m} c'_i v'_i \,,$$

we also have

$$0 = v - v = \sum_{i=1}^{n} c_i v_i - \sum_{j=1}^{m} c'_j v_j \;.$$

In view of the linear independence of the set $\{v_1, \ldots, v_n, v'_1, \ldots, v'_m\}$ whose primed and non-primed elements are not necessarily distinct, observing that $c_i \neq 0$ and $c'_j \neq 0$ the found identity implies that $n = m$, the set of v_is must coincide with the set of v'_js and the corresponding coeffients c_i and c'_j must be equal.

Theorem 1.1. *Let V be a vector space over the field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. There is a vector basis $B \subset V$. Furthermore, if $v_1, \ldots, v_n \in V$ are linearly independent, there is a vector basis $B \subset V$ such that $B \ni v_1, \ldots, v_n$.*

Proof. Consider the class $\mathscr{B}$ of all possible sets $L \subset V$ of linearly independent elements of V. Define the partial ordering relation on $\mathscr{B}$ given by the theoretical set inclusion relation $\subset$. It is clear that, if $\mathscr{T} \subset \mathscr{B}$ is totally ordered, then $\cup_{L \in \mathscr{T}} L \in \mathscr{B}$ is an upper bound for $\mathscr{T}$. Therefore Zorn's lemma entails that $\mathscr{B}$ admits maximal elements. Let $B \in \mathscr{B}$ such an element. As B is maximal, if $u \in V \setminus \{0\}$ the set $\{u\} \cup B$ cannot be made of linearly independent elements. Since B is made of linearly independent elements, it means that there is a finite number of elements $u_1, \ldots, u_m \in B$ such that $cu + \sum_{k=1}^{m} c_k u_k = 0$ holds for $c \neq 0$ and some $c_k \in \mathbb{K}$. In other words every $u \in V$ can be written as a finite linear combination of elements of B and thus B is a vector basis of V.
Regarding the proof of the second statement, it is sufficient to re-define $\mathscr{B}$ as the class of all possible sets $L \subset V$ of linearly independent elements of V which always include $v_1, \ldots, v_n$ as elements and to take advantage of Zorn's lemma. In this case a maximal element B of $\mathscr{B}$ is necessarily a vector basis of V including $v_1, \ldots, v_n$ among its elements. $\square$

Acknowledgments

The author is grateful to Mikael Ecco, Leonardo Goller, Omar Zoghlami for some useful remarks and corrections.

Chapter 2

Multi-linear Maps and Tensors

Within this section we introduce basic concepts concerning multi-linear algebra and tensors. The theory of vector spaces and linear maps is assumed to be well known.

2.1 Dual space and conjugate space, pairing, adjoint operator

As a first step we introduce the dual space and the conjugate space of a given vector space.

Definition 2.1. **(Dual Space, Conjugate Dual Space and Conjugate space.)** Let V be a vector space on the field either $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.
(a) The **dual space** of V, V^*, is the vector space of linear functionals on V, i.e., the linear maps $f : V \to \mathbb{K}$.
(b) If $\mathbb{K} = \mathbb{C}$, the **conjugate dual space** of V, $\overline{V^*}$, is the vector space of anti-linear functionals on V, i.e., the antilinear maps $g : V \to \mathbb{C}$. Finally the **conjugate space** of V, $\overline{V}$ is the space $(\overline{V^*})^*$. ∎

Remarks 2.2.
(1) If V and V' are vector spaces on $\mathbb{C}$, a map $f : V \to V'$ is called *anti linear* or *conjugate linear* if it satisfies

$$f(\alpha u + \beta v) = \overline{\alpha} f(u) + \overline{\beta} f(v)$$

for all $u, v \in V$ and $\alpha, \beta \in \mathbb{C}$, $\overline{\lambda}$ denoting the complex conjugate of $\lambda \in \mathbb{C}$. If $V' = \mathbb{C}$ the given definition reduces to the definition of *anti-linear functional*.
(2) V^*, $\overline{V^*}$ and $\overline{V}$ turn out to be vector spaces on the field $\mathbb{K}$ when the composition rule of vectors and the product with elements of the field are defined in the usual way. For instance, if $f, g \in V^*$ or $\overline{V^*}$, and $\alpha \in \mathbb{K}$ then $f + g$ and αf are functions such that:

$$(f + g)(u) := f(u) + g(u)$$

and

$$(\alpha f)(u) := \alpha f(u)$$

for all of $u \in V$.

In Definition 2.3, we do not explicitly assume that V is finite dimensional.

Definition 2.3. **(Dual Basis, Conjugate Dual Basis and Conjugate Basis.)** Let V be a vector space on either $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Let $\{e_i\}_{i\in I}$ be a vector basis of V. The set $\{e^{*j}\}_{j\in I} \subset V^*$ whose elements are defined by

$$e^{*j}(e_i) := \delta_i^j$$

for all $i, j \in I$ is called the **dual basis** of $\{e_i\}_{i\in I}$.
Similarly, if $\mathbb{K} = \mathbb{C}$, the set of elements $\{\overline{e^*}^j\}_{j\in I} \subset \overline{V^*}$ defined by:

$$\overline{e^*}^j(e_i) := \delta_i^j$$

for all $i, j \in I$ is called the **conjugate dual basis** of $\{e_i\}_{i\in I}$.
Finally the set of elements $\{\overline{e}_j\}_{j\in I} \subset \overline{V}$ defined by:

$$\overline{e}_p(\overline{e^*}^q) := \delta_p^q$$

for all $p, q \in I$ is called the **conjugate basis** of $\{e_i\}_{i\in I}$. ■

Remarks 2.4.
(1) One may wonder whether or not the *dual bases* (and the conjugate bases) are *proper* vector bases of the respective vector spaces, the following theorem gives a positive answer in the finite-dimensional case. This is not true, in general, for infinite-dimensional spaces V. We shall not try to improve the definition of dual basis in the general case, since we are interested on algebraic features only and the infinite-dimensional case should be approached by convenient topological tools which are quite far form the goals of these introductory notes.
(3) In *spinor* theory – employed, in particular, to describe the *spin* of quantum particles – there is is a complex two-dimensional space V called the *space of Weyl spinors*. It is the representation space of the group $SL(2, \mathbb{C})$ viewed as a double-valued representation of the orthochronous proper Lorentz group.
Referring to a basis in V and the associated in the spaces V^*, $\overline{V}$ and $\overline{V}^*$, the following notation is often used in physics textbooks [Wald, Streater-Wightman]. The components of the spinors, i.e. vectors in V are denoted by ξ^A. The components of dual spinors, that is vectors of V^*, are denoted by ξ_A. The components of conjugate spinors, that is vectors of $\overline{V}$, are denoted either by $\xi^{A'}$ or, using the dot index notation, $\xi^{\dot{A}}$. The components of dual conjugate spinors, that is vectors of $\overline{V}^*$, are denoted either by $\xi_{A'}$ or by $\xi_{\dot{A}}$. ■

Theorem 2.5. *If $dimV < \infty$ concerning definition 2.3, the dual basis, the conjugate dual basis and the conjugate basis of a base $\{e_i\}_{i\in I} \subset V$, are proper vector bases for V^*, $\overline{V^*}$ and $\overline{V}$ respectively. As a consequence $dimV = dimV^* = dim\overline{V^*} = dim\overline{V}$.*

Proof. Consider the dual basis in V^*, $\{e^{*j}\}_{j\in I}$. We have to show that the functionals $e^{*j} : V \to \mathbb{K}$ are generators of V^* and are linearly independent.
(Generators.) Let us show that, if $f : V \mapsto \mathbb{K}$ is linear, then there are numbers $c_j \in \mathbb{K}$ such that $f = \sum_{j\in I} c_j e^{*j}$.
To this end define $f_j := f(e_j)$, $j \in I$, then we argue that $f = f'$ where $f' := \sum_{j\in I} f_j e^{*j}$. Indeed, any $v \in V$ may be decomposed as $v = \sum_{i\in I} v^i e_i$ and, by linearity, we have:

$$f'(v) = \sum_{j\in I} f_j e^{*j}(\sum_{i\in I} v^i e_i) = \sum_{i,j\in I} f_j v^i e^{*j}(e_i) = \sum_{i,j\in I} f_j v^i \delta_i^j = \sum_{j\in I} v^j f_j = \sum_{j\in I} v^j f(e_j) = f(v)\ .$$

(Notice that above we have used the fact that one can extract the summation symbol from the argument of each e^{*j}, this is because the sum on the index i is *finite* by hypotheses it being $dim V < +\infty$.) Since $f'(v) = f(v)$ holds for all of $v \in V$, we conclude that $f' = f$.
(Linear independence.) We have to show that if $\sum_{k\in I} c_k e^{*k} = 0$ then $c_k = 0$ for all $k \in I$.
To achieve that goal notice that $\sum_{k\in I} c_k e^{*k} = 0$ means $\sum_{k\in I} c_k e^{*k}(v) = 0$ for all $v \in V$. Therefore, putting $v = e_i$ and using the definition of the dual basis, $\sum_{k\in I} c_k e^{*k}(e_i) = 0$ turns out to be equivalent to $c_k \delta_i^k = 0$, namely, $c_i = 0$. This result can be produced for each $i \in I$ and thus $c_i = 0$ for all $i \in I$. The proof for the conjugate dual basis is very similar and is left to the reader. The proof for the conjugate basis uses the fact that $\overline{V}$ is the dual space of $\overline{V^*}$ and thus the first part of the proof applies. The last statement of the thesis holds because, in the four considered cases, the set of the indices I is the same. $\square$

Remarks 2.6. Let $\{e_i\}_{i\in I}$ be a vector basis of the vector space V with field $\mathbb{K}$ and consider the general case with I *infinite*. Each linear or anti-linear map $f : V \to \mathbb{K}$ is anyway completely defined by giving the values $f(e_i)$ for all of $i \in I$. This is because, if $v \in V$ then $v = \sum_{i\in I_v} c^i e_i$ for some numbers $c^i \in \mathbb{K}$, $I_v \subset I$ being *finite*. Then the linearity of f yields $f(v) = \sum_{i\in I_v} c^i f(e_i)$. (This fact holds true no matter if $dim V < +\infty$). So, formally one may use the notation

$$f := \sum_{i\in I} f(e_i) e^{*i}$$

to expand any element $f \in V^*$ also if $\{e^{*i}\}_{i\in I}$ is not a vector basis of V^* in the proper sense (the linear combination in the right-hand side may be infinite). This can be done provided that he/she adopts the convention that, for $v = \sum_{k\in I} v^k e_k \in V$,

$$\sum_{i\in I} f(e_i) e^{*i} \sum_{j\in I} v^j e_j := \sum_{k\in I} f(e_k) v^k\ .$$

Notice that the last sum is always finite, so it makes sense. ■

Notation 2.7. From now on we take advantage of the following convention. Whenever an index appears twice, once as an upper index and once as a lower index, in whichever expression, the summation over the values of that index is understood in the notation. E.g.,

$$t_{ijkl} f^{rsil}$$

means
$$\sum_{i,l} t_{ijkl} f^{rsil} .$$
The range of the various indices should be evident from the context or otherwise specified. ■

We are naturally lead to consider the following issue. It could seem that the definition of dual space, V^* (of a vector space V) may be implemented on V^* itself producing the double dual space $(V^*)^*$ and so on, obtaining for instance $((V^*)^*)^*$ and, by that way, an infinite sequence of dual vector spaces. The theorem below shows that, *in the finite-dimensional case*, this is not the case because $(V^*)^*$ turns out to be *naturally isomorphic* to the initial space V and thus the apparently infinite sequence of dual spaces ends on the second step. We remind the reader that a vector space isomorphism $F : V \to V'$ is a linear map which is also one-to-one, i.e., injective and surjective. An isomorphism is called *natural* when it is built up using the definition of the involved algebraic structures only and it does not depend on "arbitrary choices". A more precise definition of *natural isomorphism* may be given by introducing the *theory of mathematical categories* and using the notion of *natural transformation* (between the identity functor in the category of finite dimensional vector spaces on $\mathbb{K}$ and the functor induced by F in the same category).

Theorem 2.8. *Let V be a vector space on the field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. The following holds.*
(a) *There is an injective linear map $F : V \to (V^*)^*$ given by*
$$(F(v))\,(u) := u(v) ,$$
for all $u \in V^*$ and $v \in V$, so that V identifies naturally with a subspace of $(V^*)^*$.
(b) *If V has finite dimension, the map F is a natural isomorphism and V identifies naturally with the whole space $(V^*)^*$*

Proof. Notice that $F(v) \in (V^*)^*$ because it is a linear functional on V^*:
$$(F(v))(\alpha u + \beta u') := (\alpha u + \beta u')(v) = \alpha u(v) + \beta u'(v) =: \alpha(F(v))(u) + \beta(F(v))(u') .$$
Let us prove that $v \mapsto F(v)$ with $(F(v))\,(u) := u(v)$ is linear and injective in the general case, and surjective when V has finite dimension.
(Linearity.) We have to show that, for all $\alpha, \beta \in \mathbb{K}$, $v, v' \in V$,
$$F(\alpha v + \beta v') = \alpha F(v) + \beta F(v') .$$
This is equivalent to, by the definition of F given above,
$$u(\alpha v + \beta v') = \alpha u(v) + \beta u(v') ,$$
for all $u \in V^*$. This is obvious because, u is a linear functional on V.
(Injectivity.) Due to linearity it is enough to show that $F(v) = 0$ implies $v = 0$. $(F(v))\,(u) = 0$

can be re-written as $u(v) = 0$. In our hypotheses, this holds true for all $u \in V^*$. Then, define $e_1 := v$ and notice that, if $v \neq 0$, one can complete e_1 with other vectors to get an algebraic vector basis of V, $\{e_i\}_{i \in I}$ (I being infinite in general). Since u is arbitrary, we can pick out u as the unique element of V^* such that $u(e_1) = 1$ and $u(e_i) = 0$ if $i \neq 1$. (This is possible also if V – and thus V^* – is not finite-dimensional.) It contradicts the hypothisis $u(e_1) = u(v) = 0$. By consequence $v = 0$.
(Surjectivity.) Assume that V (and thus V^*) has finite dimension. Since $dim((V^*)^*) = dim(V^*)$ (V^* having finite dimension) and so $dim((V^*)^*) = dimV < +\infty$, injectivity implies surjectivity. However it is interesting to give and explicit proof. We have to show that if $f \in (V^*)^*$, there is $v_f \in V$ such that $F(v_f) = f$.
Fix a basis $\{e_i\}_{i \in I}$ in V and the dual one in V^*. Since $\{e^{*i}\}_{i \in I}$ is a (proper) vector basis of V^* (it may be false if I is infinite!), $f \in (V^*)^*$ is completely defined by the coefficients $f(e^{*i})$. Then $v_f := f(e^{*i})e_i$ fulfills the requirement $F(v_f) = f$. $\square$

Definition 2.9. **(Pairing.)** Let V be a vector space on $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$ with dual space V^* and conjugate dual space $\overline{V^*}$ when $\mathbb{K} = \mathbb{C}$. The bi-linear map $\langle\, .,\, . \rangle : V \times V^* \to \mathbb{K}$ such that

$$\langle u, v \rangle := v(u)$$

for all $u \in V$, $v \in V^*$, is called **pairing**.
If $\mathbb{K} = \mathbb{C}$, the map, linear in the right-hand entry and anti linear in the left-hand entry $\langle\, .,\, . \rangle : V \times \overline{V^*} \to \mathbb{C}$ such that

$$\langle u, v \rangle := v(u)$$

for all $u \in V$, $v \in \overline{V^*}$, is called **(conjugate) pairing**. $\blacksquare$

Remarks 2.10. Because of the theorem proved above, we may indifferently think $\langle u, v \rangle$ as representing either the action of $u \in (V^*)^*$ on $v \in V^*$ or the action of $v \in V^*$ on $u \in V$. The same happens for V and $\overline{V^*}$ $\blacksquare$

Notation 2.11. From now on

$$V \simeq W$$

indicates that the vector spaces V and W are isomorphic under some *natural* isomorphism. If the field of V and W is $\mathbb{C}$,

$$V \overline{\simeq} W$$

indicates that there is a *natural* anti-isomorphism, i.e. there is an injective, surjective, anti-linear map $G : V \to W$ built up using the abstract definition of vector space (including the abstract definitions of dual vector space and (dual) conjugate vector space). $\blacksquare$

We finally state a theorem, concerning conjugated spaces, that is analogous to theorem 2.8 and with a strongly analogous proof. The proof is an exercise left to the reader.

Theorem 2.12. *If V is a vector space with finite dimension on the field $\mathbb{C}$, one has:*
(i) $V^* \simeq \overline{V^*}$, where the anti-isomorphism $G : V^* \to \overline{V^*}$ is defined by

$$(G(v))(u) := \overline{\langle u, v\rangle} \quad \textit{for all } v \in V^* \textit{ and } u \in V;$$

(ii) $V \simeq \overline{V}$, where the anti-isomorphism $F : V \to \overline{V}$ is defined by

$$(F(v))(u) := \langle v, G^{-1}(u)\rangle \quad \textit{for all } v \in V \textit{ and } u \in V^*;$$

(iii) $\overline{V}^* \simeq \overline{V^*}$, where the involved isomorphism $H : \overline{V}^* \to \overline{V^*}$ arises by applying theorem 2.8 to the definition of $\overline{V}$.
Finally, with respect to any fixed basis $\{e_i\}_{i\in I} \subset V$, and the canonically associated bases $\{\overline{e}_i\}_{i\in I} \subset \overline{V}$, $\{e^{*i}\}_{i\in I} \subset V^*$, $\{\overline{e^*}^i\}_{i\in I} \subset \overline{V^*}$, $\{\overline{e}^{*i}\}_{i\in I} \subset \overline{V}^*$, one also has

$$F : v^i e_i \mapsto \overline{v^i}\overline{e}_i\ , \quad G : v_i e^{*i} \mapsto \overline{v_i}\overline{e^*}^i\ , \quad H : v_i \overline{e^*}^i \mapsto v_i \overline{e}^{*i}$$

where the bar over the components denotes the complex conjugation.

Exercises 2.13.

1. Show that if $v \in V$ then $v = \langle v, e^{*j}\rangle e_j$, where $\{e_j\}_{j\in I}$ is any basis of the finite dimensional vector space V.
(*Hint*. Decompose $v = c^i e_i$, compute $\langle v, e^{*k}\rangle$ taking the linearity of the the left entrance into account. Alternatively, show that $v - \langle v, e^{*j}\rangle e_j = 0$ proving that $f(v - \langle v, e^{*j}\rangle e_j) = 0$ for every $f \in V^*$.)

2. Show that $V^* \simeq \overline{V^*}$ if $dim V < \infty$ (and the field of V is $\mathbb{C}$). Similarly, show that $V \simeq \overline{V}$ under the same hypotheses. where the anti-isomorphism $G : V^* \to \overline{V^*}$, is defined by $(G(v))(u) := \overline{\langle u, v\rangle}$ and the anti-isomorphism $F : V \to \overline{V}$, is defined by $(F(v))(u) := \langle v, G^{-1}(u)\rangle$.)

3. Show that if the finite-dimensional vector spaces V and V' are isomorphic or anti-isomorphic, then V^* and V'^* are isomorphic or anti-isomorphic respectively.
(*Hint*. If the initial (anti-) isomorphism is $F : V \to V'$ consider $G : V'^* \to V^*$ defined by $\langle F(u), v'\rangle = \langle u, G(v')\rangle$.) ■

The last notion we go to introduce is that of *adjoint operator*. We shall have very few occasions to employ this notion, however it is an important mathematical tool and it deserves mention. Consider a (linear) operator $T : V_1 \to V_2$, where V_1, V_2 are two linear spaces with the same field $\mathbb{K}$. One can define another operator: $T^* : V_2^* \to V_1^*$ completely determined by the requirement:

$$(T^* u^*)(v) := u^*(Tv)\ , \quad \text{for all } v \in V_1 \text{ and all } u^* \in V_2^*.$$

It is obvious that T^* is linear by definition.

Definition 2.14. **(Adjoint operator.)** Let V_1, V_2 be a pair of vector spaces with the same field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$ and finite dimension. If $T : V_1 \to V_2$ is any linear operator, the operator $T^* : V_2^* \to V_1^*$ completely defined by the requirement

$$\langle v, T^* u^*\rangle_1 := \langle Tv, u^*\rangle_2\ , \quad \text{for all } v \in V_1 \text{ and all } u^* \in V_2^*,$$

is called **adjoint operator** of T. ■

Remarks 2.15. Notice that we may drop the hypothesis of finite dimension of the involved spaces in the definition above without troubles: T^* would turn out to be well-defined also in that case. However this is not the full story, since such a definition, in the infinite-dimensional case would not be very useful for applications. In fact, a generalized definition of the adjoint operator can be given when V_1, V_2 are infinite dimensional and equipped with a suitable topology. In this case T is promoted to a continuous operator and thus one expects that T^* is continuous as well. This is not the case with the definition above as it stands: In general T^* would not be continuous also if starting with T continuous since the algebraic duals V_1^* and V_2^* are not equipped with any natural topology. However if one replaces the algebraic duals with *topological duals*, the definition abolve gives rise to a continuous operator T^* when T is continuous. It happens, in particular, whenever V_1 and V_2 are a Banach spaces.
In this scenario, there is another definition of adjoint operator, in the context of vector spaces with scalar product (including Hilbert spaces). It is worth stressing that the adjoint operator in the sense of Hilbert space theory is different form the adjoint in the sense of definition 2.14. However the two notions enjoy a nice interplay we shall discuss in the remarks 7.2 later. ■

2.2 Multi linearity: tensor product, tensors, universality theorem

2.2.1 Tensors as multi linear maps

Let us consider $n \geq 1$ vector spaces $V_1, \ldots, V_n$ on the common field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ and another vector space W on $\mathbb{K}$, all spaces are not necessarily finite-dimensional. In the following $\mathscr{L}(V_1, \ldots, V_n|W)$ denotes the vector space (on $\mathbb{K}$) of *multi-linear maps* from $V_1 \times \ldots \times V_n$ to W. We remind the reader that a map $f : V_1 \times \ldots \times V_n \to W$ is said to be *multi linear* if, arbitrarily fixing $n-1$ vectors, $v_i \in V_i$ for $i \neq k$, each map $v_k \mapsto f(v_1, \ldots, v_k, \ldots, v_n)$ is linear for every $k = 1, \ldots, n$. We leave to the reader the trivial proof of the fact that $\mathscr{L}(V_1, \ldots, V_n|W)$ is a *vector space* on $\mathbb{K}$, with respect to the usual sum of pair of functions and product of an element of $\mathbb{K}$ and a function. If $W = \mathbb{K}$ we use the shorter notation $\mathscr{L}(V_1, \ldots, V_n) := \mathscr{L}(V_1, \ldots, V_n|\mathbb{K})$, $\mathbb{K}$ being the common field of $V_1, \ldots, V_n$.

Exercises 2.16.

1. Suppose that $\{e_{k,i_k}\}_{i_k \in I_k}$ are bases of thel vector spaces V_k, $k = 1, \ldots, n$ on the same field $\mathbb{K}$. Suppose also that $\{e_i\}_{i \in I}$ is a basis of the vector space W on $\mathbb{K}$. Show that each $f \in \mathscr{L}(V_1, \ldots, V_n|W)$ satisfies

$$f(v_1, \ldots, v_n) = v_1^{i_1} \cdots v_n^{i_n} \langle f(e_{1,i_1}, \ldots, e_{n,i_n}), e^{*i} \rangle e_i \,,$$

for all $(v_1, \ldots, v_n) \in V_1 \times \cdots \times V_n$.

2. Endow $V_1 \times \cdots \times V_n$ with the structure of vector space over the common field $\mathbb{K}$, viewed as the direct sum of $V_1, \ldots, V_n$. In other words, that vector space structure is the unique compatible with the definition:

$$\alpha(v_1, \ldots, v_n) + \beta(v'_1, \ldots, v'_n) := (\alpha v_1 + \beta v'_1, \ldots, \alpha v_n + \beta v'_n)$$

for every numbers $\alpha, \beta \in \mathbb{K}$ and vectors $v_i, v'_i \in V_i$, $i = 1, \ldots, n$.
If $f \in \mathscr{L}(V_1, \ldots, V_n|W)$, may we say that f is a linear map from $V_1 \times \cdots \times V_n$ to W? Generally speaking, is the range of f a subspace of W?
(*Hint*. No. Is this identity $f(\alpha(v_1, \ldots, v_n)) = \alpha f(v_1, \ldots, v_n)$ true? The answer to the second question is no similarly. This is already evident, for instance, for $W = \mathbb{R}^2$ and $V_1 = V_2 = \mathbb{R}$ considering the multi linear map $f : \mathbb{R} \times \mathbb{R} \ni (a, b) \mapsto (a \cdot b, a \cdot b) \in \mathbb{R}^2$.)

We have a first foundamental and remarkable theorem. To introduce it, we pass through three steps.
(1) Take $f \in \mathscr{L}(V_1, ..., V_n|W)$. This means that f associates every string $(v_1, ..., v_n) \in V_1 \times \cdots \times V_n$ with a corresponding element $f(v_1, \ldots, v_n) \in W$, and the map

$$(v_1, \ldots, v_n) \mapsto f(v_1, \ldots, v_n)$$

is multi-linear.
(2) Since, for *fixed* $(v_1, \ldots, v_n)$, the vector $f(v_1, \ldots, v_n)$ is an element of W, the action of $w^* \in W^*$ on $f(v_1, \ldots, v_n)$ makes sense, producing $\langle f(v_1, \ldots, v_n), w^* \rangle \in \mathbb{K}$.
(3) Allowing $v_1, \ldots v_n$ and w^* to range freely in the corresponding spaces, the map Ψ_f with

$$\Psi_f : (v_1, \ldots, v_n, w^*) \mapsto \langle f(v_1, \ldots, v_n), w^* \rangle \,,$$

turns out to be multi-linear by construction. Hence, by definition $\Psi_f \in \mathscr{L}(V_1, \ldots, V_n, W^*)$.
The theorem concerns the map F which associates f with Ψ_f. ■

Theorem 2.17. *Let $V_1, \ldots, V_n$ be (not necessarily finite-dimensional) vector spaces on the common field $\mathbb{K}$ ($= \mathbb{C}$ or $\mathbb{R}$), and W is another finite-dimensional vector spaces on $\mathbb{K}$.*
The vector spaces $\mathscr{L}(V_1, \ldots, V_n|W)$ and $\mathscr{L}(V_1, \ldots, V_n, W^*)$ are naturally isomorphic by means of the map $F : \mathscr{L}(V_1, \ldots, V_n|W) \to \mathscr{L}(V_1, \ldots, V_n, W^*)$ with $F : f \mapsto \Psi_f$ defined by

$$\Psi_f(v_1, \ldots, v_n, w^*) := \langle f(v_1, \ldots, v_n), w^* \rangle \,, \quad \textit{for all } (v_1, \ldots, v_n) \in V_1 \times \ldots \times V_n \textit{ and } w^* \in W^*.$$

If $\dim W = +\infty$, the above-defined linear map F is injective in any cases.

Proof. Let us consider the map F defined above. We have only to establish that F is linear, injective and surjective. This ends the proof.
(Linearity.) We have to prove that $\Psi_{\alpha f + \beta g} = \alpha \Psi_f + \beta \Psi_g$ for all $\alpha, \beta \in \mathbb{K}$ and $f, g \in \mathscr{L}(V_1, \ldots, V_n|W)$.
In fact, making use of the left-hand linearity of the pairing, one has

$$\Psi_{\alpha f+\beta g}(v_1, \ldots, v_n, w^*) = \langle (\alpha f + \beta g)(v_1, \ldots, v_n), w^* \rangle = \alpha \langle f(v_1, \ldots, v_n), w^* \rangle + \beta \langle g(v_1, \ldots, v_n), w^* \rangle$$

and this is nothing but:

$$\Psi_{\alpha f+\beta g}(v_1,\ldots,v_n,w^*) = (\alpha\Psi_f + \beta\Psi_g)(v_1,\ldots,v_n,w^*)\,.$$

Since $(v_1,\ldots,v_n,w^*)$ is arbitrary, the thesis is proved.
(Injectivity.) Due to linearity we have only to show that if $\Psi_f = 0$ then $f = 0$.
In fact, if $\Psi_f(v_1,\ldots,v_n,w^*) = 0$ for all $(v_1,\ldots,v_n,w^*)$, using the definition of Ψ_g we have $\langle f(v_1,\ldots,v_n),w^*\rangle = 0$, for all $(v_1,\ldots,v_n)$ and w^*. Then define $e_1 := f(v_1,\ldots,v_n)$, if $e_1 \neq 0$ we can complete it to a basis of W. Fixing $w^* = e^{*1}$ we should have

$$0 = \Psi_f(v_1,\ldots,v_n,w^*) = \langle f(v_1,\ldots,v_n),w^*\rangle = 1$$

which is impossible. Therefore $f(v_1,\ldots,v_n) = 0$ for all $(v_1,\ldots,v_n) \in V_1\times\ldots\times V_n$, in other words $f = 0$.
(Surjectivity.) We have to show that for each $\Phi \in \mathscr{L}(V_1,\ldots,V_n,W^*)$ there is a $f_\Phi \in \mathscr{L}(V_1,\ldots,V_n|W)$ with $\Psi_{f_\Phi} = \Phi$. To this end fix a basis $\{e_k\}_{k\in I} \subset W$ and the associated dual one $\{e^{*k}\}_{k\in I} \subset W^*$. Then, take $\Phi \in \mathscr{L}(V_1,\ldots,V_n,W^*)$ and define the map $f_\Phi \in \mathscr{L}(V_1,\ldots,V_n|W)$ given by

$$f_\Phi(v_1,\ldots,v_n) := \Phi(v_1,\ldots,v_n,e^{*k})e_k\,.$$

By construction that map is multi-linear and, using multilinearity we find

$$\Psi_{f_\Phi}(v_1,\ldots,v_n,w^*) = \langle\Phi(v_1,\ldots,v_n,e^{*k})e_k, w^*_i e^{*i}\rangle = \Phi(v_1,\ldots,v_n,w^*_i e^{*k})\langle e_k,e^{*i}\rangle = \Phi(v_1,\ldots,v_n,w^*),$$

for all $(v_1,\ldots,v_n,w^*)$ and this is equivalent to $\Psi_{f_\Phi} = \Phi$.
If $\dim W = +\infty$ the proof of surjectivity may be false since the sum over $k \in I$ used to define f_Φ may diverge. The proofs of linearity and injectivity are independent form the cardinality of I. $\Box$

Remarks 2.18. An overall difficult point in understanding and trusting in the statement of the theorem above relies upon the fact that the function f has n entries, whereas the function Ψ_f has $n+1$ entries but f is identified to Ψ_f by means of F. This fact may seem quite weird at first glance and the statement of the theorem may seem suspicious by consequence. The difficulty can be clarified from a practical point of view as follows.
If bases $\{e_{1,i_1}\}_{i_1\in I_1}$,...,$\{e_{n,i_n}\}_{i_n\in I_n}$, for V_1,...,V_n respectively, and $\{e_k\}_{k\in I}$ for W are fixed and $\Psi \in \mathscr{L}(V_1,\ldots,V_n,W^*)$, one has, for constant coefficients $P^k_{i_1\ldots i_n}$ depending on Ψ,

$$\Psi(v_1,\ldots,v_n,w^*) = P^k_{i_1\ldots i_n} v_1^{i_1}\ldots v_n^{i_n} w^*_k\,,$$

where w^*_k are the components of w^* in the dual basis $\{e^{*k}\}_{k\in I}$ and $v_p^{i_p}$ the components of v_p in the basis $\{e_{p,i_p}\}_{i_p\in I_p}$. Now consider the other map $f \in \mathscr{L}(V_1,\ldots,V_n|W)$ whose argument is the string $(v_1,\ldots,v_n)$, *no further vectors being necessary*:

$$f : (v_1,\ldots,v_n) \mapsto P^k_{i_1\ldots i_n} v_1^{i_1}\ldots v_n^{i_n} e_k$$

The former map Ψ, which deals with $n+1$ arguments, is associated with f which deals with n arguments. The point is that we have taken advantage of the bases, $\{e_k\}_{k\in I_k}$ in particular. Within this context, theorem 2.17 just proves that *the correspondence which associates Ψ to f is linear, bijective and independent from the chosen bases.* ■

In the finite-dimensional case – and we are mainly interested in this case – Theorem 2.17 allow us to restrict our study to the spaces of multi-linear *functionals* $\mathscr{L}(V_1, \ldots, V_k)$, since the spaces of multi-linear *maps* are completely encompassed. Let us introduce the concept of *tensor product* of vector spaces. The following definitions can be extended to encompass the case of non-finite dimensional vector spaces by introducing suitable topological notions (e.g., Hilbert spaces).

Definition 2.19. **(Tensor product.)** Let $U_1, \ldots, U_n$ be $n \geq 1$ vector spaces (not necessarily finite-dimensional) on the common field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.
(1) if $(u_1, \ldots, u_n) \in U_1 \times \ldots \times U_n$, $u_1 \otimes \ldots \otimes u_n$ denotes the multi linear map in $\mathscr{L}(U_1^*, \ldots, U_n^*)$ defined by

$$(u_1 \otimes \ldots \otimes u_n)(v_1, \ldots, v_n) := \langle u_1, v_1\rangle \cdots \langle u_n, v_n\rangle \ ,$$

for all $(v_1, \ldots, v_n) \in U_1^* \times \ldots \times U_n^*$.
$u_1 \otimes \ldots \otimes u_n$ is called **tensor product of vectors** $u_1, \ldots, u_n$.
(2) The map $\otimes : U_1 \times \ldots \times U_n \to \mathscr{L}(U_1^*, \ldots, U_n^*)$ given by: $\otimes : (u_1, \ldots, u_n) \mapsto u_1 \otimes \ldots \otimes u_n$, is called **tensor product map**.
(3) The vector subspace of $\mathscr{L}(U_1^*, \ldots, U_n^*)$ *generated* by all of $u_1 \otimes \ldots \otimes u_n$ for all $(u_1, \ldots, u_n) \in U_1 \times \ldots \times U_n$ is called **tensor product of spaces** $U_1, \ldots, U_n$ and is indicated by $U_1 \otimes \ldots \otimes U_n$. The vectors in $U_1 \otimes \ldots \otimes U_n$ are called **tensors**. ■

Remarks.
(1) $U_1 \otimes \ldots \otimes U_n$ is therefore made of all the *linear combinations* of the form

$$\sum_{j=1}^{N} \alpha_j u_{1,j} \otimes \ldots \otimes u_{n,j} \ ,$$

where $\alpha_j \in \mathbb{K}$, $u_{k,j} \in U_k$ and $N = 1, 2, \ldots$.
(2) It is trivially proved that the tensor product map:

$$(u_1, \ldots, u_n) \mapsto u_1 \otimes \ldots \otimes u_n \ ,$$

is *multi linear*.
That is, for any $k \in \{1, 2, \ldots, n\}$, the following identity holds for all $u, v \in U_k$ and $\alpha, \beta \in \mathbb{K}$:

$$u_1 \otimes \ldots \otimes u_{k-1} \otimes (\alpha u + \beta v) \otimes u_{k+1} \otimes \ldots \otimes u_n$$

$$= \alpha(u_1 \otimes \ldots \otimes u_{k-1} \otimes u \otimes u_{k+1} \otimes \ldots \otimes u_n)$$

$$+\beta(u_1 \otimes \ldots \otimes u_{k-1} \otimes v \otimes u_{k+1} \otimes \ldots \otimes u_n) \ .$$

As a consequence, it holds

$$(\alpha u) \otimes v = u \otimes (\alpha v) = \alpha(u \otimes v) ,$$

and similar identities hold considering whichever number of factor spaces in a tensor product of vector spaces.

(3) From the given definition, if $U_1, \ldots, U_n$ are given vector spaces on the same field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, what we mean by $U_1 \otimes \ldots \otimes U_n$ or $U_1^* \otimes \ldots \otimes U_n^*$ but also, for instance, $U_1^* \otimes U_2 \otimes U_3$ or $U_1 \otimes U_2^* \otimes U_3^*$ should be obvious. One simply has to take the action of U^* on $(U^*)^*$ into account.

By definition $U_1 \otimes \ldots \otimes U_n \subset \mathscr{L}(U_1^*, \ldots, U_n^*)$, more strongly, $U_1 \otimes \ldots \otimes U_n$ is, in general, a *proper* subspace of $\mathscr{L}(U_1^*, \ldots, U_n^*)$. The natural question which arises is if there are hypotheses which entail that $U_1 \otimes \ldots \otimes U_n$ coincides with the whole space $\mathscr{L}(U_1^*, \ldots, U_n^*)$. The following theorem gives an answer to that question. ■

Theorem 2.20. *If all the spaces U_i have finite dimension, one has:*

$$U_1 \otimes \ldots \otimes U_n = \mathscr{L}(U_1^*, \ldots, U_n^*) .$$

Proof. It is sufficient to show that if $f \in \mathscr{L}(U_1^*, \ldots, U_n^*)$ then $f \in U_1 \otimes \ldots \otimes U_n$. To this end fix vector bases $\{e_{k,i}\}_{i \in I_k} \subset U_k$ for $k = 1, \ldots, n$ and consider also the associated dual bases $\{e_k^{*i}\}_{i \in I_k}$. f above is completely determined by coefficients (their number is finite since $dim U_i^* = dim U_i < +\infty$ by hypotheses!) $f_{i_1,\ldots,i_n} := f(e_1^{*i_1}, \ldots, e_n^{*i_n})$. Every $v_k \in U_k^*$ can be decomposed as $v_k = v_{k,i_k} e_k^{*i_k}$, where each sum is finite, and thus, by multi linearity:

$$f(v_1, \ldots, v_n) = v_{1,i_1} \cdots v_{n,i_n} f(e_1^{*i_1}, \ldots, e_n^{*i_n}) .$$

Then consider the tensor $t_f \in U_1 \otimes \ldots \otimes U_n$ defined by:

$$t_f := f(e_1^{*i_1}, \ldots, e_n^{*i_n}) e_{1,i_1} \otimes \ldots \otimes e_{n,i_n} .$$

Then, by def.2.19, one can directly prove that, by multi linearity

$$t_f(v_1, \ldots, v_n) = v_{1,i_1} \cdots v_{n,i_n} f(e_1^{*i_1}, \ldots, e_n^{*i_n}) = f(v_1, \ldots, v_n) ,$$

for all of $(v_1, \ldots, v_n) \in U_1^* \times \ldots \times U_n^*$. This is nothing but $t_f = f$. □

Another relevant result is stated by the theorem below.

Theorem 2.21. *Consider vector spaces U_i, $i = 1, \ldots, n$ with the same field $\mathbb{K}$. The following statements hold.*

(a) *If all spaces U_i are finite-dimensional, the dimension of $U_1 \otimes \ldots \otimes U_n$ is:*

$$dim(U_1 \otimes \ldots \otimes U_n) = \prod_{k=1}^{n} dim\, U_k .$$

(b) *If* $\{e_{k,i_k}\}_{i_k\in I_k}$ *is a basis of* U_k, $k=1,\ldots,n$, *then* $\{e_{1,i_1}\otimes\ldots\otimes e_{n,i_n}\}_{(i_1,\ldots,i_n)\in I_1\times\ldots\times I_n}$ *is a vector basis of* $U_1\otimes\ldots\otimes U_n$.
(c) *If* $t\in U_1\otimes\ldots\otimes U_n$ *and all spaces* U_i *are finite-dimensional, the components of* t *with respect to a basis* $\{e_{1,i_1}\otimes\ldots\otimes e_{n,i_n}\}_{(i_1,\ldots,i_n)\in I_1\times\ldots\times I_n}$ *are given by:*

$$t^{i_1\ldots i_n}=t(e_1^{*i_1},\ldots,e_n^{*i_n})$$

and thus it holds

$$t=t(e_1^{*i_1},\ldots,e_n^{*i_n})e_{1,i_1}\otimes\ldots\otimes e_{n,i_n}\,.$$

Proof. Notice that, in the finite-dimensional case, **(b)** trivially implies **(a)** because the elements $e_{1,i_1}\otimes\ldots\otimes e_{n,i_n}$ are exactly $\prod_{k=1}^n dimU_k$. So it is sufficient to show that the second statement holds true. To this end, since elements $e_{1,i_1}\otimes\ldots\otimes e_{n,i_n}$ are generators of $U_1\otimes\ldots\otimes U_n$, it is sufficient to show that they are linearly independent. Consider the generic vanishing linear combination

$$C^{i_1\ldots i_n}e_{1,i_1}\otimes\ldots\otimes e_{n,i_n}=0\,,$$

We want to show that all of the coefficients $C^{i_1\ldots i_n}$ vanish. The action of that linear combination of multi-linear functionals on the generic element $(e_1^{*j_1},\ldots,e_n^{*j_n})\in U_1^*\times\cdots\times U_n^*$ produces the result

$$C^{j_1\ldots j_n}=0\,.$$

Since we can arbitrarily fix the indices $j_1,\ldots j_n$ this proves the thesis.
The proof of **(c)** has already be done in the proof of theorem 2.20. However we repeat it for sake of completeness. By uniqueness of the components of a vector with respect to a basis, it is sufficient to show that, defining (notice that the number of vectors $e_{1,i_1}\otimes\ldots\otimes e_{n,i_n}$ is finite by hypotheses since $dimU_k$ is finite in the considered case)

$$t':=t(e_1^{*i_1},\ldots,e_n^{*i_n})e_{1,i_1}\otimes\ldots\otimes e_{n,i_n}\,,$$

it holds

$$t'(v_1,\ldots,v_n)=t(v_1,\ldots,v_n)\,,$$

for all $(v_1,\ldots,v_n)\in U_1^*\times\ldots\times U_n^*$. By multi linearity,

$$t(v_1,\ldots,v_n)=v_{1\,i_1}\cdots v_{n\,i_n}t(e_1^{*i_1},\ldots,e_n^{*i_n})=t'(v_1,\ldots,v_n)\,.$$

$\square$

We notice that, obviously, if some of the spaces U_i has infinite dimension $U_1\otimes\ldots\otimes U_n$ cannot have finite dimension.

In case of *finite dimension* of all involved spaces, there is an important result which, *together with the identification of* V *and* $(V^*)^*$, imply that all of the spaces which one may build up using the symbols $\otimes$, V_k, * and () coherently are naturally isomorphic to spaces which are of the form

$V_{i_1}^{(*)} \otimes \ldots \otimes V_{i_n}^{(*)}$. The rule to produce spaces naturally isomorphic to a given initial space is that one has to (1) ignore parentheses – that is $\otimes$ *is associative* –, (2) assume that * is *distributive with respect to* $\otimes$ and (3) assume that * is *involutive* (i.e. $(X^*)^* \simeq X$). For instance, one has:

$$((V_1^* \otimes V_2) \otimes (V_3 \otimes V_4^*))^* \simeq V_1 \otimes V_2^* \otimes V_3^* \otimes V_4 \,.$$

Let us state the theorems corresponding to the rules (2) and (1) respectively (the rule (3) being nothing but theorem 2.8).
If not every V_i has finite dimension the distributivity and involutivity properties of * may fail to be fulfilled, but associativity of $\otimes$ halds true anyway.

Theorem 2.22. *If $V_1 \ldots V_n$ are finite-dimensional vector spaces on the same field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, * is* **distributive** *with respect to $\otimes$ by means of* **natural isomorphisms** $F : V_1^* \otimes \ldots \otimes V_n^* \to (V_1 \otimes \ldots \otimes V_n)^*$. *In other words it holds:*

$$(V_1 \otimes \ldots \otimes V_n)^* \simeq V_1^* \otimes \ldots \otimes V_n^* \,.$$

under F. The isomorphism F is uniquely determined by the requirement

$$\langle u_1 \otimes \ldots \otimes u_n, F(v_1^* \otimes \ldots \otimes v_n^*)\rangle = \langle u_1, v_1^*\rangle \cdots \langle u_n, v_n^*\rangle \,, \tag{2.1}$$

for every choice of $v_i^* \in V_i^*$, $u_i \in V_i$ and $i = 1, 2, \ldots, n$.

Proof. The natural isomorphism is determined by (2.1), thus employing only the general mathematical structures of dual space and tensor product. Consider, *if it exists* (see the discussion below), the linear map $F : V_n^* \otimes \ldots \otimes V_n^* \to (V_1 \otimes \ldots \otimes V_n)^*$ which satisfies, for all $u_1 \otimes \ldots \otimes u_n \in V_1 \otimes \ldots \otimes V_n$,

$$\langle u_1 \otimes \ldots \otimes u_n, F(v_1^* \otimes \ldots \otimes v_n^*)\rangle := \langle u_1, v_1^*\rangle \ldots \langle u_n, v_n^*\rangle \,. \tag{2.2}$$

Using the linearity and the involved definitions the reader can simply prove that the map F is surjective. Indeed, equipping each space V_k with a basis $\{e_{k,i_k}\}_{i_k \in I_k}$ and the dual space V_k^* with the corresponding duel basis $\{e_k^{*i_k}\}_{i_k \in I_k}$, if $f \in (V_1 \otimes \cdots \otimes V_n)^*$, defining the numbers $f_{i_1 \cdots i_n} := f(e_{1,i_1} \otimes \cdots \otimes e_{n,i_n})$, by direct inspection one finds that:

$$\langle u_1 \otimes \ldots \otimes u_n, F(f_{i_1 \cdots i_n} e_1^{*i_1} \otimes \cdots \otimes e_n^{*i_n})\rangle = \langle u_1 \otimes \ldots \otimes u_n, f\rangle \,,$$

for all $u_1 \otimes \ldots \otimes u_n$. By linearity this result extends to any element $u \in V_1 \otimes \ldots \otimes V_n$ in place of $u_1 \otimes \ldots \otimes u_n$ so that $f(u) = \left(F(f_{i_1 \cdots i_n} e_1^{*i_1} \otimes \cdots \otimes e_n^{*i_n})\right)(u)$ for every $u \in V_1 \otimes \cdots \otimes V_n$. Therefore the functionals $f, F(f_{i_1 \cdots i_n} e_1^{*i_1} \otimes \cdots \otimes e_n^{*i_n}) \in (V_1 \otimes \ldots \otimes V_n)^*$ concide:

$$f = F(f_{i_1 \cdots i_n} e_1^{*i_1} \otimes \cdots \otimes e_n^{*i_n}) \,.$$

Finally, by the theorems proved previously,

$$dim(V_1 \otimes \cdots \otimes V_n)^* = dim(V_1 \otimes \cdots \otimes V_n) = dim(V_1^* \otimes \cdots \otimes V_n^*)$$

As a consequence, F must be also injective and thus it is an isomorphism. $\square$

Theorem 2.23. *The tensor product is* **associative** *by means of* **natural isomorphisms**. *In other words considering a space which is made of tensor products of vector spaces on the same field $\mathbb{R}$ or $\mathbb{C}$, one may omit parenthesis everywhere obtaining a space which is naturally isomorphic to the initial one. So, for instance*

$$V_1 \otimes V_2 \otimes V_3 \simeq (V_1 \otimes V_2) \otimes V_3 \,.$$

Where the natural isomorphism $F_1 : V_1 \otimes V_2 \otimes V_3 \to (V_1 \otimes V_2) \otimes V_3$ satisfies

$$F_1 : v_1 \otimes v_2 \otimes v_3 \to (v_1 \otimes v_2) \otimes v_3$$

for every choice of $v_i \in V_i$, $i = 1, 2, 3$.
Consequently, if the space $\{e_{i,j_i}\}_{j_i \in J_i} \subset V_i$ is a basis for $i = 1, 2, 3$, the action of F can be written down as:

$$F : t^{j_1 j_2 j_3} e_{1,j_1} \otimes e_{2,j_2} \otimes e_{3,j_3} \mapsto t^{j_1 j_2 j_3} (e_{1,j_1} \otimes e_{2,j_2}) \otimes e_{3,j_3} \,.$$

Proof. The natural isomorphism is constructed as a linear mapping, *if it exists* (see discussion below), which linearly extends the action $F_1 : V_1 \otimes V_2 \otimes V_3 \to (V_1 \otimes V_2) \otimes V_3$ such that

$$F_1 : v_1 \otimes v_2 \otimes v_3 \mapsto (v_1 \otimes v_2) \otimes v_3 \,,$$

for all $v_1 \otimes v_2 \otimes v_3 \in V_1 \otimes V_2 \otimes V_3$.
Injectivity and surjectiviy can be proved by considering the linear map *if it exists* (see discussion below), $F'_1 : (V_1 \otimes V_2) \otimes V_3 \to V_1 \otimes V_2 \otimes V_3$ such that

$$F'_1 : (v_1 \otimes v_2) \otimes v_3 \mapsto v_1 \otimes v_2 \otimes v_3 \,,$$

for all $(v_1 \otimes v_2) \otimes v_3 \in (V_1 \otimes V_2) \otimes V_3$. By construction $F_1 \circ F'_1 = id_{(V_1 \otimes V_2) \otimes V_3}$ and $F'_1 \circ F_1 = id_{V_1 \otimes V_2 \otimes V_3}$ are the identity maps so that F_1 is linear and bijective. The last statement immediately follows from the defintion of F. $\square$

Remarks 2.24. The central point is that the maps F and F_1 (and F'_1) above have been given by specifying their action on tensor products of elements (e.g, $F(v_1 \otimes \ldots \otimes v_n)$) and not on *linear combinations* of these tensor products of elements. Recall that, for instance $V_1^* \otimes \ldots \otimes V_n^*$ is not the set of products $u_1^* \otimes \ldots \otimes u_n^*$ but it is the set of *linear combinations* of those products. Hence, in order to completely define F and F_1, one must require that F and F_1 admit *uniquely determined linear extensions* on their initial domains in order to encompass the whole tensor spaces generated by linear combinations of simple tensor products. In other words one has to complete the given definition, in the former case, by adding the further requirement

$$F(\alpha\, u_1^* \otimes \ldots \otimes u_n^* + \beta\, v_1^* \otimes \ldots \otimes v_n^*) = \alpha F(u_1^* \otimes \ldots \otimes u_n^*) + \beta F(v_1^* \otimes \ldots \otimes v_n^*) \,,$$

Actually, one has to assume in addition that, for any fixed $v_1^* \otimes \cdots \otimes v_n^*$, there is a unique linear extension, of the application $F(v_1^* \otimes \cdots \otimes v_n^*)$ defined in (2.2), when the argument is a linear combination of vectors of type $u_1 \otimes \cdots \otimes u_n$.
Despite these could seem trivial requirements (and they are, in the present case) they are not trivial in general. The point is that the class of all possible products $v_1^* \otimes \ldots \otimes v_n^*$ is a system of generators of $V_1^* \otimes \ldots \otimes V_n^*$, but it is by no means a basis *because these elements are linearly dependent*. So, in general any attempt to assign a linear application with domain $V_1^* \otimes \ldots \otimes V_n^*$ by assigning the values of it on all the elements $v_1^* \otimes \ldots \otimes v_n^*$ may give rise to contraddictions. For instance, $v_1^* \otimes \ldots \otimes v_n^*$ can be re-written using linear combinations:

$$v_1^* \otimes \ldots \otimes v_n^* = [(v_1^* + u_1^*) \otimes \ldots \otimes v_n^*] - [u_1^* \otimes \ldots \otimes v_n^*] .$$

Now consider the identities, which has to hold as a consequence of the assumed linearity of F, such as:

$$F(v_1^* \otimes \ldots \otimes v_n^*) = F((v_1^* + u_1^*) \otimes \ldots \otimes v_n^*) - F(u_1^* \otimes \ldots \otimes v_n^*)$$

Above $F(v_1^* \otimes \ldots \otimes v_n^*)$, $F(u_1^* \otimes \ldots \otimes v_n^*)$, $F((v_1^* + u_1^*) \otimes \ldots \otimes v_n^*)$ are *independently* defined as we said at the beginning and there is no reason, in principle, for the validity of the constraint:

$$F(v_1^* \otimes \ldots \otimes v_n^*) = F((v_1^* + u_1^*) \otimes \ldots \otimes v_n^*) - F(u_1^* \otimes \ldots \otimes v_n^*) .$$

Similar problems may arise concerning F_1. ■

The general problem which arises by the two considered cases can be stated as follows. Suppose we are given a tensor product of vector spaces $V_1 \otimes \ldots \otimes V_n$ and we are interested in the possible *linear extensions* of a map f on $V_1 \otimes \ldots \otimes V_n$, with values in some vector space W, when f is initially defined on the whole class of the simple products $v_1 \otimes \ldots \otimes v_n \in V_1 \otimes \ldots \otimes V_n$ only. *Is there any general prescription on the specification of values $f(v_1 \otimes \ldots \otimes v_n)$ which assures that f can be extended,* uniquely*, to a linear map from $V_1 \otimes \ldots \otimes V_n$ to W?*
The general obstruction is evident: if the class of all possible tensors $v_1 \otimes \ldots \otimes v_n$ were a basis of $V_1 \otimes \ldots \otimes V_n$, the class of values $f(v_1 \otimes \ldots \otimes v_n)$ would determine a (unique) linear map from $V_1 \otimes \ldots \otimes V_n$ to W. *Instead we have here a set which is much larger than a basis, so assigning independent values thereon for a linear function is impossible, in view of the linearity constraint.*
An answer is given by the following very important **universality theorem** discussed in the following subsection.

2.2.2 Universality theorem and its applications

Theorem 2.25. **(Universality Theorem.)** *Given $n \geq 1$ vector spaces $U_1, \ldots, U_n$ (not necessarily finite-dimensional) on the same field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, the following so-called* **universal property** *holds for the pair $(\otimes, U_1 \otimes \ldots \otimes U_n)$. For any vector space W (not necessarily finite-dimensional) and any multi-linear map $f : U_1 \times \ldots \times U_n \to W$, there is a unique linear map*

$f^{\otimes} : U_1 \otimes \ldots \otimes U_n \to W$ *such that the diagram below* **commutes** *(in other words,* $f^{\otimes} \circ \otimes = f$*).*

$$\begin{tikzcd} U_1 \times \ldots \times U_n \arrow[r, "\otimes"] \arrow[dr, "f"'] & U_1 \otimes \ldots \otimes U_n \arrow[d, "f^{\otimes}"] \\ & W \end{tikzcd}$$

Proof. Fix bases $\{e_{k,i_k}\}_{i_k \in I_k} \subset U_k$, $k = 1, \ldots, n$. By linearity, a linear map $g : U_1 \otimes \ldots \otimes U_n \to W$ is uniquely assigned by fixing the set of vectors

$$\{g(e_{1,i_1} \otimes \cdots \otimes e_{n,i_n})\}_{i_1 \in I_1, \ldots, i_n \in I_n} \subset W \,.$$

This is obviously true also if some of the spaces U_k, W have infinite dimension. Define the linear function $f^{\otimes} : U_1 \otimes \ldots \otimes U_n \to W$ as the unique linear map with

$$f^{\otimes}(e_{1,i_1} \otimes \cdots \otimes e_{n,i_n}) := f(e_{1,i_1}, \ldots, e_{n,i_n}) \,.$$

By construction, using linearity of $f^{\otimes}$ and multi-linearity of f, one gets immediately:

$$f^{\otimes}(v_1 \otimes \ldots \otimes v_n) = f(v_1, \ldots, v_n)$$

for all $(v_1, \ldots, v_n) \in U_1 \times \ldots \times U_n$. In other words $f^{\otimes} \circ \otimes = f$.
The uniqueness of the map $f^{\otimes}$ is obvious: suppose there is another map $g^{\otimes}$ with $g^{\otimes} \circ \otimes = f$ then $(f^{\otimes} - g^{\otimes}) \circ \otimes = 0$. This means in particular that

$$(f^{\otimes} - g^{\otimes})(e_{1,i_1} \otimes \ldots \otimes e_{n,i_n}) = 0 \,,$$

for all $i_k \in I_k$, $k = 1, \ldots, n$. Since the coefficients above completely determine a map, the considered map must be the null map and thus: $f^{\otimes} = g^{\otimes}$. $\square$

Let us now explain how the universality theorem gives a precise answer to the question formulated before the universality theorem. The theorem says that a *linear extension* of any function f with values in W, initially defined on simple tensor products only, $f(v_1 \otimes \ldots \otimes v_n)$ with $v_1 \otimes \ldots \otimes v_n \in U_1 \otimes \ldots \otimes U_n$, does *exist* on the whole domain space $U_1 \otimes \ldots \otimes U_n$ and it is *uniquely* determined provided $f(v_1 \otimes \ldots \otimes v_n) = g(v_1, \ldots, v_n)$ where $g : U_1 \times \ldots \times U_n \to W$ is some *multi-linear function*.
Concerning the maps F and F_1 introduced above, we may profitably use the universality theorem to show that they are well-defined on the whole domain made of linear combinations of simple tensor products. In fact, consider F for example. We can define the *multi-linear* map $G : V_1^* \times \ldots \times V_n^* \to (V_1 \otimes \ldots \otimes V_n)^*$ such that $G(v_1^*, \ldots, v_n^*) : V_1 \otimes \ldots \otimes V_n \to \mathbb{C}$ is the unique linear function (assumed to exist, see below) which linearly extends the requirement

$$(G(v_1^*, \ldots, v_n^*))\,(u_1 \otimes \cdots \otimes u_n) := \langle u_1, v_1^* \rangle \cdots \langle u_n, v_n^* \rangle, \quad \text{for all } (u_1, \ldots, u_n) \in U_1 \times \cdots \times U_n \quad (2.3)$$

to the whole space $V_1 \otimes \ldots \otimes V_n$. Then the universality theorem with $U_k = V_k^*$ and $W = (V_1 \otimes \ldots \otimes V_n)^*$ assures the existence of a linear map $F : V_1^* \otimes \ldots \otimes V_n^* \to (V_1 \otimes \ldots \otimes V_n)^*$ with the required properties because $F := G^{\otimes}$ is such that

$$F(v_1^* \otimes \ldots \otimes v_n^*) = (G^{\otimes} \circ \otimes)(v_1^*, \ldots, v_n^*) = G(v_1^*, \ldots, v_n^*) \,,$$

so that (2.2) is fulfilled due to (2.3). Finally, the existence of the linear function $G(v_1^*, \ldots, v_n^*) : V_1 \otimes \ldots \otimes V_n \to \mathbb{K}$ fulfilling (2.3) can be proved, once again, employing the universality theorem for $U_k := V_k$ and $W := \mathbb{K}$. Starting form the multi linear function $G'(v_1^*, \ldots, v_n^*) : V_1 \times \ldots \times V_n \to \mathbb{C}$ such that

$$\left(G'(v_1^*, \ldots, v_n^*)\right)(u_1, \cdots, u_n) := \langle u_1, v_1^* \rangle \cdots \langle u_n, v_n^* \rangle$$

it is enough to define $G(v_1^*, \ldots, v_n^*) := G'(v_1^*, \ldots, v_n^*)^{\otimes}$.
A similar multi-linear map G_1 can be found for F_1:

$$G_1 : (v_1, v_2, v_3) \mapsto (v_1 \otimes v_2) \otimes v_3 \,.$$

Then: $F_1 := G_1^{\otimes}$ can be used in the proof of theorem 2.23. If V_1, V_2, V_3 have finite dimension, the proof of theorem 2.23 ends because the map F_1 so obtained is trivially surjective and it is also injective since

$$dim((V_1 \otimes V_2) \otimes V_3) = dim(V_1 \otimes V_2) \times dim V_3 = dim(V_1) \times dim(V_2) \times dim(V_3) = dim(V_1 \otimes V_2 \otimes V_3)$$

by Theorem 2.21.
If not every V_1, V_2, V_3 has finite dimension the proof Theorem 2.23 is more complicated and the existence of the function F_1' used in its proof has to be explicitly checked. The construction of the function F_1' used in the proof of theorem 2.23 can be obtained with a little change in the first statement of the universality theorem. With trivial changes one gets that:

Proposition 2.26. *Given $n \geq 1$ vector spaces $U_1, \ldots, U_n$ on the same field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, the following statements hold.*
If $h < k = 2, \ldots, n$ are fixed, for any vector space W and any multi-linear map

$$f : U_1 \times \ldots \times U_n \to W \,,$$

there is a unique linear map

$$f^{\dot{\otimes}} : U_1 \otimes \cdots \otimes U_{h-1} \otimes (U_h \otimes \cdots \otimes U_k) \otimes U_{k+1} \otimes \cdots \otimes U_n \to W$$

such that the diagram below commute.

$$\begin{tikzcd}
U_1 \times \ldots \times U_n \arrow[r, "\dot{\otimes}"] \arrow[dr, "f"'] & U_1 \otimes \cdots \otimes U_{h-1} \otimes (U_h \otimes \cdots \otimes U_k) \otimes U_{k+1} \otimes \cdots \otimes U_n \arrow[d, "f^{\dot{\otimes}}"] \\
& W
\end{tikzcd}$$

Above

$$\dot{\otimes} : U_1 \times \ldots \times U_n \to U_1 \otimes \cdots \otimes U_{h-1} \otimes (U_h \otimes \cdots \otimes U_k) \otimes U_{k+1} \otimes \cdots \otimes U_n \,.$$

is defined as the multi-linear function such that:

$$\dot{\otimes} : (u_1, \ldots, u_n) \mapsto u_1 \otimes \cdots \otimes u_{h-1} \otimes (u_h \otimes \cdots \otimes u_k) \otimes u_{k+1} \otimes \cdots \otimes u_n \,.$$

Now the linear map $F'_1 : (V_1 \otimes V_2) \otimes V_3 \to V_1 \otimes V_2 \otimes V_3$ such that linearly extends

$$(v_1 \otimes v_2) \otimes v_3 \to v_1 \otimes v_2 \otimes v_3$$

is nothing but the function $H_1^{\dot{\otimes}}$

$$H : (v_1, v_2, v_3) \to v_1 \otimes v_2 \otimes v_3$$

when $n = 3$, $h = 1$, $k = 2$ and $W := V_1 \otimes V_2 \otimes V_3$.

Exercises 2.27.

1. Consider a finite-dimensional vector space V and its dual V^*. Show by the universality theorem that there is a natural isomorphism such that

$$V \otimes V^* \simeq V^* \otimes V \,.$$

(*Hint*. Consider the bilinear map $f : V \times V^* \to V^* \otimes V$ with $f : (v_1, v_2^*) \mapsto v_2^* \otimes v_1$. Show that $f^{\otimes}$ is injective and thus surjective because $dim(V^* \otimes V) = dim(V \otimes V^*)$.)

2. Extend the result in exercise 1 to the infinite-dimensional case.

3. If $\psi_i : V_i \to U_i$ are $\mathbb{K}$-vector space isomorphisms for $i = 1, \ldots, n$, prove that

$$\Psi^{\otimes} : \bigotimes_{i=1}^{n} V_i \to \bigotimes_{i=1}^{n} U_i$$

is a $\mathbb{K}$-vector space isomorphism if

$$\Psi : V_1 \times \cdots \times V_n \ni (v_1, \ldots, v_n) \mapsto \psi_1(v_1) \otimes \cdots \otimes \psi_n(v_n) \in U_1 \otimes \cdots \otimes U_n \,.$$

(*Hint*. Prove that $(\Psi^{\otimes})^{-1} = \Phi^{\otimes}$ for

$$\Phi : U_1 \times \cdots \times U_n \ni (u_1, \ldots, u_n) \mapsto \psi_1^{-1}(u_1) \otimes \cdots \otimes \psi_n^{-1}(u_n) \in V_1 \otimes \cdots \otimes V_n \,,$$

using the universal property of the tensor product.) ■

2.2.3 Abstract definition of the tensor product of vector spaces

The property of the pair $(\otimes, U_1 \otimes \ldots \otimes U_n)$ stated in the theorem 2.25, the *universality property (of the tensor product)*, is very important from a theoretical point of view. As a matter of fact, it can be used, adopting a more advanced theoretical point of view, to *define* the tensor product of given vector spaces. This is due to the following important result.

Theorem 2.28. *Given $n \geq 1$ vector spaces $U_1, \ldots, U_n$ (not necessarily finite dimensional) on the same field $\mathbb{K}$ (not necessarily $\mathbb{R}$ or $\mathbb{C}$)[1] suppose there is a pair (T, U_T) where U_T is a vector space on $\mathbb{K}$ and $T : U_1 \times \ldots \times U_n \to U_T$ a multi linear map, fulfilling the universal property. That is, for any vector space W on $\mathbb{K}$ and any multi linear map $f : U_1 \times \ldots \times U_n \to W$, there is a unique linear map $f^T : U_T \to W$ such that the diagram below commute,*

$$\begin{tikzcd} U_1 \times \ldots \times U_n \arrow[r, "T"] \arrow[dr, "f"'] & U_T \arrow[d, "f^T"] \\ & W \end{tikzcd}$$

Under these hypotheses the pair (T, U_T) is determined up to vector space isomorphisms. In other words, for any other pair (S, V_S) fulfilling the universality property with respect to $U_1, \ldots, U_n$, there is a unique isomorphism $\phi : V_S \to U_T$ such that $f^T \circ \phi = f^S$ for every multi linear map $f : U_1 \times \ldots \times U_n \to W$ and every vector space W.

Proof. Suppose that there is another pair (S, V_S) fulfilling the universality property with respect to $U_1, \ldots, U_n$. Then, using the universal property with $f = T$ and $W = U_T$ we have the diagram below, with the former diagram commutation relation:

$$T^S \circ S = T \,. \tag{2.4}$$

$$\begin{tikzcd} U_1 \times \ldots \times U_n \arrow[r, "S"] \arrow[dr, "T"'] & V_S \arrow[d, "T^S"] \\ & U_T \end{tikzcd}$$

On the other hand, using the analogous property of the pair (T, U_T) with $f = S$ and $W = V_S$

[1] The theorem can be proved assuming that $U_1, \ldots, U_n, U_T$ are *modules* on a common *commutative ring* R. A module is defined as vector space with the exception that the field of scalars is here replaced by a *ring*, which may be also non-commutative.

we also have the commutative diagram

$$\begin{tikzcd}
U_1 \times \ldots \times U_n \arrow[r, "T"] \arrow[dr, "S"'] & U_T \arrow[d, "S^T"] \\
 & V_S
\end{tikzcd}$$

which involves a second diagram commutation relation:

$$S^T \circ T = S \,.$$

The two obtained relations imply

$$(T^S \circ S^T) \circ T = T$$

and

$$(S^T \circ T^S) \circ S = S \,,$$

Now observe that also the following corresponding identities are trivially valid

$$id_{U_T} \circ T = T$$

and

$$id_{V_S} \circ S = S \,.$$

the former can be seen as a special case of $f^T \circ T = f$ for $f = T$ and the latter a special case of $f^S \circ S = f$ for $f = S$. Since these idetities uniquely define the respective function f due to the universal propery, we conclude that $T^S \circ S^T = Id_{U_T}$ and $S^T \circ T^S = Id_{V_S}$. Summarizing, we have found that the linear map $\phi := T^S : V_S \to U_T$ is a vector-space isomorphism whose inverse is S^T. As far as the property $f^T \circ \phi = f^S$ is concerned, we observe that, by definition of f^S and f^T, the maps f^S and $f^T \circ \phi$ satisfy the universal property for f with respect to the pair (S, V_S):

$$f^S \circ S = f \quad \text{and} \quad (f^T \circ \phi) \circ S = f \,, \tag{2.5}$$

and thus f^S and $f^T \circ \phi$ have to coincide.
Indeed, the former in (2.5) is valid by definition. Furthermore, using (2.4), one achieves the latter:

$$(f^T \circ \phi) \circ S = f^T \circ T^S \circ S = f^T \circ (T^S \circ S) = f^T \circ T = f \,.$$

To conclude we prove that $\phi : V_S \to U_T$ verifying $f^T \circ \phi = f^S$ is unique. Indeed if $\phi' : V_S \to U_T$ is another isomorphism verifying $f^T \circ \phi' = f^S$, we have $f^T \circ (\phi' - \phi) = 0$ for every $f : U_1 \times \cdots \times U_n \to W$ multilinear. Choose $W = U_T$ and $f = T$ so that $f^T = id_{U_T}$ and $f^T \circ (\phi' - \phi) = 0$ implies $\phi' - \phi = 0$, that is $\phi' = \phi$. $\square$

Obviously, first of all this theorems applies on the pair $(\otimes, U_1 \otimes, \cdots \otimes U_n)$ on the field $\mathbb{R}$ or $\mathbb{C}$. In fact, we wish to stress that, taking advantage of such a result, given the generic vector spaces

$U_1, \dots, U_n$ on the generic field $\mathbb{K}$, one may says that a vector space U_T on $\mathbb{K}$, equipped with a multi linear map $T : U_1 \times \dots \times U_n \to U_T$, is the **tensor product** of the spaces and T is the **tensor product map**, if the pair (T, U_T) fulfills the universality property. In this way (T, U_T) turns out to be defined *up to vector-space isomorphisms*. Obviously, if one aims to follow this very general way to define tensor products, he/she still has to show that a pair (T, U_T) exists for the considered set of vector spaces $U_1, \dots, U_n$. This is exactly what we done within our constructive approach case when $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. However the existence of a pair (T, U_T) satisfying the universality property can be given similarly also for a larger class of vector spaces than that of finite-dimensional ones on the field $\mathbb{R}$ or $\mathbb{C}$ and also for *modules* over *commutative* rings. Moreover the explicit construction of a pair (T, U_T) fulfilling the universality property can be produced, within a more abstract approach, by taking the quotient of a suitable freely generated module with respect to suitable sub modules [Lang2].

2.2.4 Hilbert tensor product of Hilbert spaces

This subsection is devoted to define the Hilbert tensor product of Hilbert spaces which plays a key role in applications to Quantum Mechanics in reference to composite systems (see, e.g., [Moretti-a]). The issue is quite far from the main stream of these lecture notes and it could be omitted by the reader who is not strictly interested, also because it involves some advanced mathematical tools [Rudin] which will not be treated in detail within these notes (however see Chapter 5 for some elementary pieces of information about Hilbert spaces). We assume here that the reader is familiar with the basic notions of complex (generally nonseparable) Hilbert space and the notion of H bases (also called orthonormal systems).

Consider a set of complex Hilbert spaces $(H_i, (\cdot|\cdot)_i)$, where $i = 1, \dots, n < +\infty$ and the spaces are not necessarily separable. Here $(\cdot|\cdot)_i$ denotes the Hermitian scalar product on H_i. (In these lectures, concerning Hermitian scalar products, we always assume the convention that each $(\cdot|\cdot)_i$ is antilinear in the left entry.) Since the H_i are vector spaces on $\mathbb{C}$, their tensor product $H_1 \otimes \cdots \otimes H_n$ is defined as pure algebraic object, but it does not support any preferred Hermitian scalar product. We wish to specialize the definition of tensor product in order to endow $H_1 \otimes \cdots \otimes H_n$ with the structure of Hilbert space, induced by the structures of the $(H_i, (\cdot|\cdot)_i)$ naturally. Actually, this specialization requires an extension of $H_1 \otimes \cdots \otimes H_n$, to assure the completeness of the space. The key tool is the following theorem whose nature is very close to that of the universality theorem (Theorem 2.25). This result allow us to define a preferred hermitean product $(\cdot|\cdot)$ on $H_1 \otimes \cdots \otimes H_n$ induced by the $(\cdot|\cdot)_i$. The Hilbert structure generated by that scalar product on $H_1 \otimes \cdots \otimes H_n$ will be the wanted one.

Theorem 2.29. *Given $n \geq 1$ complex Hilbert spaces $(H_i, (\cdot|\cdot)_i)$, $i = 1, \dots, n < +\infty$ the following holds.*
(a) *For any complex vector space W and a map $f : (H_1 \times \cdots \times H_n) \times (H_1 \times \cdots \times H_n) \to W$ that is anti linear in each one of the first n entries and linear in each one of the remaining n ones,*

there is a unique map $f^{\boxtimes} : (H_1 \otimes \ldots \otimes H_n) \times (H_1 \otimes \ldots \otimes H_n) \to W$ which is anti linear in the left-hand argument and linear in the right-hand one and such that the diagram below commutes (in other words, $f^{\boxtimes} \circ \boxtimes = f$), where we have defined:

$$\boxtimes : (v_1, \ldots, v_n, u_1, \ldots, u_n) \mapsto (v_1 \otimes \cdots \otimes v_n, u_1 \otimes \cdots \otimes u_n) \quad \textit{for all } u_i, v_i \in H_i,\ i = 1, \ldots, n.$$

$$\begin{tikzcd}
(H_1 \times \cdots \times H_n) \times (H_1 \times \cdots \times H_n) \arrow[r, "\boxtimes"] \arrow[dr, "f"] & (H_1 \otimes \cdots \otimes H_n) \times (H_1 \otimes \cdots \otimes H_n) \arrow[d, "f^{\boxtimes}"] \\
& W
\end{tikzcd}$$

(b) *If $W := \mathbb{C}$ and it holds*

$$f((v_1, \ldots, v_n), (u_1, \ldots, u_n)) = \overline{f((u_1, \ldots, u_n), (v_1, \ldots, v_n))} \quad \textit{for all } u_i, v_i \in H_i,\ i = 1, \ldots, n,$$

then the map $f^{\boxtimes}$ as in (a) fulfills

$$f^{\boxtimes}(t, s) = \overline{f^{\boxtimes}(s, t)} \quad \text{for all } s, t \in H_1 \otimes \cdots \otimes H_n.$$

Proof. (a) Fix (algebraic) bases $\{e_{k,i_k}\}_{i_k \in I_k} \subset H_k$, $k = 1, \ldots, n$. By (anti)linearity, a map $g : (H_1 \otimes \ldots \otimes H_n) \times (H_1 \otimes \ldots \otimes H_n) \to W$, which is anti linear in the left-hand entry and linear in the right-hand one, is uniquely assigned by fixing the set of vectors

$$\{g(e_{1,i_1} \otimes \cdots \otimes e_{n,i_n}, e_{1,j_1} \otimes \cdots \otimes e_{n,j_n})\}_{i_1 \in I_1, \ldots, i_n \in I_n, j_1 \in I_1, \ldots, j_n \in I_n} \subset W\,.$$

Define the function $f^{\boxtimes} : (H_1 \otimes \ldots \otimes H_n) \times (H_1 \otimes \ldots \otimes H_n) \to W$ as the unique, anti linear in left-hand argument and linear in the reamining one, map with

$$f^{\boxtimes}(e_{i_1} \otimes \cdots \otimes e_{i_n}, e_{j_1} \otimes \cdots \otimes e_{j_n}) := f(e_{i_1}, \ldots, e_{i_n}, e_{j_1}, \ldots, e_{j_n})\,.$$

By construction, using (anti)linearity of $f^{\boxtimes}$ and multi-(anti)linearity of f, one gets immediately:

$$f^{\boxtimes}((v_1 \otimes \ldots \otimes v_n), (u_1 \otimes \ldots \otimes u_n)) = f(v_1, \ldots, v_n, u_1, \ldots, u_n)$$

for all $(v_1, \ldots, v_n), (u_1, \ldots, u_n) \in H_1 \times \ldots \times H_n$. In other words $f^{\boxtimes} \circ \boxtimes = f$.
The uniqueness of the map $f^{\boxtimes}$ is obvious: suppose there is another map $g^{\boxtimes}$ with $g^{\boxtimes} \circ \boxtimes = f$ then $(f^{\boxtimes} - g^{\boxtimes}) \circ \boxtimes = 0$. This means in particular that

$$(f^{\boxtimes} - g^{\boxtimes})(e_{i_1} \otimes \ldots \otimes e_{i_n}, e_{j_1} \otimes \ldots \otimes e_{j_n}) = 0\,,$$

for all $i_k \in I_k$, $j_k \in I_k$ $k = 1, \ldots, n$. Since the coefficients above completely determine a map, the considered map must be the null map and thus: $f^{\boxtimes} = g^{\boxtimes}$. The proof of (b) follows straightforwardly by the construction of $f^{\boxtimes}$. $\Box$

Equipped with the established result, consider the map $f : (H_1 \times \cdots \times H_n) \times (H_1 \times \cdots \times H_n) \to \mathbb{C}$ that is anti linear in the first n entries and linear in the remaining n entries and it is defined by the requirement:

$$f((v_1, \ldots, v_n), (u_1, \ldots, u_n)) := (v_1|u_1)_1 (v_2|u_2)_2 \cdots (v_n|u_n)_n \,, \quad \text{for all } v_i, u_i \in H_i, i = 1, \ldots, n.$$

As $(v_k|u_k)_k = \overline{(u_k|v_k)_k}$ by definition of Hermitian scalar product, the hypotheses in (b) of the theorem above is verified and thus the application

$$f^{\boxtimes} : (H_1 \otimes \cdots \otimes H_n) \times (H_1 \otimes \cdots \otimes H_n) \to \mathbb{C}$$

defined in (a) of the mentioned theorem, is linear in the right-hand argument, antilinear in the left-hand one and fulfils

$$f^{\boxtimes}(s|t) = \overline{f^{\boxtimes}(t|s)} \,, \quad \text{for all } s, t \in H_1 \otimes \cdots \otimes H_n.$$

If we were able to prove that $f^{\boxtimes}(\cdot|\cdot)$ is positively defined, i.e $f^{\boxtimes}(s|s) \geq 0$ for every $s \in H_1 \otimes \cdots \otimes H_n$ as well as $f^{\boxtimes}(s|s) = 0$ implies $s = 0$, then $f^{\boxtimes}(\cdot|\cdot)$ would define a Hermitian scalar product on $H_1 \otimes \cdots \otimes H_n$. This is just the last ingredient we need to give the wanted defintion of H tensor product. Let us prove it in form of a proposition.

Proposition 2.30. *Given $n \geq 1$ complex Hilbert spaces $(H_i, (\cdot|\cdot)_i)$, $i = 1, \ldots, n < +\infty$, the unique map $(\cdot|\cdot) : (H_1 \otimes \cdots \otimes H_n) \times (H_1 \otimes \cdots \otimes H_n) \to \mathbb{C}$ which is linear in the right-hand argument and anti linear in the left-hand one and verify*

$$(v_1 \otimes \cdots \otimes v_n | u_1 \otimes \cdots \otimes u_n) := (v_1|u_1)_1 (v_2|u_2)_2 \cdots (v_n|u_n)_n \,, \quad \textit{for all } v_i, u_i \in H_i, i = 1, \ldots, n,$$

is a Hermitian scalar product on $H_1 \otimes \cdots \otimes H_n$.

Proof. The only point we have to demonstrate is that $(\cdot|\cdot)$ is positively defined. To this end, fix a algebraic basis $\{e_{k,i_k}\}_{i_k \in I_k}$ for every H_k. If $S = S^{i_1 \cdots i_n} e_{1,i_1} \otimes \cdots \otimes e_{n,i_n} \in H_1 \otimes \cdots \otimes H_n$, the sum over repeated indices is always finite in view of the algebraic nature of the bases, also if each I_k may be infinite and uncountable. Similarly, fix a Hilbert base $\{h_{k,\alpha_k}\}_{\alpha_k \in A_k}$ for every H_k. Notice that also the set of indices A_k may be uncountable, however, as is well known, in every Hilbert decomposition $(u|v)_k = \sum_{\alpha_k \in A_k} (u|h_{k,\alpha_k})_k (h_{k,\alpha_k}|v)_k$ the set of nonvanishing terms $(u|h_{k,\alpha_k})_k$ and $(h_{k,\alpha_k}|v)_k$ is coutable and the sum reduces to a (absolutely convergent) standard series. In the following we do *not* omit the symbol of summation over the indices varying in every A_k, but we do follow that convention concerning algebraic indices $i_k, j_k \in I_k$. Notice also that the order of infinite summs cannot be interchanged in general. With our hypotheses we have for every $S \in H_1 \otimes \cdots \otimes H_n$:

$$(S|S) = \overline{S^{i_1 \cdots i_n}} S^{j_1 \cdots j_n} (e_{1,i_1}|e_{1,j_1})_1 \cdots (e_{n,i_n}|e_{n,j_n})_n$$

$$= \sum_{\alpha_1\in A_1}\cdots\sum_{\alpha_n\in A_n} \overline{S^{i_1\cdots i_n}} S^{j_1\cdots j_n}(e_{1,i_1}|h_{1,\alpha_1})_1(h_{1,\alpha_1}|e_{1,j_1})_1\cdots(e_{n,i_n}|h_{n,\alpha_n})_n(h_{n,\alpha_n}|e_{n,j_n})_n$$

$$= \sum_{\alpha_1\in A_1}\cdots\sum_{\alpha_n\in A_n} \overline{S^{i_1\cdots i_n}} S^{j_1\cdots j_n}\overline{(h_{1,\alpha_1}|e_{1,i_1})_1}(h_{1,\alpha_1}|e_{1,j_1})_1\cdots\overline{(h_{n,\alpha_n}|e_{n,i_n})_n}\cdots(h_{n,\alpha_n}|e_{n,j_n})_n$$

$$= \sum_{\alpha_1\in A_1}\cdots\sum_{\alpha_n\in A_n} \overline{S^{i_1\cdots i_n}(h_{1,\alpha_1}|e_{1,i_1})_1\cdots(h_{n,\alpha_n}|e_{n,i_n})_n}\; S^{j_1\cdots j_n}(h_{1,\alpha_1}|e_{1,j_1})_1\cdots(h_{n,\alpha_n}|e_{n,j_n})_n$$

$$= \sum_{\alpha_1\in A_1}\cdots\sum_{\alpha_n\in A_n} \left|S^{j_1\cdots j_n}(h_{1,\alpha_1}|e_{1,j_1})_1\cdots(h_{n,\alpha_n}|e_{n,j_n})_n\right|^2 \geq 0\,.$$

Semi positivity is thus established. We have to prove that $(S|S) = 0$ entails $S = 0$. Form the last line of the expansion of $(S|S)$ obtained above, we conclude that $(S|S) = 0$ implies $S^{i_j\cdots j_n}(h_{1,\alpha_1}|e_{1,j_1})_1(h_{n,\alpha_n}|e_{n,j_n})_n = 0$ for every given set of elements $h_{1,\alpha_1},\ldots,h_{n,\alpha_n}$. Now fix $\alpha_2,\ldots,\alpha_n$ to (arbitrarily) assigned values. By linearity and continuity of the scalar product, we have,

$$\sum_{\alpha_1\in A_1} S^{j_1 j_2\cdots j_n}c^{\alpha_1}(h_{1,\alpha_1}|e_{1,j_1})_1(h_{2,\alpha_1}|e_{2,j_2})_2\cdots(h_{n,\alpha_n}|e_{n,j_n})_n = 0\,,$$

provided $\sum_{\alpha_1\in A_1} c^{\alpha_1}h_{1,\alpha_1}$ converges to some element of H_1. Choosing $e_{1,i_1} = \sum_{\alpha_1\in A_1} c^{\alpha_1}h_{1,\alpha_1}$, for any fixed $i_1 \in I_1$, the identity above reduces to

$$(e_{1,i_1}|e_{1,i_1})_1 S^{i_1 j_2\cdots j_n}(h_{2,\alpha_1}|e_{2,j_2})_2\cdots(h_{n,\alpha_n}|e_{n,j_n})_n = 0 \quad \text{so that} \quad S^{i_1 j_2\cdots j_n}(h_{2,\alpha_1}|e_{2,j_2})_2\cdots(h_{n,\alpha_n}|e_{n,j_n})_n = 0.$$

Iterating the procedure we achieve the final identity:

$$S^{i_1 i_2\cdots i_n} = 0\,, \quad \text{for every } i_1 \in I_1,\ldots,i_n \in I_n,$$

that is $S = 0$. □

We have just established that the algebraic tensor product $H_1\otimes\cdots\otimes H_n$ can be equipped with a natural Hermitian scalar product $(\cdot|\cdot)$ naturally induced by the Hermitian scalar products in the factors H_i. At this point, it is worth reminding the reader that, given a complex vector space V with a Hermitian scalar product, $(\cdot,\cdot)$, there is a unique (up to Hilbert space isomorphisms) Hilbert space which admits V as a dense subspace and whose scalar product is obtained as a continuous extension of $(\cdot,\cdot)_V$. This distinguished Hilbert space is the **Hilbert completion** of $(V, (\cdot,\cdot)_V)$ [Rudin]. We have all the ingredients to state the definition of H tensor product.

Definition 2.31. (**Hilbert tensor product**.) Given $n \geq 1$ complex Hilbert spaces $(H_i, (\cdot|\cdot)_i)$, $i = 1,\ldots,n < +\infty$, the **Hilbert tensor product** $H_1\otimes_H\cdots\otimes_H H_n$ of those spaces, is the Hilbert completion of the space $H_1\otimes\cdots\otimes H_n$ with respect to the unique scalar product $(\cdot|\cdot) : H_1\otimes\cdots\otimes H_n \times H_1\otimes\cdots\otimes H_n \to \mathbb{C}$ verifying

$$(v_1\otimes\cdots\otimes v_n|u_1\otimes\cdots\otimes u_n) := (v_1|u_1)_1(v_2|u_2)_2\cdots(v_n|u_n)_n\,, \quad \text{for all } v_i, u_i \in H_i, i = 1,\ldots,n\,.$$

$\blacksquare$

To conclude, we prove the following useful result.

Proposition 2.32. *If $\{h_{k,\alpha_k}\}_{\alpha_k \in A_k}$ is a Hilbert basis of the complex Hilbert space H_k, for every $k = 1, \ldots, n$, then $\{h_{1,\alpha_1} \otimes \cdots \otimes h_{n,\alpha_n}\}_{\alpha_k \in A_k, k=1,\ldots,n}$ is a Hilbert basis of $H_1 \otimes_H \cdots \otimes_H H_n$. As a consequence $H_1 \otimes_H \cdots \otimes_H H_n$ is separable whenever all the factors H_k are separable.*

Proof. In the following $B_k := \{h_{k,\alpha_k}\}_{\alpha_k \in A_k}$, $< B_k >$ denotes the dense subspace of H_k generated by B_k, $B := \{h_{1,\alpha_1} \otimes \cdots \otimes h_{n,\alpha_n}\}_{\alpha_k \in A_k, k=1,\ldots,n}$, and $< B >$ is the subspace of $H_1 \otimes_H \cdots \otimes_H H_n$ generated by B. Finally, in the rest of the proof, the isometric linear map with dense range $j : H_1 \otimes \cdots \otimes H_n \to H_1 \otimes_H \cdots \otimes_H H_n$ defining the Hilbert completion of $H_1 \otimes \cdots \otimes H_n$ will be omitted in order to semplify the notation (one can always reduce to this case). Using the definition of the scalar product $(\cdot|\cdot)$ in $H_1 \otimes_H \cdots \otimes_H H_n$ given in definition 2.31 one sees that, trivially, B is an orthonormal set. Therefore, to conclude the proof, it is sufficient to prove that $\overline{< B >} = H_1 \otimes_H \cdots \otimes_H H_n$, where the bar indicates the topological closure. To this end consider also algebraic bases $D_k := \{e_{k,i_k}\}_{i_k \in I_k} \subset H_k$ made of *normalized vectors*. Since $\overline{< B_k >} = H_k$, given $m = 1, 2, \ldots$, there is $u_{k,i_k} \in < B_k >$ such that $||e_{k,i_k} - u_{k,i_k}||_k < 1/m$. Thus

$$||u_{k,i_k}||_k \leq ||e_{k,i_k}||_k + ||e_{k,i_k} - u_{k,i_k}||_k < 1 + 1/m\,.$$

Therfore, using $||v_1 \otimes \cdots \otimes v_n|| = ||v_1||_1 \cdots ||v_n||_n$ and the triangular inequality (see the remark after definition 5.3), we achieve: $||e_{1,i_1} \otimes \cdots \otimes e_{n,i_n} - u_{1,i_1} \otimes \cdots \otimes u_{n,i_n}||$

$$= ||(e_{1,i_1} - u_{1,i_1}) \otimes e_{2,i_2} \otimes \cdots \otimes e_{n,i_n} + u_{1,i_1} \otimes (e_{2,i_2} - u_{2,i_2}) \otimes \cdots \otimes e_{n,i_n} + \cdots + u_{1,i_1} \otimes u_{2,i_2} \otimes \cdots \otimes (e_{n,i_n} - u_{n,i_n})||$$

$$\leq ||(e_{1,i_1} - u_{1,i_1}) \otimes e_{2,i_2} \otimes \cdots \otimes e_{n,i_n}|| + ||u_{1,i_1} \otimes (e_{2,i_2} - u_{2,i_2}) \otimes \cdots \otimes e_{n,i_n}|| + \cdots$$

$$+ ||u_{1,i_1} \otimes u_{2,i_2} \otimes \cdots \otimes (e_{n,i_n} - u_{n,i_n})|| \leq n(1 + 1/m)^{n-1} \leq n\frac{2^{n-1}}{m}\,.$$

We conclude that, fixing $\epsilon > 0$, choosing m large enough, there is $y_{i_1 \ldots i_n} \in < B >$ such that

$$||e_{1,i_1} \otimes \cdots \otimes e_{n,i_n} - y_{i_1 \ldots i_n}|| < \epsilon\,. \tag{2.6}$$

The vectors $e_{1,i_1} \otimes \cdots \otimes e_{n,i_n}$ generates $H_1 \otimes \cdots \otimes H_n$ and linear combinations of elements $y_{i_1 \ldots i_n}$ as above belong to $< B >$. Hence, (2.6) together with homogeneity property of the norm as well as the triangular inequality (see the remarks after definition 5.3), imply that, for every $v \in H_1 \otimes \cdots \otimes H_n$ and every $\epsilon > 0$, there is $u \in < B >$ such that:

$$||u - v|| \leq \epsilon/2\,. \tag{2.7}$$

Now notice that, for every $v_0 \in H_1 \otimes_H \cdots \otimes_H H_n$ and any fixed $\epsilon > 0$, there is $v \in H_1 \otimes \cdots \otimes H_n$ with $||v_0 - v|| \leq \epsilon/2$, because $\overline{H_1 \otimes \cdots \otimes H_n} = H_1 \otimes_H \cdots \otimes_H H_n$. Therefore, fixing $u \in < B >$ as in (2.7), we have that, for every $v_0 \in H_1 \otimes_H \cdots \otimes_H H_n$ and any fixed $\epsilon > 0$, there is $u \in < B >$

with $||v_0 - u|| \leq \epsilon$. In other words $\overline{< B >} = H_1 \otimes_H \cdots \otimes_H H_n$ as wanted. $\square$

Remarks 2.33. Consider the case, quite usual in quantum mechanics, where the relevant Hilbert spaces are spaces of *square-integrable functions* [Rudin], $H_i := L^2(X_i, d\mu_i)$, where every X_i is a σ-finite space with measure μ_i, $i = 1, \ldots, n$. For instance $X_i = \mathbb{R}^3$ and $\mu_i = \mu$ is the standard Lebesgue measure (in this case every H_i is the state space of a spinless quantum particle). In the considered case, on a hand one can define the tensor product $H_1 \otimes_H \cdots \otimes_H H_n = L^2(X_1, d\mu_1) \otimes_H \cdots \otimes_H L^2(X_1, d\mu_1)$ on the other hand another natural object, relying on the given mathematical structures, is the Hilbert space $H := L^2(X_1 \times \cdots \times X_n, d\mu_1 \otimes \cdots \otimes d\mu_n)$, where the product measure $\mu_1 \otimes \cdots \otimes \mu_n$ is well defined since the spaces X_i are σ-finite. It turns out that these two spaces are naturally isomorphic and the isomorphism is the unique linear and continuous extension of the map:

$$L^2(X_1, d\mu_1) \otimes_H \cdots \otimes_H L^2(X_1, d\mu_1) \ni \psi_1 \otimes \cdots \otimes \psi_n \mapsto \psi_1 \cdots \psi_n \in L(X_1 \times \cdots \times X_n, d\mu_1 \otimes \cdots \otimes d\mu_n)$$

where

$$\psi_1 \cdots \psi_n(x_1, \ldots, x_n) := \psi_1(x_1) \cdots \psi_n(x_n) \quad \text{for all } x_i \in X_i,\ i = 1, \ldots, n..$$

Some textbook on quantum mechanics define the Hilbert space of a composite system of n spinless particles as $L(\mathbb{R}^3 \times \cdots \times \mathbb{R}^3, d\mu \otimes \cdots \otimes d\mu)$ rather than $L^2(\mathbb{R}^3, d\mu) \otimes_H \cdots \otimes_H L^2(\mathbb{R}^3, d\mu)$. Actually, the two definitions are mathematically equivalent. $\blacksquare$

Chapter 3

Tensor algebra, abstract index notation and some applications

3.1 Tensor algebra generated by a vector space

Let V be vector space on the field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Equipped with this algebraic structure, in principle, we may build up several different tensor spaces. Notice that we have to consider also $\mathbb{K}$, $\mathbb{K}^*$ and V^* as admissible tensor factors. (In the following we shall not interested in conjugate spaces.) Obviously, we are interested in tensor products which are not identifiable by some natural isomorphism.
First consider the dual $\mathbb{K}^*$ when we consider $\mathbb{K}$ as a vector space on the field $\mathbb{K}$ itself. $\mathbb{K}^*$ is made of linear functionals from $\mathbb{K}$ to $\mathbb{K}$. Each $c^* \in \mathbb{K}^*$ has the form $c^*(k) := c \cdot k$, for all $k \in \mathbb{K}$, where $c \in \mathbb{K}$ is a fixed field element which completely determines c^*. The map $c \mapsto c^*$ is a (natural) vector space isomorphism. Therefore

$$\mathbb{K} \simeq \mathbb{K}^* .$$

The natural isomorfism $\mathbb{K} \simeq \mathbb{K}^*$, taking advantage of the result in (3) of exercises 2.27, immediately induces an analogous isomorphism $\mathbb{K}\otimes\ldots\otimes\mathbb{K} \simeq \mathbb{K}^*\otimes\ldots\otimes\mathbb{K}^*$. We pass therefore to consider products $\mathbb{K}\otimes\ldots\otimes\mathbb{K} \simeq \mathbb{K}^*\otimes\ldots\otimes\mathbb{K}^* = \mathscr{L}(\mathbb{K},\ldots,\mathbb{K})$. Each multi-linear map $f \in \mathscr{L}(\mathbb{K},\ldots,\mathbb{K})$ is completely determined by the number $f(1,\ldots,1)$, since $f(k_1,\ldots,k_n) = k_1\cdots k_n f(1,\ldots,1)$. One can trivially show that the map $f \mapsto f(1,\ldots,1)$ is a (natural) vector space isomorphism between $\mathbb{K}\otimes\ldots\otimes\mathbb{K}$ and $\mathbb{K}$ itself. Therefore

$$\mathbb{K}\otimes\ldots\otimes\mathbb{K} \simeq \mathbb{K} .$$

Also notice that the found isomorphism trivially satisfies $c_1\otimes\ldots\otimes c_n \mapsto c_1\cdots c_n$, and thus the tensor product map reduces to the ordinary product of the field.
We pass now to consider the product $\mathbb{K}\otimes V \simeq \mathbb{K}^*\otimes V = \mathscr{L}(\mathbb{K}, V^*)$. Each multi-linear functional f in $\mathscr{L}(\mathbb{K}, V^*)$ is completely determined by the element of V, $f(1,\cdot) : V^* \to \mathbb{K}$, which maps each $v^* \in V^*$ in $f(1, v^*)$. Once again, it is a trivial task to show that $f \mapsto f(1,\cdot)$ is a (natural)

vector space isomorphism between $\mathbb{K} \otimes V$ and V itself.

$$\mathbb{K} \otimes V \simeq V \, .$$

Notice that the found isomorphism satisfies $k \otimes v \mapsto kv$ and thus the tensor product map reduces to the ordinary product of a field element and a vector.
Reminding that $\otimes$ is associative and * is involutive and, in the finite-dimensional case, it is also distributive, as established in chapter 2, we conclude that only the spaces $\mathbb{K}$, V, V^* and all the tensor products of V and V^* (in whichever order and number) may be significantly different. This result leads us to the following definition which we extend to the infinite dimensional case (it is however worth stressing that, in the infinite-dimensional case, some of the natural isomorphisms encountered in the finite-dimensional case are not valid, for instance $(V \otimes V)^*$ may be larger than $V^* \otimes V^*$).

Definition 3.1. **(Tensor Algebra generated by V.)** Let V be a vector space with field $\mathbb{K}$.

(1) The **tensor algebra** $\mathscr{A}_{\mathbb{K}}(V)$ generated by V with field $\mathbb{K}$ is the disjoint union of the vector spaces: $\mathbb{K}$, V, V^* and all of tensor products of factors V and V^* in whichever order and number.

(2)
- (i) The tensors of $\mathbb{K}$ are called **scalars**,
- (i) the tensors of V are called **contravariant vectors**,
- (iii) the tensors of V^* are called **covariant vectors**,
- (iv) the tensors of spaces $V^{n\otimes} := \underbrace{V \otimes \ldots \otimes V}_{n\ times}$, times, are called **contravariant tensors of order** n or **tensors of order** $(n, 0)$,
- (v) the tensors of spaces $V^{*n\otimes} := \underbrace{V^* \otimes \ldots \otimes V^*}_{n\ times}$ are called **covariant tensors of order** n or **tensors of order** $(0, n)$,
- (iv) the remaining tensors which belong to spaces containing n consecutive factors V and m subsequent consecutive factors V^* are called **tensors of order** (n, m). (The order of the remaining type of tensors is defined analogously taking the order of occurrence of the factors V and V^* into account.)

■

Remarks 3.2. Obviously, pairs of tensors spaces made of the same number of factors V and V^* in different order, are naturally isomorphic (see exercises 2.27.1). However, for practical reasons it is convenient to consider these spaces as different spaces and use the identifications when and if necessary. ■

Definition 3.3. **(Associative algebra.)** An *associative algebra* over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ is a vector space $\mathscr{A}$ over $\mathbb{K}$ equipped with a map $\circ : \mathscr{A} \times \mathscr{A} \to \mathscr{A}$ which is an *associative product*, i.e.,

$$A \circ (B \circ C) = (A \circ B) \circ C$$

if $A, B, C \in \mathscr{A}$, such that it is *two-side distributive* with respect to the vector space sum:

$$A \circ (B + C) = A \circ B + A \circ C\,, \quad (B + C) \circ A = B \circ A + C \circ A$$

if $A, B, C \in \mathscr{A}$ and finally satisfying

$$a(A \circ B) = (aA) \circ B = A \circ (aB)$$

for all $A, B \in \mathscr{A}$ and $a \in \mathbb{K}$.
The algebra $\mathscr{A}$ is said to be **commutative** or **Abelian** if $A \circ B = B \circ A$ for every $A, B \in \mathscr{A}$.
The algebra $\mathscr{A}$ is said to be **unital** if there is $\mathbb{I} \in \mathscr{A}$, called the **unit element**, such that $A \circ \mathbb{I} = \mathbb{I} \circ A$ for all $A \in \mathscr{A}$. ■

A simple example of (non-commutative for $n > 1$) unital algebra over $\mathbb{K}$ is the set of the $n \times n$ matrices with coefficients in $\mathbb{K}$ and equipped with the standard structure of vector space over $\mathbb{K}$, the associative product being the standard product of matrices and the unit element being the identity matrix I.

According to our previous definition $\mathscr{A}_{\mathbb{K}}(V)$ is *not* an *associative algebra*. Indeed, even if it has a notion of product, it does not enjoy a structure of linear space over $\mathbb{K}$, since linear combinations of tensors of different type are not defined. For this reason $\mathscr{A}_{\mathbb{K}}(V)$ is sometimes called "weak tensor algebra".
It is however possible to make stronger this structure obtaining a proper associative algebra in the sense of the stated definition, if suitably interpreting the tensor product. In this way, the various spaces of tensors turn out to be proper subspaces of the constructed structure. However linear combinations of tensors of different type have no natural interpretation especially in physical applications.
As a matter of fact, let us denote by S_α the generic space of tensors in $\mathscr{A}_{\mathbb{K}}(V)$. More precisely, if $\mathscr{A}_{\mathbb{K}}(V) = \bigsqcup_{\alpha\in A} S_\alpha$ where A is a set of indices with $S_\alpha \neq S_\beta$ if $\alpha \neq \beta$. With this notation, we can first define a natural $\mathbb{K}$-vector space structure over $\bigtimes_{\alpha\in A} S_\alpha$ with the notion of linear combination given by, for $a, b \in \mathbb{K}$,

$$a \oplus_{\alpha\in A} v_\alpha + b \oplus_{\beta\in A} u_\beta := \oplus_{\alpha\in A} (a v_\alpha + b u_\beta)$$

where $\oplus_{\alpha\in A} v_\alpha := \{v_\alpha\}_{\alpha\in A}$ and $\oplus_{\beta\in A} u_\beta := \{u_\beta\}_{\beta\in A}$ are generic elements of

$$\bigoplus_{\alpha\in A} S_\alpha := \bigtimes_{\alpha\in A} S_\alpha\,,$$

the notation stressing that the Cartesian product it is now equipped with a vector space structure. Finally the associative algebra structure arises by defining the associative product over $\bigoplus_{\alpha\in A} S_\alpha$:

$$\oplus_{\alpha\in A} v_\alpha \bigotimes \oplus_{\beta\in A} u_\beta := \oplus_{\gamma\in A}\left(\sum_{v_\alpha\otimes u_\beta\in S_\gamma} v_\alpha\otimes u_\beta\right)$$

where we have identified elements $v_\alpha \otimes u_\beta$ with corresponding elements of tensor spaces S_γ in $\mathscr{A}_{\mathbb{K}}(V)$ making use of the vector space isomorphisms in theorem 2.23. Notice that products $v_\alpha \otimes u_\beta$ and $v_{\alpha'} \otimes u_{\beta'}$ with $\alpha \neq \alpha'$ and $\beta \neq \beta'$ may belong to the same tensor space S_γ in $\mathscr{A}_{\mathbb{K}}(V)$.

3.2 The abstract index notation and rules to handle tensors

3.2.1 Canonical bases and abstract index notation

Let us introduce the **abstract index notation**. Consider a finite-dimensional vector space V with field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. After specification of a basis $\{e_i\}_{i\in I} \subset V$ and the corresponding basis in V^*, each tensor is completely determined by giving its components with respect to the induced basis in the corresponding tensor space in $\mathscr{A}_{\mathbb{K}}(V)$. We are interested in the transformation rule of these components under change of the base in V. Suppose to fix another basis $\{e'_j\}_{j\in I} \subset V$ with $e_i = A^j{}_i e'_j$. The coefficients $A^j{}_i$ determine a matrix $A := [A^j{}_i]$ in the matrix group $GL(dimV, \mathbb{K})$, i.e., the group (see section 4.1) of $dimV \times dimV$ matrices with coefficients in $\mathbb{K}$ and non-vanishing determinant. First consider a contravariant vector $t = t^i e_i$, passing to the other basis, we have $t = t^i e_i = t'^j e'_j$ and thus $t'^j e'_j = t^i A^j{}_i e'_j$. This is equivalent to $(t'^j - A^j{}_i t^i)e'_j = 0$ which implies

$$t'^j = A^j{}_i t^i \ ,$$

because of the linear independence of vectors e'_j. Similarly, if we specify a set of components in $\mathbb{K}$, $\{t^i\}_{i\in I}$ for each basis $\{e_i\}_{i\in I} \subset V$ and these components, changing the basis to $\{e'_j\}_{j\in I}$, transform as

$$t'^j = A^j{}_i t^i \ ,$$

where the coefficients $A^j{}_i$ are defined by

$$e_i = A^j{}_i e'_j \ ,$$

then a contravariant tensor t is defined. It is determined by $t := t^i e_i$ in each basis. The proof is self evident.

Concerning covariant vectors, a similar result holds. Indeed a covariant vector $u \in V^*$ is completely determined by the specification of a set of components $\{u_i\}_{i\in I}$ for each basis $\{e^{*i}\}_{i\in I} \subset V^*$ (dual basis of $\{e_i\}_{i\in I}$ above) when these components, changing the basis to $\{e'^{*j}\}_{j\in I}$ (dual base of $\{e'_j\}_{j\in I}$ above), transform as

$$u'_j = B_j{}^i u_i \ ,$$

where the coefficients ${B_l}^r$ are defined by

$$e^{*i} = {B_j}^i e'^{*j} .$$

What is the relation between the matrix $A = [{A^j}_i]$ and the matrix $B := [{B_k}^h]$? The answer is obvious: it must be

$$\delta^i_j = \langle e_j, e^{*i} \rangle = {A^l}_j {B_k}^i \langle e'_l, e'^{*k} \rangle = {A^l}_j {B_k}^i \delta^k_l = {A^k}_j {B_k}^i .$$

In other words it has to hold $I = A^t B$, which is equivalent to

$$B = A^{-1t} .$$

(Notice that ^{t} and $^{-1}$ commute.)
Proper tensors have components which transform similarly to the vector case. For instance, consider $t \in V \otimes V^*$, fix a basis $\{e_i\}_{i \in I} \subset V$, the dual one $\{e^{*i}\}_{i \in I} \subset V^*$ and consider that induced in $V \otimes V^*$, $\{e_j \otimes e^{*j}\}_{(i,j) \in I \times I}$. Then $t = {t^i}_j e_i \otimes e^{*j}$. By bi linearity of the tensor product map, if we pass to consider another basis $\{e'_i\}_{i \in I} \subset V$ and those associated in the relevant spaces as above, concerning the components ${t'^k}_l$ of t in the new tensor space basis, one trivially gets

$${t'^k}_l = {A^k}_i {B_l}^j {t^i}_j ,$$

where the matrices $A = [{A^j}_i]$ and $B := [{B_k}^h]$ are those considered above. It is obvious that the specification of a tensor of $V \otimes V^*$ is completely equivalent to the specification of a set of components for each basis of $V \otimes V^*$, $\{e_j \otimes e^{*j}\}_{(i,j) \in I \times I}$, provided these components transform as specified above under change of basis.
We can generalize the obtained results after a definition.

Definition 3.4. **(Canonical bases.)** Let $\mathscr{A}_{\mathbb{K}}(V)$ be the tensor algebra generated by the finite-dimensional vector space V on the field $\mathbb{K}$ ($= \mathbb{C}$ or $\mathbb{R}$). If $B = \{e_i\}_{i \in I}$ is a basis in V with dual basis $B^* = \{e^{*i}\}_{i \in I} \subset V^*$, the **canonical bases** associated to the former are the bases in the tensor spaces of $\mathscr{A}_{\mathbb{K}}(V)$ obtained by tensor products of elements of B and B^*. ■

Remarks 3.5. Notice that also $\{e_i \otimes e'_j\}_{i,j \in I}$ is a basis of $V \otimes V$ if $\{e_i\}_{i \in I}$ and $\{e'_j\}_{j \in I}$ are bases of V. However, $\{e_i \otimes e'_j\}_{i,j \in I}$ is *not* canonical unless $e_i = e'_i$ for all $i \in I$. ■

Theorem 3.6. *Consider the tensor algebra $\mathscr{A}_{\mathbb{K}}(V)$ generated by a finite-dimension vector space V with field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ and take a tensor space $V^{n\otimes} \otimes V^{*m\otimes} \in \mathscr{A}_{\mathbb{K}}(V)$. Defining of a tensor $t \in V^{n\otimes} \otimes V^{*m\otimes}$ is completely equivalent to assigning a set of components*

$$\{{t^{i_1 \ldots i_n}}_{j_1 \ldots j_m}\}_{i_1, \ldots, i_n, j_1, \ldots, j_m \in I}$$

with respect to each canonical basis of $V^{n\otimes} \otimes V^{*m\otimes}$,

$$\{e_{i_1} \otimes \ldots \otimes e_{i_n} \otimes e^{*j_1} \otimes \ldots \otimes e^{*j_n}\}_{i_1, \ldots i_n, j_1, \ldots j_m \in I}$$

which, under change of basis:

$$\{e'_{i_1} \otimes \ldots \otimes e'_{i_n} \otimes e'^{*j_1} \otimes \ldots \otimes e'^{*j_n}\}_{i_1,\ldots i_n, j_1,\ldots j_m \in I}$$

transform as:

$$t'^{i_1\ldots i_n}{}_{j_1\ldots j_m} = A^{i_1}{}_{k_1} \ldots A^{i_n}{}_{k_n} B_{j_1}{}^{l_1} \ldots B_{j_m}{}^{l_m} t^{k_1\ldots k_n}{}_{l_1\ldots l_m} \, ,$$

where

$$e_i = A^j{}_i e'_j \, ,$$

and the coefficients $B_j{}^l$ are those of the matrix:

$$B = A^{-1t} \, ,$$

with $A := [A^j{}_i]$. The associated tensor t is represented by

$$t = t^{i_1\ldots i_n}{}_{j_1\ldots j_m} e_{i_1} \otimes \ldots \otimes e_{i_n} \otimes e^{*j_1} \otimes \ldots \otimes e^{*j_m}$$

for each considered canonical basis. Analogous results hold for tensor spaces whose factors V and V^* take different positions.

Notation 3.7. In the **abstract index notation** a tensor is indicated by writing its generic component in a non-specified basis. E.g. $t \in V^* \otimes V$ is indicated by $t_i{}^j$. ■

We shall occasionally adopt a *cumulative index notation*, i.e., letters $A, B, C, \ldots$ denote set of covariant, contravariant or mixed indices. For instance $t^{ijk}{}_{lm}$ can be written as t^A with ${}^A = {}^{ijk}{}_{lm}$. Similarly e_A denotes the element of a canonical basis $e_i \otimes e_j \otimes e_k \otimes e^{*l} \otimes e^{*m}$. Moreover, if ^{A} and ^{B} are cumulative indices, the indices of the cumulative index ^{AB} are those of ^{A} immediately followed by those of ^{B}, e.g, if ${}^A = {}^{ijk}{}_{lm}$, and ${}^B = {}^{pq}{}_u{}^n$, ${}^{AB} = {}^{ijk}{}_{lm}{}^{pq}{}_u{}^n$.

3.2.2 Rules to compose, decompose, produce tensors form tensors

Let us specify the allowed mathematical rules to produces tensors from given tensors. To this end we shall use both the synthetic and the index notation.

Linear combinations of tensors of a fixed tensor space. Take a tensor space S in $\mathscr{A}_{\mathbb{K}}(V)$. This is a vector space by definition, and thus picking out $s, t \in S$, and $\alpha, \beta \in \mathbb{K}$, linear combinations can be formed which still belong to S. In other words we may define the tensor of S

$$u := \alpha s + \beta t \quad \text{or, in the abstract index notation} \quad u^A = \alpha s^A + \beta t^A \, .$$

The definition of u above given by the abstract index notation means that the components of u are related to the components of s and t by a linear combination which has *the same form in every canonical basis of the space S* and the coefficients α,β *do not depend on the basis.*

Products of tensors of generally different tensor spaces. Take two tensor spaces S, S' in $\mathscr{A}_{\mathbb{K}}(V)$ and pick out $t \in S$, $t' \in S'$. Then the tensor $t \otimes t' \in S \otimes S'$ is well defined. Using the associativity of the tensor product by natural isomorphisms, we find a unique tensor space S'' in $\mathscr{A}_{\mathbb{K}}(V)$ which is isomorphic to $S \otimes S'$ and thus we can identify $t \otimes t'$ with a tensor in S'' which we shall indicate by $t \otimes t'$ once again with a little misuse of notation. $t \otimes t'$ is called the **product** of tensors t and t'. Therefore, the product of tensors coincides with the usual tensor product of tensors up to a natural isomorphism. What about the abstract index notation? By theorem 2.23 one sees that, fixing a basis in V, the natural isomorphism which identify $S \otimes S$ with S'' transforms the products of elements of canonical bases in S and S' in the corresponding elements of the canonical basis of S'' obtained by cancelling every parantesis; e.g. for $S = V \otimes V^*$ and $S' = V^*$, the natural isomorphism from $(V \otimes V^*) \otimes V^*$ to $V \otimes V^* \otimes V^*$ transforms each $(e_k \otimes e^{*h}) \otimes e^{*r}$ to $e_k \otimes e^{*h} \otimes e^{*r}$. As a consequence

$$(t \otimes t')^{AB} = t^A t'^B \,.$$

Therefore, for instance, if $S = V \otimes V^*$ and $S' = V^*$ the tensors t and t' are respectively indicated by $t^i{}_j$ and s_k and thus $(t \otimes s)^i{}_{jk} = t^i{}_j s_k$.

Contractions. Consider a tensor space of $\mathscr{A}_{\mathbb{K}}(V)$ of the form

$$U_1 \otimes \ldots \otimes U_k \otimes V \otimes U_{k+1} \otimes \ldots \otimes U_l \otimes V^* \otimes U_{l+1} \otimes \ldots \otimes U_n$$

where U_i denotes either V or V^*. Everything we are going to say can be re-stated for the analogous space

$$U_1 \otimes \ldots \otimes U_k \otimes V^* \otimes U_{k+1} \otimes \ldots \otimes U_l \otimes V \otimes U_{l+1} \otimes \ldots \otimes U_n \,.$$

Then consider the *multi-linear* map C with domain

$$U_1 \times \ldots \times U_k \times V \times U_{k+1} \times \ldots \times U_l \times V^* \times U_{l+1} \times \ldots \times U_n \,,$$

and values in

$$U_1 \otimes \ldots \otimes U_k \otimes U_{k+1} \otimes \ldots \otimes U_l \otimes U_{l+1} \otimes \ldots \otimes U_n$$

defined by:

$$(u_1, \ldots, u_k, v, u_{k+1}, \ldots, u_l, v^*, u_{l+1}, \ldots, u_n) \mapsto \langle v, v^* \rangle u_1 \otimes \ldots \otimes u_k \otimes u_{k+1} \otimes \ldots \otimes u_l \otimes u_{l+1} \otimes \ldots \otimes u_n \,.$$

By the universality theorem there is a *linear* map $C^{\otimes}$, called **contraction** of V and V^*, defined on the whole tensor space

$$U_1 \otimes \ldots \otimes U_k \otimes V \otimes U_{k+1} \otimes \ldots \otimes U_l \otimes V^* \otimes U_{l+1} \otimes \ldots \otimes U_n$$

taking values in

$$U_1 \otimes \ldots \otimes U_k \otimes U_{k+1} \otimes \ldots \otimes U_l \otimes U_{l+1} \otimes \ldots \otimes U_n$$

such that, on simple products of vectors reduces to

$$u_1\otimes\ldots\otimes u_k\otimes v\otimes u_{k+1}\otimes\ldots\otimes u_l\otimes v^*\otimes u_{l+1}\otimes\ldots\otimes u_n \mapsto \langle v, v^*\rangle u_1\otimes\ldots\otimes u_k\otimes u_{k+1}\otimes\ldots\otimes u_l\otimes u_{l+1}\otimes\ldots\otimes u_n.$$

This linear map takes tensors in a tensor product space with $n+2$ factors and produces tensors in a space with n factors. The simplest case arises for $n = 0$. In that case $C : V \times V^* \to \mathbb{K}$ is nothing but the bilinear pairing $C : (v, v^*) \mapsto \langle v, v^*\rangle$ and $C^\otimes$ is the linear associated map by the universality theorem.
Finally, let us represent the contraction map within the abstract index picture. It is quite simple to show that, $C^\otimes$ takes a tensor $t^{AiB}{}_j{}^C$ where ^{A}, ^{B} and ^{C} are arbitrary cumulative indices, and produces the tensor $(C^\otimes t)^{ABC} := t^{AkB}{}_k{}^C$ where we remark the convention of *summation of the twice repeated index* k. To show that the abstract-index representation of contractions is that above notice that the contractions are linear and thus

$$C^\otimes(t^{AiB}{}_j{}^C e_A\otimes e_i\otimes e_B\otimes e^{*j}\otimes e_C) = t^{AiB}{}_j{}^C C^\otimes(e_A\otimes e_i\otimes e_B\otimes e^{*j}\otimes e_C) = t^{AiB}{}_j{}^C \delta^j_i e_A\otimes e_B\otimes e_C\,,$$

and thus

$$C^\otimes(t^{AiB}{}_j{}^C e_A\otimes e_i\otimes e_B\otimes e^{*j}\otimes e_C) = t^{AkB}{}_k{}^C e_A\otimes e_B\otimes e_C\,.$$

This is nothing but:

$$(C^\otimes t)^{ABC} := t^{AkB}{}_k{}^C\,.$$

3.2.3 Linear transformations of tensors are tensors too

To conclude we pass to consider a final theorem which shows that there is a one-to-one correspondence between linear maps on tensors and tensors them-selves.

Theorem 3.8. **(Linear maps and tensors.)** *Let S, S' be a pair of tensor spaces of a tensor algebra $\mathscr{A}_{\mathbb{K}}(V)$ where V is finite dimensional. The vector space of linear maps from S to S' is naturally isomorphic to $S^*\otimes S'$ (which it is naturally isomorphic to the corresponding tensor space of $\mathscr{A}_{\mathbb{K}}(V)$). The isomorphism $F : S^*\otimes S' \to \mathscr{L}(S|S')$ such that $F : t \mapsto f_t$, where the linear function $f_t : S \to S'$ is defined by:*

$$f_t(s) := C^\otimes(s\otimes t)\,, \quad \textit{for all } s \in S,$$

where $C^\otimes$ is the contraction of all the indices of s and the corresponding indices in S^* of t. Moreover, fixing a basis $\{e_i\}_{i\in I}$ in V, let $\{e_A\}$ denote the canonical basis induced in S, $\{e'_B\}$ that induced in S' and $\{e'^{*C}\}$ that induced in S'^*. With those definitions if $s = s^A e_A \in S$,

$$f_t(s)^C = s^A t_A{}^C$$

and

$$t_A{}^C = \langle f_t(e_A), e'^{*C}\rangle\,.$$

The isomorphism F^{-1} is the composition of the isomorphisms in theorems 2.17 and 2.8: $\mathscr{L}(S|S') \simeq \mathscr{L}(S, S'^*) = S^* \otimes (S'^*)^* \simeq S^* \otimes S'$.

Proof. Defining f_t as said above it is evident that the map F is linear. The given definition implies that, in components $f_t(s)^C = s^A t_A{}^C$. Since $S^* \otimes S'$ and $\mathscr{L}(S|S') \simeq \mathscr{L}(S, S'^*) \simeq S^* \otimes S'$ have the same dimension, injectivity of F implies surjectivity. Injectivity has a straightforward proof: $f_t = 0$ is equivalent to, fixing canonical bases as indicated in the hypoteses, $s^A t_A{}^C = 0$ for every choice of coefficients s^A. This is equivalent, in turn, to $t_A{}^C = 0$ for every component of t and thus $t = 0$. The formula $t_A{}^C = \langle f_t(e_A), e'^{*C}\rangle$ can be proved as follows: if $s = e_A$, the identity $f_t(s)^C = s^A t_A{}^C$ reduces to $f_t(e_A)^C = t_A{}^C$, therefore $f_t(e_A) = f_t(e_A)^C e'_C = t_A{}^C e'_C$ and so $\langle f_t(e_A), e'^{*C}\rangle = t_A{}^C$. The last statement in the thesis can be obtained by direct inspection. □

Let us illustrate how one may use that theorem. For instance consider a linear map $f : V \to V$ where V is finite-dimensional with field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$. Then f defines a tensor $t \in V^* \otimes V$. In fact, fixing a basis $\{e_i\}_{i\in I}$in V and considering those canonically associated, by the linearity of f and the pairing:

$$f(v) = \langle f(v), e^{*k}\rangle e_k = v^i \langle f(e_i), e^{*k}\rangle e_k \,.$$

The tensor $t \in V^* \otimes V$ associated with f in view of the proved theorem has components

$$t_i{}^k := \langle f(e_i), e^{*k}\rangle \,.$$

Moreover, using the abstract index notation, one has

$$(f(v))^k = v^i t_i{}^k \,.$$

In other words the action of f on v reduces to (1) a product of the involved tensors:

$$v^i t_j{}^k \,,$$

(2) followed by a convenient contraction:

$$(f(v))^k = v^i t_i{}^k \,.$$

More complicate cases can be treated similarly. For example, linear maps f from $V^* \otimes V$ to $V^* \otimes V \otimes V$ are one-to-one with tensors $t^i{}_{jk}{}^{lm}$ of $V \otimes V^* \otimes V^* \otimes V \otimes V$ and their action on tensors $u_p{}^q$ of $V^* \otimes V$ is

$$f(u)_k{}^{lm} = u_i{}^j t^i{}_{jk}{}^{lm}$$

i.e., a product of tensors and two contractions. Obviously

$$t^i{}_{jk}{}^{lm} = \langle f(e^{*i} \otimes e_j), e_k \otimes e^{*l} \otimes e^{*m}\rangle \,.$$

Remarks 3.9. In the applications it is often convenient defining the tensor t in $S' \otimes S^*$ instead of $S^* \otimes S'$. In this case the isomorfism between $S' \otimes S^*$ and $\mathscr{L}(S|S')$ associates $t \in S' \otimes S^*$ with $g_t \in \mathscr{L}(S|S')$ where

$$g_t(s) := C^{\otimes}(t \otimes s)\,, \quad \text{for all } s \in S,$$

where $C^{\otimes}$ is the contraction of all the indices of s and the corresponding indices in S^* of t. ■

3.3 Physical invariance of the form of laws and tensors

A physically relevant result is that the rules given above to produce a new tensor from given tensors have the same form independently from the basis one use to handle tensors. For that reason the abstract index notation makes sense. In physics the choice of a basis is associated with the choice of a reference frame. As is well known, various relativity principles (*Galileian Principle*, *Special Relativity Principle* and *General Relativity Principle*) assume that
"*the law of Physics can be written in the same form in every reference frame*".
The allowed reference frames range in a class singled out by the considered relativity principle, for instance in Special Relativity the relevant class is that of inertial reference frames.
It is obvious that the use of tensors and rules to compose and decompose tensors to represent physical laws is very helpful in implementing relativity principles. In fact the theories of Relativity can be profitably formulated in terms of tensors and operations of tensors just to assure the invariance of the physical laws under change of the reference frame. When physical laws are given in terms of tensorial relations one says that those laws are *covariant*. It is worthwhile stressing that covariance is *not* the only way to state physical laws which preserve their form under changes of reference frames. For instance the Hamiltonian formulation of mechanics, in relativistic contexts, is invariant under change of reference frame but it is not formulated in terms of tensor relations (in spacetime).

3.4 Tensors on Affine and Euclidean spaces

We remind the reader the definitions of *affine space* and **Euclidean space**. We assume that the reader is familiar with these notions and in particular with the notion of scalar product and orthonormal basis. [Sernesi]. A general theory of scalar product from the point of view of tensor theory will be developed in chapter 5.

Definition 3.10. **(Affine space).** A **real** n**-dimensional affine space** [Sernesi] is a triple $(\mathbb{A}^n, V, \vec{\cdot\cdot})$ where $\mathbb{A}^n$ is a set whose elements are called **points**, V is a real n-dimensional vector space called **space of translations** and $\vec{\cdot\cdot} : \mathbb{A}^n \times \mathbb{A}^n \to V$ is a map such that the two following requirements are fulfilled.
(i) For each pair $p \in \mathbb{A}^n$, $v \in V$ there is a *unique* point $q \in \mathbb{A}^n$ such that $\overrightarrow{pq} = v$.
(ii) $\overrightarrow{pq} + \overrightarrow{qr} = \overrightarrow{pr}$ for all $p, q, r \in \mathbb{A}^n$.
$\overrightarrow{pq}$ is called vector with **initial point** p and **final point** q. ■

Definition 3.11. **(Euclidean space).** A (real) **Euclidean** space is a pair $(\mathbb{E}^n, (\cdot|\cdot))$, where $\mathbb{E}^n$ is an affine space with space of translations V and $(\cdot|\cdot) : V \times V \to \mathbb{R}$ is a scalar product on V. ■

A **Cartesian coordinate systems** on the affine space $\mathbb{A}^n$ is a bijective map $f : \mathbb{A}^n \to \mathbb{R}^n$

defined as follows. Fix a point $O \in \mathbb{A}^n$ called **origin** of the coordinates and fix a vector basis $\{e_i\}_{i=1,\ldots,n} \subset V$ whose elements are called **axes** of the coordinates. Varying $p \in \mathbb{A}^n$, the components $(x^1(p), \ldots, x^n(p)) \in \mathbb{R}^n$ of any vector $\overrightarrow{OP}$ with respect the chosen basis, define a map $f : \mathbb{A}^n \to \mathbb{R}^n$. Property (i) above implies that f is bijective.
It is simply proved that, if f' is another Cartesian coordinate system with origin O' and axes $\{e'_j\}_{j=1,\ldots,n}$, the following coordinate transformation law holds:

$$x'^j = \sum_{i=1}^{n} A^j{}_i x^i + b^j \;, \tag{3.1}$$

where $e_i = \sum_j A^j{}_i e'_j$ and $\overrightarrow{O'O} = \sum_j b^j e'_j$.
An **orthonormal Cartesian coordinate systems** on the Euclidean space $(\mathbb{E}^n, (\cdot|\cdot))$ is a Cartesian coordinate systems whose axes form an orthonormal basis of the vector space V associated with $\mathbb{E}^n$ with respect to $(\cdot|\cdot)$, namely $(e_i|e_j) = \delta_{ij}$.
In that case the matrices A of elements $A^j{}_i$ employed to connect different orthonormal Cartesian coordinate system are all of the elements of the **orthogonal group** of order n:

$$O(n) = \{R \in M(n, \mathbb{R}) \mid R^t R = I \, .\}$$

This is evident from the requirement that both the bases appearing in the identities $e_i = \sum_j A^j{}_i e'_j$ are orthonormal.

Remarks 3.12.
(1) The simplest example of affine space is $\mathbb{R}^n$ itself with $\mathbb{A}^n = \mathbb{R}^n$ and $V = \mathbb{R}^n$. The map $\vec{\cdot} : \mathbb{A}^n \times \mathbb{A}^n \to V$ in this case is trivial: $\overrightarrow{(x^1, \ldots, x^n)(y^1, \ldots, y^n)} = (y^1 - x^1, \ldots, y^n - x^n)$.
From a physical point of view affine and Euclidean spaces are more interesting than $\mathbb{R}^n$ to describe physical space (of inertial reference frames) because the affine structure can naturally be used to represent homogeneity and isotropy of space. Conversely $\mathbb{R}^n$ has a unphysical structure breaking those symmetries: it admits a preferred coordinate system!
(2) Three-dimensional Euclidean spaces are the spaces of usual Euclidean geometry.
(3) The whole class of Cartesian coordinate systems, allow the introduction of a Hausdorff second-countable topology on $\mathbb{A}^n$. This is done by fixing a Cartesian coordinate system $f : \mathbb{A}^n \to \mathbb{R}^n$ and defining the open sets on $\mathbb{A}^n$ as the set of the form $f^{-1}(B)$, $B \subset \mathbb{R}^n$ being any open set of $\mathbb{R}^n$. Since affine transformation (3.1) from $\mathbb{R}^n$ to $\mathbb{R}^n$ are homomorphisms, the given definition does not depend on the used Cartesian coordinate system. In this way, every Cartesian coordinate system becomes a homeomorphism and thus Hausdorff and second countability properties are inherited from the topology on $\mathbb{R}^n$.
(4) it is possible to give the notion of a *differentiable function* $h : A \to B$, where $A \subset \mathbb{A}^r$, $B \subset \mathbb{A}^s$ are open sets. In particular it could be either $\mathbb{A}^r = \mathbb{R}^r$ or $\mathbb{A}^s = \mathbb{R}^s$. ■

Definition 3.13. If $A \subset \mathbb{A}^r$, $B \subset \mathbb{A}^s$ are open sets in the corresponding affine spaces, we say that $h : A \to B$ is of **class** $C^k(A)$ if $f' \circ h \circ f^{-1} : f(A) \to \mathbb{R}^s$ is of class $C^k(f(A))$ for a pair of

Cartesian coordinate systems $f : \mathbb{A}^r \to \mathbb{R}^r$ and $f' : \mathbb{A}^s \to \mathbb{R}^s$. $\blacksquare$

The definition is well-posed, since affine transformations (3.1) are C^∞, the given definition does not matter the choice of f and f'.

Remarks 3.14. As a final remark we notice that, from a more abstract point of view, the previous comments proves that every n-dimensional affine (or euclidean) space has a natural structure of *n-dimensional C^∞ differentiable manifold* induced by the class of all Cartesian coordinate systems. $\blacksquare$

3.4.1 Tensor spaces, Cartesian tensors

Consider an $\mathbb{A}^n$ is an affine space and a Cartesian coordinate system $f : \mathbb{A}^n \ni p \mapsto (x^1(p), \ldots, x^n(p)) \in \mathbb{R}^n$ with origin O and axis $e_1, \ldots, e_n$. A standard notation to denote the elements of that basis and those of the associated dual basis is:

$$\frac{\partial}{\partial x^i} := e_i \;, \quad \text{and} \quad dx^i := e^{*i} \;, \quad \text{for } i = 1, \ldots, n. \tag{3.2}$$

This notation is standard notation in differential geometry, however we want not to address here the general definition of vector tangent and cotangent to a manifold. The simplest explanation of such a notation is that, under changes of coordinates (3.1), formal computations based on the fact that:

$$\frac{\partial x'^j}{\partial x^i} = A^j{}_i \;,$$

show that (writing the summation over repeated indices explicitly)

$$\frac{\partial}{\partial x^i} = \sum_{j=1}^n A^j{}_i \frac{\partial}{\partial x'^j} \;, \quad \text{and} \quad dx'^j = \sum_{i=1}^n A^j{}_i dx^i \;. \tag{3.3}$$

And, under the employed notation, those correspond to the *correct* relations

$$e_i = \sum_{j=1}^n A^j{}_i e'_j \;, \quad \text{and} \quad e'^{*j} = \sum_{i=1}^n A^j{}_i e^{*i} \;.$$

Very often, the notation for $\frac{\partial}{\partial x^i}$ is shortened to ∂_{x^i}. We make use of the shortened notation from now on.

Let us pass to consider tensors in the algebra $\mathscr{A}_{\mathbb{R}}(V)$. A Change of Cartesian coordinates

$$x'^j = \sum_{i=1}^n A^j{}_i x^i + b^j \;,$$

involves a corresponding change in the associated bases of V and V^*. Under these changes of bases, the transformation of components of contravariant tensors uses the same coefficients $A^j{}_i$ appearing in the relation connecting different Cartesian coordinate systems: if $t \in V \otimes V$ one has

$$t = t^{ij}\partial_{x^i} \otimes \partial_{x^j} = t'^{kl}\partial_{x'^i} \otimes \partial_{x'^l} \ , \quad \text{where } t'^{kl} = A^k{}_i A^l{}_j t^{ij}.$$

Covariant tensors uses the coefficients of the matrix A^{-1t} where A is the matrix of coefficients $A^j{}_i$. Therefore, if we restrict ourselves to employ only bases in V and V^* associated with Cartesian coordinate systems, the transformation law of components of tensors use the same coefficients as those appearing in the transformation law of Cartesian coordinates. For this reason tensors in affine spaces are said *Cartesian tensors*. A straightforward application of the given notation is the tangent vector to a differentiable curve in $\mathbb{A}^n$.

Definition 3.15. **(Tangent vector.)** Let $\mathbb{A}^n$ be an affine space with space of translations V. If $\gamma : (a, b) \to \mathbb{A}^n$ is $C^1((a, b))$, one defines the **tangent vector** $\dot\gamma(t)$ at γ in $\gamma(t)$ as

$$\dot\gamma(t) := \frac{dx^i}{dt}\partial_{x^i} \ ,$$

where $f : \mathbb{A}^n \ni p \mapsto (x^1(p), \ldots, x^n(p)) \in \mathbb{R}^n$ is any Cartesian coordinate system on $\mathbb{A}^n$. ■

The definition is well-posed because $\dot\gamma(t)$ does not depend on the chosen Cartesian coordinate system. Indeed, under a changes of coordinates (3.1) one finds

$$\frac{dx'^j}{dt} = A^j{}_i \frac{dx^i}{dt} \ ,$$

and thus

$$\frac{dx'^j}{dt}\partial_{x'^j} = A^j{}_i \frac{dx^i}{dt}\partial_{x'^j} \ ,$$

so that, by (3.3),

$$\frac{dx'^j}{dt}\partial_{x'^j} = \frac{dx^i}{dt}\partial_{x^i} \ .$$

■

3.4.2 Applied tensors

In several physical applications it is convenient to view a tensor as applied in a point $p \in \mathbb{A}^n$. This notion is based on the following straightforward definition which is equivalent to that given using the structure of differentiable manifold of an affine space.

Definition 3.16. **(Tangent and cotangent space.)** If $\mathbb{A}^n$ is an affine space with space of translations V and $p \in \mathbb{A}^n$, the **tangent space** at $\mathbb{A}^n$ in p is the vector space

$$T_p\mathbb{A}^n := \{(p, v) \mid v \in V\}$$

with vector space structure naturally induced by that of V (i.e. by the definition of linear composition of vectors

$$a(p,v)+b(p,u) := (p, au+bv)\ , \quad \text{for all } a,b \in \mathbb{R} \text{ and } u,v \in V.)$$

The **cotangent space** at $\mathbb{A}^n$ in p is the vector space

$$T^*_P\mathbb{A}^n := \{(p,v^*) \mid v^* \in V^*\}$$

with vector space structure naturally induced by that of V^*.
The bases of $T_p\mathbb{A}^n$ and $T^*_p\mathbb{A}^n$ associated with a Cartesian coordinate system $f : \mathbb{A}^n \ni p \mapsto (x^1(p),\ldots,x^n(p)) \in \mathbb{R}^n$ are respectively denoted by $\{\partial_{x^i}|_p\}_{i=1,\ldots,n}$ and $\{dx^i|_p\}_{i=1,\ldots,n}$ (but the indication of the application point p is very often omitted).
■

It is obvious that $T^*_p\mathbb{A}^n$ is canonically isomorphic to the dual space $(T_p\mathbb{A}^n)^*$ and every space of tensor $S_p \in \mathscr{A}_{\mathbb{R}}(T_p\mathbb{A}^n)$ is canonically isomorphic to a vector space

$$\{(p,t) \mid t \in S\}\ ,$$

where $S \in \mathscr{A}_{\mathbb{R}}(V)$.

Definition 3.17. **(Tensor fields.)** Referring to definition 3.16, the **tensor algebra of applied tensors** at $\mathbb{A}^n$ in p is the algebra $\mathscr{A}(T_p\mathbb{A}^n)$.
An assignment $\mathbb{A}^n \ni p \mapsto t_p \in S_p$ such that, every space $S_p \in \mathscr{A}(T_p\mathbb{A}^n)$ is of the same order, and the components of t_p define C^k functions with respect to the bases of tensors applied at every $p \in \mathbb{A}^n$ associated with a fixed Cartesian coordinate system, is a C^k **tensor field** on $\mathbb{A}^n$. The order of the tensor field is defined as the order of each tensor t_p.
■

Another remarkable aspect of Cartesian tensor fields id the fact that it is possible to coherently define a notion of derivative of each of them with respect to a vector field.

Definition 3.18. **(Derivative of tensor fields.)** Referring to definitions 3.16 and 3.17, given a C^k tensor field $\mathbb{A}^n \ni p \mapsto t_p \in S_p$, with $k \geq 1$ and contravariant vector field $\mathbb{A}^n \ni p \mapsto v_p$, the **derivative of** t **with respect to** v is the tensor field $\mathbb{A}^n \ni p \mapsto \nabla_v t_p \in S_p$ (so of the same type as t) that in every Cartesian coordinate system takes the components:

$$(\nabla_v t_p)^A = v^a(p)\, \frac{\partial t^A}{\partial x^a}\bigg|_p$$

The **derivative** of t (without fixing any vector field) is the tensor field $\mathbb{A}^n \ni p \mapsto \nabla_v t_p \in S_p$

such that, in every Cartesian coordinate system takes the components:

$$\nabla_a t^A(p) := (\nabla t_p)_a{}^A := t(p)^A{}_{,a} := \left.\frac{\partial t^A}{\partial x^a}\right|_p ,$$

so that:

$$(\nabla_v t)^A = v^a (\nabla_a t)^A = v^a t^A{}_{,a} .$$

■

The given definitions are well-posed in the sense that they do not depend on the choice of the used Cartesian coordinate system. This is due to the fact that the components of a fixed tensor field transforms, under changes of *Cartesian coordinates*, by means of a linear transformation with constant coefficients.The fact that these coefficients are *constant* implies that the definition of $(\nabla t)_a{}^A$ is well posed:

$$\frac{\partial t^{ij}}{\partial x^k} = \frac{\partial}{\partial x^k}\left(A^i{}_p A^j{}_q t'^{pq}\right) = B_k{}^l \frac{\partial}{\partial x'^l}\left(A^i{}_p A^j{}_q t'^{pq}\right) = B_k{}^l A^i{}_p A^j{}_q \frac{\partial}{\partial x'^l} t'^{pq} ,$$

so that:

$$\frac{\partial t^{ij}}{\partial x^k} = B_k{}^l A^i{}_p A^j{}_q \frac{\partial t'^{pq}}{\partial x'^l} .$$

3.4.3 The natural isomorphism between V and V^* for Cartesian vectors in Euclidean spaces

Let us consider an Euclidean space $\mathbb{E}^n$ and, from now on, *we restrict ourselves to only use orthonormal Cartesian coordinate frames*. We are going to discuss the existence of a natural isomorphism between V and V^* that we shall consider much more extensively later from a wider viewpoint.
If $\{e_i\}_{i=1,\ldots,n}$ and $\{e^{*j}\}_{j=1,\ldots,n}$ are an orthonormal basis in V and its associated basis in V^* respectively, we can consider the isomorphism:

$$H : V \ni v = v^i e_i \mapsto v_j e^{*j} \in V^* \quad \text{where } v_j := v^j \text{ for } j = 1, 2, \ldots, n.$$

In principle that isomorphism may seem to depend on the chosen basis $\{e_i\}_{i=1,\ldots,n}$, actually it does not! Indeed, if we chose another *orthonormal* basis and its dual one $\{e'_k\}_{k=1,\ldots,n}$ and $\{e'^{*h}\}_{h=1,\ldots,n}$, so that $e_i = A^j{}_i e'_j$ and $e^{*j} = B_h{}^j e'^{*h}$, the isomorphism above turns out to be written:

$$H : V \ni v = v'^j e'_j \mapsto v'_h e'^{*h} \in V^* \quad \text{where } v'_h := B_h{}^k (A^{-1})^k{}_j v'^j \text{ for } j = 1, 2, \ldots, n.$$

In fact, since we are simply writing down the same isomorphism H defined above, using another basis, v'_h are the components of $v_j e^{*j}$ but computed with respect to the basis $\{e'^{*h}\}_{h=1,\ldots,n}$ and so $v'_h = B_h{}^k v_k = B_h{}^k v^k = B_h{}^k (A^{-1})^k{}_j v'^j$.

We know that $B = A^{-1t}$, however, in our case $A \in O(n)$, so that $A^t = A^{-1}$ and thus $B = A$. Consequently:

$$B_h{}^k (A^{-1})^k{}_j = (AA^{-1})_{hj} = \delta_{hj} .$$

We conclude that the isomorphism H takes always the form

$$H : V \ni v = v^i f_i \mapsto v_j f^{*j} \in V^* \quad \text{where } v_j := v^j \text{ for } j = 1, 2, \ldots, n,$$

for every choice of the *orthonormal* basis $\{f_i\}_{i=1,\ldots,n} \subset V$. So H is actually independent from the initially chosen orthonormal basis and, in that sense, is natural. In other words, in the presence of a scalar product there is a natural identification of V and V^* that can be written down as, in the abstract index notation:

$$v^i \mapsto v_j \quad \text{where } v_j := v^j \text{ for } j = 1, 2, \ldots, n,$$

provided we confine ourselves to handle vectors by employing orthonormal basis only. Exploiting the universality property, this isomorphism easily propagates in the whole tensor algebra generated by V. So, for instance there exists natural isomorphisms between $V \otimes V^*$ and $V \otimes V$ or $V \otimes V^*$ and $V^* \otimes V$ acting as follows in the abstract index notation and referring to canonical bases associated with orthogonal bases in V:

$$t_i{}^j \mapsto t_{ij} := t_i{}^j ,$$

or, respectively

$$t_i{}^j \mapsto t^i{}_j := t_i{}^j .$$

In practice the height of indices does not matter in vector spaces equipped with a scalar product, provided one uses orthonormal bases only.
All that said immediately applies to Cartesian tensors in Euclidean spaces when V is the space of translations or is the tangent space $T_p\mathbb{E}^n$ at some point $p \in \mathbb{E}^n$.

Examples 3.19. In classical continuum mechanics, consider an internal portion $\mathscr{C}$ of a continuous body $\mathscr{B}$. Suppose that $\mathscr{C}$ is represented by a regular set in the Euclidean space $\mathbb{E}^3$: the physical space where the body is supposed to be at rest. Let $\partial\mathscr{C}$ be the boundary of $\mathscr{C}$ assumed to be a regular surface. If $p \in \partial\mathscr{C}$, the part of $\mathscr{B}$ external to $\mathscr{C}$ acts on $\mathscr{C}$ through the element of surface $dS \subset \partial\mathscr{C}$ about p, by a density of force called *stress vector* $\mathbf{s}(p, \mathbf{n})$. $\mathbf{n}$ denotes the versor orthogonal to dS at p. In this context, it is convenient to assume that: $\mathbf{s}(p, \mathbf{n}), \mathbf{n}_p \in T_p\mathbb{E}^3$.
In general one would expect a complicated dependence of the stress vector from $\mathbf{n}$. Actually a celebrated theorem due to Cauchy proves that (under standard hypotheses concerning the dynamics of the continuous body) this dependence must be linear. More precisely, if one defines: $\mathbf{s}(p, \mathbf{0}) := \mathbf{0}$ and $\mathbf{s}(p, \mathbf{v}) := |\mathbf{v}|\mathbf{s}(p, \mathbf{v}/|\mathbf{v}|)$, it turns out that $T_p\mathbb{E}^3 \ni \mathbf{v} \mapsto \mathbf{s}(p, \mathbf{v})$ is linear. As a consequence, there is a tensor $\sigma \in T_p\mathbb{E}^3 \otimes T_p^*\mathbb{E}^3$ (depending on the nature of $\mathscr{B}$ and on p) called *Cauchy stress tensor*, such that

$$\mathbf{s}(p, \mathbf{n}) = \sigma(\mathbf{n}) . \tag{3.4}$$

In order to go on with continuum mechanics it is assumed that, varying $p \in \mathbb{E}^3$, the map $p \mapsto \sigma(p)$ gives rise to a sufficiently regular (usually C^2) tensor field of order $(1,1)$.
Exploiting the abstract index notation:

$$s^i(p, \mathbf{n}) = \sigma(p)^i{}_j n^j .$$

It is possible to prove, taking the equation for the angular momentum into account that the stress tensor is symmetric. In other words, passing to the completely covariant expression for it (taking advantage) of the existence of the natural isomorphism between $V \otimes V^*$ and $V^* \otimes V^*$ – for $V = T_p\mathbb{E}^3$ – discussed above, one has:

$$\sigma(p)_{ij} = \sigma(p)_{ji} .$$

Chapter 4

Some application to general group theory

In this chapter we present a few applications of the theory developed previously in relation to group theory.

4.1 Groups

4.1.1 Basic notions on groups and matrix groups

As is known a **group** is an algebraic structure, $(G, \circ)$, where G is a set and $\circ : G \times G \to G$ is a map called the **composition rule** of the group or **group product**. Moreover the following three conditions have to be satisfied.
(1) $\circ$ is *associative*, i.e.,

$$g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3 \,, \quad \text{for all} \quad g_1, g_2, g_3 \in G \,.$$

(2) There is a **group unit** or **unit element**, i.e., there is $e \in G$ such that

$$e \circ g = g \circ e = g \,, \quad \text{for all} \quad g \in G \,.$$

(3) Each element $g \in G$ admits an **inverse element**, i.e.,

$$\text{for each} \quad g \in G \quad \text{there is} \quad g^{-1} \in G \quad \text{with} \quad g \circ g^{-1} = g^{-1} \circ g = e \,.$$

We remind the reader that the unit element turns out to be unique and so does the inverse element for each element of the group (the reader might show those uniqueness properties as an exercise). A group $(G, \circ)$ is said to be **commutative** or **Abelian** if $g \circ g' = g' \circ g$ for each pair of elements, $g, g' \in G$; otherwise it is said to be **non-commutative** or **non-Abelian**. A subset $G' \subset G$ of a group is called **subgroup** if it is a group with respect to the restriction to $G' \times G'$ of the composition rule of G. A subgroup N of a group G is called a **normal subgroup** if it is invariant under **conjugation**, that is, for each element $n \in N$ and each $g \in G$, the element $g \circ n \circ g^{-1}$ is still in N.

If $(G_1, \circ_1)$ and $(G_2, \circ_2)$ are groups, a group **homomorphism** from G_1 to G_2 is a map $h : G_1 \to G_2$ which *preserves the group structure*, i.e., the following requirement has to be fulfilled:

$$h(g \circ_1 g') = h(g) \circ_2 h(g') \quad \text{for all} \quad g, g' \in G_1 \,,$$

As a consequence, they also hold with obvious notations:

$$h(e_1) = e_2 \,,$$

and

$$h(g^{-1_1}) = (h(g))^{-1_2} \quad \text{for each} \quad g \in G_1 \,.$$

Indeed, if $g \in G_1$ one has $h(g) \circ e_2 = h(g) = h(g \circ e_1) = h(g) \circ h(e_1)$. Applying $h(g)^{-1}$ on the left, one finds $e_2 = h(e_1)$. On the other hand $h(g)^{-1} \circ h(g) = e_2 = h(e_1) = h(g^{-1} \circ g) = h(g^{-1}) \circ h(g)$ implies $h(g)^{-1} = h(g^{-1})$.
The **kernel** of a group homomorphism, $h : G \to G'$, is the subgroup $K \subset G$ whose elements g satisfy $h(g) = e'$, e' being the unit element of G'. Obviously, h is injective if and only if its kernel contains the unit element only. A group **isomorphism** is a *bijective* group homomorphism. A group isomorphism $h : G \to G$, so that the domain and the co-domain are the same group, is called group **automorphism** on G. The set of group automorphism on a given group G is denoted by $Aut(G)$ and it is a group in its own right when the group product is the standard composition of maps.

Examples 4.1. The algebra $M(n, \mathbb{K})$, that is the vector space of the matrices $n \times n$ with coefficients in $\mathbb{K}$ equipped with the matrix product, contains several interesting *matrix groups.* If $n > 0$, most interesting groups in $M(n, \mathbb{K})$ are non-commutative.
(1) The first example is $GL(n, \mathbb{K})$ which is the set of the $n \times n$ matrices A with components in the field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$ and $\det A \neq 0$. It is a group with group composition rule given by the usual product of matrices.
(2) An important subgroup of $GL(n, \mathbb{K})$ is the *special group* $SL(n, \mathbb{K})$, i.e., the set of matrices A in $GL(n, \mathbb{K})$ with $\det A = 1$. (The reader might show that $SL(n, \mathbb{K})$ is a subgroup of $GL(n, \mathbb{K})$.)
(3) If $\mathbb{K} = \mathbb{R}$ another interesting subgroup is the *orthogonal group* $O(n)$ containing all of the matrices satisfying $R^t R = I$.
(4) If $\mathbb{K} = \mathbb{C}$ an interesting subgroup of $GL(n, \mathbb{C})$ is the *unitary group* $U(n)$ (and the associated subgroup $SU(n) := SL(n, \mathbb{C}) \cap U(n)$), containing the matrices A satisfying $\overline{A^t} A = I$ (the bar denoting the complex conjugation) is of great relevance in quantum physics.
(5) In section 5.2.1 we shall encounter the pseudo orthogonal groups $O(m, p)$ and, one of them, the Lorentz group $O(1, 3)$ will be studied into details in the last two chapters.

Notice that there are groups which *are not defined* as group of matrices, e.g., $(\mathbb{Z}, +)$. Some of them can be represented in terms of matrices anyway (for example $(\mathbb{Z}, +)$ can be represented as a matrix subgroup of $GL(2,\mathbb{R})$, we leave the proof as an exercise). There are however groups which cannot be represented in terms of matrices as the so-called *universal covering of* $SL(2,\mathbb{R})$. An example of a group which is not defined as a group of matrices (but it admits such representations) is given by the *group of permutations of* n *elements* which we shall consider in the next subsection.

Exercises 4.2.

1. Prove the uniqueness of the unit element and the inverse element in any group.

2. Show that in any group G the unique element e such that $e^2 = e$ is the unit element.

3. Show that if G' is a subgroup of G the unit element of G' must coincide with the unit element of G, and, if $g \in G'$, the inverse element g^{-1} in G' coincides with the inverse element in G.

4. Show that if $h : G_1 \to G_2$ is a group homomorphism, then $h(G_1)$ is a subgroup of G_2.

4.1.2 Direct product and semi-direct product of groups

We remind here two important elementary notions of group theory, the direct product and the semi-direct product of groups.

If G_1 and G_2 are groups, their **direct product**, $G_1 \times G_2$, is another group defined as follows. The elements of $G_1 \times G_2$ are, as the notation suggests, the elements (g_1, g_2) of the *Cartesian product* of the *sets* G_1 and G_2. Moreover the composition rule is $(g_1, g_2) \circ (f_1, f_2) := (g_1 \circ_1 f_1, g_2 \circ_2 f_2)$ for all $(g_1, g_2), (f_1, f_2) \in G_1 \times G_2$. Obviously the unit of $G_1 \times G_2$ is (e_1, e_2), where e_1 and e_2 are the unit elements of G_1 and G_2 respectively. The proof that these definitions determine a correct structure of group on the set $G_1 \times G_2$ is trivial and it is left to the reader. Notice that, with the given definition of direct product, G_1 and G_2 turn out to be normal subgroups of the direct product $G_1 \times G_2$.

Sometimes there are gropus with a structure close to that of direct product but a bit more complicated. These structure are relevant, especially in physics and are captured within the following definition.
Suppose that $(G_1, \circ_1)$ and $(G_2, \circ_2)$ are groups and for every $g_1 \in G_1$ there is a group isomorphism $\psi_{g_1} : G_2 \to G_2$ satisfying the property: $\psi_{g_1} \circ \psi_{g_1'} = \psi_{g_1 \circ_1 g_1'}$, which easily implies $\psi_{e_1} = id_{G_2}$ as previously discussed, where $\circ$ is the usual composition of functions and e_1 the unit element of G_1. In other words, $\psi_g \in Aut(G_2)$ for every $g \in G_1$ and the map $\psi : G_1 \ni g \mapsto \psi_g$ is a group homomorphism from G_1 to $Aut(G_2)$ viewed as a group with respect to the composition of maps. In this case, a natural structure of group can be assigned on the Cartesian product $G_1 \times G_2$. This is done by defining the composition rule between a pair of elements, $(g_1, g_2), (f_1, f_2) \in G_1 \times G_2$, as

$$(g_1, g_2) \circ_\psi (f_1, f_2) := (g_1 \circ_1 f_1,\ g_2 \circ_2 \psi_{g_1}(f_2)) .$$

We leave to the reader the strightforward proof of the well posedness of the given group composition rule. Again, the unit element turns out to be (e_1, e_2). The obtained group structure

$(G_1 \times_\psi G_2, \circ_\psi)$, is called the **semi-direct product** of G_1 and G_2.

Examples 4.3.
(1) Consider the group $O(3)$ of orthogonal matrices in $\mathbb{R}^3$, and $\mathbb{R}^3$ itself viewed as additive Abelian group. Define, for every $R \in O(3)$, the automorphism $\psi_R \in Aut(\mathbb{R}^3)$ as $\psi_R : \mathbb{R}^3 \ni \mathbf{v} \mapsto R\mathbf{v} \in \mathbb{R}^3$. We can endow $O(3) \times \mathbb{R}^3$ with the structure of semi-direct product group $O(3) \times_\psi \mathbb{R}^3$ when definig the composition rule

$$(R, \mathbf{v}) \circ_\psi (R', \mathbf{v}') := (RR', \mathbf{v} + R\mathbf{v}') .$$

The meaning of the structure so defined should be evident: A pair $(R, \mathbf{v})$ is a roto-translation, i.e. a generic isometry of the space $\mathbb{R}^3$ (equipped with the standard metric structure). In other words $(R, \mathbf{v})$ acts of a point $\mathbf{u} \in \mathbb{R}^3$ with a rotation followed by a translation, $(R, \mathbf{v}) : \mathbf{u} \mapsto \mathbf{v} + R\mathbf{u}$. The composition rule $\circ_\psi$ is nothing but a composition of two such transformations. It is possible to prove that every isometry of $\mathbb{R}^3$ equipped with the standard structure of Euclidean space, is necessarily an element of $O(3) \times_\psi \mathbb{R}^3$.
(2) As further examples, we may mention the Poincaré group as will be presented in definition 8.26, the group of isometries of an Eucliedean (affine) space, the Galileo group in classical mechanics, the BMS group in general relativity, the inhomogeneous $SL(2, \mathbb{C})$ group used in relativistic spinor theory. Similar structures, in physics, appear very frequently.

Exercises 4.4.

1. Consider a semi-dirct product of groups $(G \times_\psi N, \circ_\psi)$ as defined beforehand (notice that the role of G and N is not interchangeable). Prove that N is a normal subgroup of $G \times_\psi N$ and that

$$\psi_g(n) = g \circ_\psi n \circ_\psi g^{-1} \quad \text{for every } g \in G \text{ and } n \in N.$$

2. Consider a group $(H, \circ)$, let N be a normal subgroup of H and let G be a subgroup of H. Suppose that $N \cap G = \{e\}$, where e is the unit element of H, and that $H = GN$, in the sense that, for every $h \in H$ there are $g \in G$ and $n \in N$ such that $h = gn$. Prove that (g, n) is uniquely determined by h and that H is isomorphic to $G \times_\psi N$, where

$$\psi_g(n) := g \circ n \circ g^{-1} \quad \text{for every } g \in G \text{ and } n \in N.$$

4.2 Tensor products of group representations

4.2.1 Linear representation of groups and tensor representations

We are interested in the notion of (linear) *representation of a group on a vector space*. In order to state the corresponding definition, notice that, if V is a (not necessarily finite-dimensional) vector space, $\mathscr{L}(V|V)$ contains an important group. This is $GL(V)$ which is the set of both injective and surjective elements of $\mathscr{L}(V|V)$ equipped with the usual composition rule of maps. We can give the following definition.

Definition 4.5. **(Linear group on a vector space.)** If V is a vector space, $GL(V)$ denotes the group of linear maps $f : V \to V$ such that f is injective and surjective, with group composition rule given by the usual maps composition. $GL(V)$ is called the **linear group on** V. ■

Remarks 4.6.
(1) If $V := \mathbb{K}^n$ then $GL(V) = GL(n, \mathbb{K})$.
(2) If $V \neq V'$ it is not possible to define the analogue of the group $GL(V)$ considering some subset of $\mathscr{L}(V|V')$. (The reader should explain the reason.)
(3) Due to theorem 3.8, if $dimV < +\infty$, we have that $GL(V)$ coincides to a subset of $V \otimes V^*$.■

Definition 4.7. **(Linear group representation on a vector space.)** Let $(G, \circ)$ be a group and V a vector space. A (linear group) **representation of** G **on** V is a homomorphism $\rho : G \to GL(V)$. Moreover a representation $\rho : G \to GL(V)$ is called:
(1) faithful if it is injective,
(2) free if, for any $v \in V \setminus \{0\}$, the subgroup of G made of the elements h_v such that $\rho(h_v)v = v$ contains only the unit element of G,
(3) transitive if for each pair $v, v' \in V \setminus \{0\}$ there is $g \in G$ with $v' = \rho(g)v$.
(4) irreducible if there is no *proper* vector subspace $S \subset V$ which is **invariant** under the action of $\rho(G)$, i.e., which satisfies $\rho(g)S \subset S$ for all $g \in G$.
Finally, the representation $\rho^d : G \to GL(V^*)$ defined by

$$\rho^d(g) := \rho(g^{-1})^* \quad \text{for all } g \in G,$$

where $\rho(g^{-1})^* : V^* \to V^*$ is the adjoint operator of $\rho(g^{-1}) : V \to V$, as definied in definition 2.14, is called the **dual representation** associated with ρ. ■

Remarks 4.8.
(1) The definitions given above representations can be extended to the case of general (nonlinear) representations of a group G and replacing $GL(V)$ for a group of bijective transformations $f : V \to V$. Assuming that V is a differential manifold $GL(V)$ can be replaced for the group $Diff(V)$ of d*iffeomorphisms* of V. In that case the homomorphism $\rho : G \to Diff(V)$ is a representation of G in terms of diffeomorphisms. When V is a Riemannian manifold, representations in terms of *isometries* can be similarly defined. In these extended contexts the definitions of faithful, free (dropping the constraint $v \neq 0$) and transitive (dropping the constraint $v, v' \neq 0$)representations can be equally stated.
(2) The presence of the inverse element in the definition of $\rho^d(g) := \rho(g^{-1})^*$ accounts for the fact that adjoint operators compose in the reverse order, i.e. $(T_1T_2)^* = T_2^*T_1^*$. Thanks to the inverse element g^{-1} in the definition of $\rho^d(g)$, one gets $\rho^d(g_1)\rho^d(g_2) = \rho^d(g_1g_2)$ as requested for representations, instead of $\rho^d(g_1)\rho^d(g_2) = \rho^d(g_2g_1)$ as it would be if simply defining $\rho^d(g) := \rho(g)^*$.
■

Equipped with the above-given definitions we are able to study the simplest interplay of tensors and group representations. We want to show that the notion of tensor product allows the definitions of *tensor products of representations.* That mathematical object is of fundamental importance in applications to Quantum Mechanics, in particular as far as systems with many components are concerned.
Consider a group G (from now on we omit to specify the symbol of the composition rule whenever it does not produces misunderstandings) and several representations of the group $\rho_i : G \to GL(V_i)$, where $V_1, \ldots, V_n$ are finite-dimensional vector spaces on the same field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$. For each $g \in G$, we may define a multi-linear map $[\rho_1(g), \ldots, \rho_n(g)] \in \mathscr{L}(V_1, \ldots, V_n | V_1 \otimes \ldots \otimes V_n)$ given by, for all $(\rho_1(g), \ldots, \rho_n(g)) \in V_1 \times \ldots \times V_n$,

$$[\rho_1(g), \ldots, \rho_n(g)] : (v_1, \ldots, v_n) \mapsto (\rho_1(g)v_1) \otimes \ldots \otimes (\rho_n(g)v_n) .$$

That map is multi linear because of the multi linearity of the tensor-product map and the linearity of the operators $\rho_k(g)$. Using the universality theorem, we *uniquely* find a linear map which we indicate by $\rho_1(g) \otimes \ldots \otimes \rho_n(g) : V_1 \otimes \ldots \otimes V_n \to V_1 \otimes \ldots \otimes V_n$ such that:

$$\rho_1(g) \otimes \ldots \otimes \rho_n(g)(v_1 \otimes \ldots \otimes v_n) = (\rho_1(g)v_1) \otimes \ldots \otimes (\rho_n(g)v_n) .$$

Definition 4.9. **(Tensor product of representations.)** Let $V_1, \ldots V_n$ be finite-dimensional vector spaces on the same field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$ and suppose there are n representations $\rho_i : G \to GL(V_k)$ of the same group G on the given vector spaces. The set of linear maps

$$\{\rho_1(g) \otimes \ldots \otimes \rho_n(g) : V_1 \otimes \ldots \otimes V_n \to V_1 \otimes \ldots \otimes V_n \mid g \in G\} ,$$

defined above is called **tensor product of representations** $\rho_1, \ldots, \rho_n$. ■

The relevance of the definition above is evident because of the following theorem,

Theorem 4.10. *Referring to the definition above, the map $G \ni g \mapsto \rho_1(g) \otimes \ldots \otimes \rho_n(g)$ is a* **linear group representation of G on the tensor-product space** $V_1 \otimes \ldots \otimes V_n$.

Proof. We have to show that the map

$$g \mapsto \rho_1(g) \otimes \ldots \otimes \rho_n(g) ,$$

is a group homomorphism from G to $GL(V_1 \otimes \ldots \otimes V_n)$. Using the fact that each ρ_i is a group homomorphism, if $g, g' \in G$, one has

$$\rho_1(g') \otimes \ldots \otimes \rho_n(g')(\rho_1(g) \otimes \ldots \otimes \rho_n(g)(v_1 \otimes \ldots \otimes v_n)) = \rho_1(g') \otimes \ldots \otimes \rho_n(g')((\rho_1(g)v_1) \otimes \ldots \otimes (\rho_n(g)v_n))$$

and this is

$$(\rho_1(g' \circ g)v_1) \otimes \ldots \otimes (\rho_n(g' \circ g)v_n) .$$

The obtained result holds true also using a canonical basis for $V_1 \otimes \ldots \otimes V_n$ made of usual elements $e_{1,i_1} \otimes \ldots \otimes e_{n,i_n}$ in place of $v_1 \otimes \ldots \otimes v_n$. By linearity, it means that the found identity is valid if replacing $v_1 \otimes \cdots \otimes v_n$ for every tensor in $V_1 \otimes \cdots \otimes V_n$. Therefore,

$$(\rho_1(g') \otimes \ldots \otimes \rho_n(g'))(\rho_1(g) \otimes \ldots \otimes \rho_n(g)) = \rho_1(g' \circ g) \otimes \ldots \otimes \rho_n(g' \circ g) .$$

To conclude, notice that $\rho_1(g) \otimes \ldots \otimes \rho_n(g) \in GL(V_1 \otimes \ldots \otimes V_n)$ because $\rho_1(g) \otimes \ldots \otimes \rho_n(g)$ is (1) linear and furthermore it is (2) bijective. The latter can be proved as follows:

$$\rho_1(g^{-1}) \otimes \ldots \otimes \rho_n(g^{-1}) \circ (\rho_1(g) \otimes \ldots \otimes \rho_n(g)) = (\rho_1(g) \otimes \ldots \otimes \rho_n(g)) \circ \rho_1(g^{-1}) \otimes \ldots \otimes \rho_n(g^{-1})$$

$$= \rho_1(e) \otimes \ldots \otimes \rho_n(e) = I .$$

The last identity follows by linearity form $(\rho_1(e) \otimes \ldots \otimes \rho_n(e))(v_1 \otimes \cdots \otimes v_n) = v_1 \otimes \cdots \otimes v_n$. $\square$

More generally, if $A_k : V_k \to U_k$ are n linear maps (operators), and all involved vector spaces are finite dimensional and with the same field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$, it is defined the *tensor product of operators*.

Definition 4.11. **(Tensor Product of Operators.)** If $A_k : V_k \to U_k$, $k = 1, \ldots, n$ are n linear maps (operators), and all the vector spaces U_i, V_j are finite dimensional with the same field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$, the **tensor product of** $A_1, \ldots, A_n$ is the linear map

$$A_1 \otimes \ldots \otimes A_n : V_1 \otimes \ldots \otimes V_n \to U_1 \otimes \ldots \otimes U_n$$

uniquely determined by the universality theorem and the requirement:

$$(A_1 \otimes \ldots \otimes A_n) \circ \otimes = A_1 \times \ldots \times A_n ,$$

where

$$A_1 \times \ldots \times A_n : V_1 \times \ldots \times V_n \to U_1 \otimes \ldots \otimes U_n ,$$

is the multi-linear map such that, for all $(v_1, \ldots, v_n) \in V_1 \times \ldots \times V_n$:

$$A_1 \times \ldots \times A_n : (v_1, \ldots, v_n) \mapsto (A_1 v_1) \otimes \ldots \otimes (A_n v_n) .$$

$\blacksquare$

Remarks 4.12.
(1) Employing the given definition and the same proof used for the relevant part of the proof of theorem 4.10, it is simply shown that if $A_i : V_i \to U_i$ and $B_i : U_i \to W_i$, $i = 1, \ldots, n$ are $2n$ linear maps and all involved spaces V_i, U_j, W_k are finite dimensional with the same field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$, then

$$B_1 \otimes \ldots \otimes B_n \;\; A_1 \otimes \ldots \otimes A_n = (B_1 A_1) \otimes \ldots \otimes (B_n A_n) .$$

(2) Suppose that $A : V \to V$ and $B : U \to U$ are linear operators. How does $A \otimes B$ work in components? Or, that is the same, what is the explicit expression of $A \otimes B$ when employing the abstract index notation?
To answer we will make use of Theorem 3.8. Fix a basis $\{e_i\}_{i\in I} \subset V$ and another $\{f_j\}_{j\in J} \subset U$ and consider the associated canonical bases in tensor spaces constructed out of V, U and their dual spaces. In view of the mentioned theorem we know that if $v \in V$ and $u \in U$ then indicating again by $A \in V \otimes V^*$ and $B \in U \otimes U^*$ the tensors associated with the linear maps A and B, respectively:

$$(Av)^i = A^i{}_j v^j \quad \text{and} \quad (Bu)^r = B^r{}_s u^s \,.$$

Passing to the tensor products and making use of the definition of tensor product of linear operators:

$$((A \otimes B)(v \otimes u))^{ir} = ((Av) \otimes (Bu))^{ir} = A^i{}_j v^j B^r{}_s u^s = A^i{}_j B^r{}_s v^j u^s \,.$$

Now, since $A \otimes B : V \otimes U \to V \otimes U$ is linear, we can write, in general for $t = t^{rs} e_r \otimes f_s$:

$$((A \otimes B)t)^{ir} = A^i{}_j B^r{}_s t^{js} \,. \tag{4.1}$$

We conclude that the tensor in $V \otimes V^* \otimes U \otimes U^*$ associated with $A \otimes B$ has components

$$A^i{}_j B^r{}_s \,,$$

and, as usually, the action of $A \otimes B : V \otimes U \to V \otimes U$ on $t \in V \otimes U$ is obtained by the tensor product of t and the tensor associated to $A \otimes B$, followed by a suitable contraction, as written in (4.1). ■

4.2.2 An example from Continuum Mechanics

Referring to the example 3.19, consider once more an internal portion $\mathscr{C}$ of a continuous body $\mathscr{B}$. Suppose that $\mathscr{C}$ is represented by a regular set in the Euclidean space $\mathbb{E}^3$. As we said there is a tensor $\sigma_p \in T_p\mathbb{E}^3 \otimes T_p^*\mathbb{E}^3$ (depending on the nature of $\mathscr{B}$ and on p) called Cauchy stress tensor, such that

$$\mathbf{s}(p, \mathbf{n}) = \sigma_p(\mathbf{n}) \,. \tag{4.2}$$

We henceforth write σ in place of σ_p for shortness, as p is supposed to be fixed.
Suppose now to rotate $\mathscr{B}$ about p with a rotation $R \in SO(3)$. If the body is isolated from any other external system, and the rest frame is inertial, assuming *isotropy of the space*, we expect that the relationship between $\mathbf{s}, \mathbf{n}$ and σ "remains fixed" under the rotation R, provided one replaces the vectors $\mathbf{s}$, $\mathbf{n}$ and the tensor σ with the corresponding objects transformed under R. How does the rotation R act on the tensor σ?
We start from the action of a rotation on vectors. We stress that we are assuming here an *active* point of view, a vector is transformed into another different vector under the rotation. It is not a

passive change of basis for a fixed vector. Fix an orthonormal basis $e_1, e_2, e_3 \in T_p\mathbb{E}^3$. Referring to that basis, the action of $SO(3)$ is, if $\mathbf{v} = v^i e_i$ and R is the matrix of coefficients $R^i{}_j$:

$$\mathbf{v} \mapsto \rho_R \mathbf{v}$$

where

$$(\rho_R \mathbf{v})^i := R^i{}_j v^j \ .$$

It is a trivial task to show that

$$\rho_R \rho_{R'} = \rho_{RR'} \ , \quad \text{for all } R, R' \in SO(3) \ ,$$

therefore $SO(3) \ni R \mapsto \rho_R$ defines, in fact, a (quite trivial) representation of $SO(3)$ on $T_p\mathbb{E}^3$ called the **fundamental representation** of $SO(3)$. This representation acts both on $\mathbf{n}$ and $\mathbf{s}$ and enjoys a straightforward physical meaning.

Let us pass to the tensor $\sigma \in T_p\mathbb{E}^3 \otimes T_p^*\mathbb{E}^3$. In the bases $\{e_i\}_{i=1,2,3}$ and the associated dual one $\{e^{*j}\}_{j=1,2,3}$, the relation (3.4) reads:

$$s^i = \sigma^i{}_j n^j \ ,$$

and thus, by action of a rotation R, one gets

$$(R\mathbf{s})^i = R^i{}_k \sigma^k{}_j n^j \ ,$$

or also,

$$(R\mathbf{s})^i = R^i{}_k \sigma^k{}_h (R^{-1})^h{}_j (R\mathbf{n})^j \ .$$

We conclude that, if we require that the relation (4.2) is preserved under rotations, we must define the action of $R \in SO(3)$ on the tensors $\sigma \in T_p\mathbb{E}^3 \otimes T_p^*\mathbb{E}^3$ given by:

$$\sigma \mapsto \gamma_R \sigma$$

where, *in the said bases*:

$$(\gamma_R \sigma)^i{}_j = R^i{}_k \, (R^{-1})^h{}_j \, \sigma^k{}_h \ . \tag{4.3}$$

It is a trivial task to show that

$$\gamma_R \gamma_{R'} = \gamma_{RR'} \ , \quad \text{for all } R, R' \in SO(3) \ ,$$

therefore $SO(3) \ni R \mapsto \gamma_R$ defines a representation of $SO(3)$ on $T_p\mathbb{E}^3 \otimes T_p^*\mathbb{E}^3$. This representation is the tensor product of ρ acting on $T_p\mathbb{E}^3$ and another representation of $SO(3)$, ρ' acting on $T_p^*\mathbb{E}^3$ and defined as follows. If $\mathbf{w} = w_j e^{*j} \in T_p^*\mathbb{E}^3$ and $R \in SO(3)$ is the matrix of coefficients $R^i{}_j$:

$$\mathbf{w} \mapsto \rho'_R \mathbf{w}$$

where, referring to the dual basis $e^{*1}, e^{*2}, e^{*3} \in T_p^*\mathbb{E}^3$,

$$(\rho'_R \mathbf{w})_j := w_i (R^t)^i{}_j \ .$$

Indeed one has:

$$\gamma_R = \rho_R \otimes \rho'_R \ , \quad \text{for all } R, R' \in SO(3) \ .$$

It is simply proved that ρ' is nothing but the dual representation associated with ρ as defined in definition 4.7. We leave the immediate proof to the reader.

4.2.3 The Schur lemma and an application to continuum mechanics

The Shur lemma actually consists of a number of different but similar statements which characterize irreducibile linear representations of groups on vecot spaces. Here we deal with a basic version of the lemma. There are many proofs, the proof we present here can be easily extended to the theory of unitary-projective representatios in Hilbert space [Moretti-a, Moretti-b].

Theorem 4.13. **(Schur's lemma in $\mathbb{R}^n$ and $\mathbb{C}^n$.)** *Assume that $\mathbb{K} := \mathbb{R}$ or $\mathbb{C}$. Let $\rho : G \ni g \mapsto GL(n,\mathbb{K})$ be a group representation in terms of isometric matrices (i.e., orthogonal if $\mathbb{K} = \mathbb{R}$ or unitary if $\mathbb{K} = \mathbb{C}$.). The representation ρ is irreducible if and only if the* **commutant** *of ρ*

$$\rho' := \{A \in M(n,\mathbb{K}) \mid A\rho_g = \rho_g A\,, \quad \forall g \in G\}$$

is trivial, i.e., it satiasfies

$$\rho' = \mathbb{K}I := \{cI \mid c \in \mathbb{K}\}\,.$$

Proof. We prove the thesis in the case $\mathbb{K} = \mathbb{C}$ (the other case being trivially analogous) and, to this end, we equipp $\mathbb{C}^n$ with the standard Hermitian scalar porduct $\langle u|v\rangle := \sum_{k=1}^n \overline{u^k} v^k$. The Hermitian adjoint operator of $A \in M(n,\mathbb{C})$ will be denoted by $A^\dagger := \overline{A}^t$, where the bar indicates the complex conjugation. By definition $\langle u|Av\rangle = \langle A^\dagger u|v\rangle$ for all $u, v \in \mathbb{C}^n$. By hypothesis $\rho_g^\dagger = (\rho_g)^{-1} = \rho_{g^{-1}}$, since the representation is made of unitary operators. Notice that, if $A \in \rho'$, taking the Hermitian adjoint of $A\rho_g = \rho_g A$ we have that $\rho_{g^{-1}} A^\dagger = A^\dagger \rho_{g^{-1}}$ so that $\rho_h A^\dagger = A^\dagger \rho_h$ for all $h \in G$ (since g^{-1} ranges throughout G when g does) and thus $A \in \rho'$ implies $A^\dagger \in \rho'$. The commutant is closed with respect the Hermitian conjugation. As a consequence, since the commutant is a complex vector space, $B := A + A^\dagger \in \rho'$ and $C := i(A - A^\dagger) \in \rho'$ provided $A \in \rho'$.

We now prove that, if ρ' is not trivial, then ρ is not irreducible.
Suppose that $A \in \rho'$ is not of the form aI. It is not possible that both $B = bI$ and $C = cI$ for some $b, c \in \mathbb{C}$ since it would imply $A = aI$ for some $a \in \mathbb{C}$ and this is not permitted. Suppose that B is not of the said form, the other case is similar. As $B = B^\dagger$, it admits an orthogonal decomposition in terms of eigenvectors. Let $\lambda \neq 0$ an eigenvalue (if all eigenvalues are 0, $B = 0 = 0I$ which is not possible). Notice that $H_\lambda \neq \{0\}$ as we said, but also $H_\lambda \neq H$, otherwise $B = \lambda I$ which we have already excluded. Hence H_λ is a proper non-trivial subspace. On the other hand $u \in H_\lambda$ satisfies $A\rho_g u = \rho_g A u = \lambda \rho_g u$ so that $\rho_g u \in H_\lambda$ and $\rho_g(H_\lambda) \subset H_\lambda$ for every $g \in G$. In summary, ρ cannot be irreducible since it admits a proper non-trivial invariant subspace.

We now prove that, if ρ is not irreducible, then the commutant ρ' is not trivial.
Let $H \subset \mathbb{C}^n$ a proper non trivial subspace which is invariant under ρ: $\rho_g(H) \subset H$. If $u_, \ldots u_m$ is an orthonormal basis of H, define $P\cdot := \sum_{j=1}^m \langle \cdot|u_j\rangle u_j$. This operator is different from I (because $\dim(H) = m < n$) and $P \neq 0$ because $\dim(H) > 0$. P is the orthogonal projector onto H. By construction $P^\dagger = P$, $PP = P$, $P(\mathbb{C}^n) = H$. Saying that $x \in H$ is equivalent to writing that $Px = x$. Since H is invariant, $\rho_g P u \in H$ for every $u \in \mathbb{C}^n$, so that $P\rho_g P v = \rho_g P v$. Hence $P\rho_g P = \rho_g P$ because v is arbitrary. Taking the Hermitian adjoint of both sides,

$P^\dagger \rho_{g^{-1}} P^\dagger = P^\dagger \rho_{g^{-1}}$. But, since, $P^\dagger = P$ and g^{-1} is arbitray, the obtained identity boils down to $P\rho_g P = P\rho_g$. Comparing with $P\rho_g P = \rho_g P$, we conclude that $\rho_g P = P\rho_g$ where P is not of the form cI by construction. $\square$

Let us come back to the previous example of the representation (4.3) γ_R, for $R \in SO(3)$, on the stress-tensor of a continuous body. We assume that not only the space but also the body is *isotropic*, in the sense that

$$R\sigma_p(\mathbf{n}) = \sigma_p(R\mathbf{n})$$

for every point p in the body, $R \in SO(3)$ and unit vector $\mathbf{n}$. *Notice that R acts here around p, which is viewed as the rigin of the rotation of the vectors $\mathbf{n}$ applied to p.* All *ideal fluids at rest in a reference space* satisfy (by definition) this hypothesis. If representing everything in an orthonormal basis of $\mathbb{R}^3$, taking into account linearity of σ_p and R, we can extend the above identity to

$$R\sigma_p(\mathbf{v}) = \sigma_p(R\mathbf{v})$$

for every $\mathbf{v} \in \mathbb{R}^3$ and $R \in SO(3)$. Namely

$$R\sigma_p = \sigma_p R\,, \quad \forall R \in SO(3)\,.$$

Now observe that the fundamental representation $\rho : SO(3) \ni R \mapsto R \in GL(3,\mathbb{R})$ is irreducible. That is because from every fixed unit vector we produce the entiere 2-sphere by applying all rotations of $SO(3)$, so, no proper non-trivial $SO(3)$-invariant subspaces can exist. Schur's lemma immediately entails that

$$\sigma_{p\,j}^{i} = -\mathcal{P}(p)\delta^i{}_j\,,$$

for some scalar function of the place $\mathcal{P}$ (the sign above is conventional and $\mathcal{P}$ is positive if a volume of the continuum is compressed). This function is the known *pressure* at p.

4.2.4 An example from Quantum Mechanics

Physicists are involved with Hilbert spaces whenever they handle quantum mechanics. A Hilbert space is noting but a complex vector space equipped with a Hermitian scalar product (see the next chapter) such that it is complete with respect to the norm topology induced by that scalar product. However, here we analyze the structure of vector space only. Physically speaking, the vectors of the Hilbert space represent the states of the considered physical system (actually things are more complicated but we do not matter). To consider the simplest case we assume that the vector space which describes the states of the system is finite-dimensional (that is the case for the spin part of a quantum particle). Moreover, physics implies that the space $\mathcal{H}$ of the states of a composite system S made of two systems S_1 and S_2 associated with Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, is the Hilbert tensor product (see section 2.2.4 for more details) $\mathcal{H} = \mathcal{H}_1 \otimes_H \mathcal{H}_2$. As far as this example is concerned, the reader may omit the adjective "Hilbert" and think of $\mathcal{H}$ as the standard algebraic tensor product $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, it being correct when $\mathcal{H}_1, \mathcal{H}_2$ are finite-dimensional. Let the system S_1 be described by a state $\psi \in \mathcal{H}_1$ and suppose to transform the

system by the action of an element R of some physical group of transformations $\mathcal{G}$ (e.g. $SO(3)$). The transformed state ψ' is given by $U_R^{(1)}\psi$ where $\mathcal{G} \ni R \mapsto U_R^{(1)}$ is a representation of $\mathcal{G}$ in terms of linear transformations $U_R^{(1)} : \mathcal{H}_1 \to \mathcal{H}_1$. Actually, physics and the celebrated *Wigner's theorem* in particular, requires that every $U_R^{(1)}$ be a *unitary* (or *anti unitary*) transformation but this is not relevant for our case. The natural question concerning the representation of the action of $\mathcal{G}$ on the compositor system S is:
"If we know the representations $\mathcal{G} \ni R \mapsto U_R^{(1)}$ and $\mathcal{G} \ni R \mapsto U_R^{(2)}$, what about the representation of the action of $\mathcal{G}$ on S in terms of linear transformations in the Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$?"
The answer given by physics, at least when the systems S_1 and S_2 do not interact, is that $U_R := U_R^{(2)} \otimes U_R^{(1)}$.

4.3 Permutation group and symmetry of tensors

We remind the definition of the *group of permutations of n objects* and give some know results of basic group theory whose proofs may be found in any group-theory textbook.

Definition 4.14. **(Group of permutations.)** Consider the set $I_n := \{1, \ldots, n\}$, the **group of permutations of n objects**, $\mathscr{P}_n$ is the set of the bijective maps $\sigma : I_n \to I_n$ equipped with the composition rule given by the usual composition rule of functions. Moreover,
(a) the elements of $\mathscr{P}_n$ are called **permutations** (of n objects);
(b) a permutation of $\mathscr{P}_n$ with $n \geq 2$ is said to be a **transposition** if differs from the identity map and reduces to the identity map when restricted to some subset of I_n containing $n-2$ elements. ■

Remarks 4.15.
(1) $\mathscr{P}_n$ contains $n!$ elements.
(2) Each permutation $\sigma \in \mathscr{P}_n$ can be represented by a corresponding string $(\sigma(1), \ldots, \sigma(n))$.
(3) If, for instance $n = 5$, with the notation above $(1, 2, 3, 5, 4)$, $(5, 2, 3, 4, 1)$, $(1, 2, 4, 3, 5)$ are transpositions, $(2, 3, 4, 5, 1)$, $(5, 4, 3, 2, 1)$ are not.
(4) It is possible to show that each permutation $\sigma \in \mathscr{P}_n$ can be decomposed as a product of transpositions $\sigma = \tau_1 \circ \ldots \circ \tau_k$. In general there are several different transposition-product decompositions for each permutation, however it is possible to show that if $\sigma = \tau_1 \circ \ldots \circ \tau_k = \tau'_1 \circ \ldots \circ \tau'_r$, where τ_i and τ'_j are transpositions, then $r + k$ is even. Equivalently, r is even or odd if and only if k is such. This defines the **parity**, $\epsilon_\sigma \in \{-1, +1\}$, of a permutation σ, where, $\epsilon_\sigma = +1$ if σ can be decomposed as a product of an *even* number of transpositions and $\epsilon_\sigma = -1$ if σ can be decomposed as a product of an *odd* number of transpositions.
(5) If $A = [A_{ij}]$ is a real or complex $n \times n$ matrix, it is possible to show (by induction) that:

$$\det A = \sum_{\sigma \in \mathscr{P}_n} \epsilon_\sigma A_{1\sigma(1)} \cdots A_{n\sigma(n)} .$$

Alternatively, the identity above may be used to define the determinant of a matrix. ■

We pass to consider the action of $\mathscr{P}_n$ on tensors. Fix $n > 1$, consider the tensor algebra $\mathscr{A}_{\mathbb{K}}(V)$ and single out the tensor space $V^{n\otimes} := V \otimes \ldots \otimes V$ where the factor V appears n times. Then consider the following action of $\mathscr{P}_n$ on $V^{n\times} := V \times \ldots \times V$ where the factor V appears n times. For each $\sigma \in \mathscr{P}_n$ consider the map

$$\hat{\sigma} : V^{n\times} \to V^{n\otimes} \quad \text{such that} \quad \hat{\sigma} : (v_1, \ldots, v_n) \mapsto v_{\sigma^{-1}(1)} \otimes \ldots \otimes v_{\sigma^{-1}(n)} \,.$$

It is quite straightforward to show that $\hat{\sigma}$ is multi linear. Therefore, let

$$\sigma^{\otimes} : V^{n\otimes} \to V^{n\otimes} \,,$$

be the linear map uniquely determined by $\hat{\sigma}$ by means of the universality theorem. By definition, it is completely determined by linearity and the requirement

$$\sigma^{\otimes} : v_1 \otimes \ldots \otimes v_n \mapsto v_{\sigma^{-1}(1)} \otimes \ldots \otimes v_{\sigma^{-1}(n)} \,.$$

Theorem 4.16. *The above-defined linear map*

$$\sigma \mapsto \sigma^{\otimes}$$

with $\sigma \in \mathscr{P}_n$, is a group representation of $\mathscr{P}_n$ on $V^{n\otimes}$.

Proof. First we show that if $\sigma, \sigma' \in \mathscr{P}_n$ then

$$\sigma^{\otimes}(\sigma'^{\otimes}(v_1 \otimes \ldots \otimes v_n)) = (\sigma \circ \sigma')^{\otimes}(v_1 \otimes \ldots \otimes v_n) \,,$$

This follows from the definition: $\sigma^{\otimes}(\sigma'^{\otimes}(v_1 \otimes \ldots \otimes v_n)) = \sigma^{\otimes}(v_{\sigma'^{-1}(1)} \otimes \ldots \otimes v_{\sigma'^{-1}(n)})$. Re-defining $u_i := v_{\sigma'^{-1}(i)}$ so that $u_{\sigma^{-1}(j)} := v_{\sigma'^{-1}(\sigma^{-1}(j))}$, one finds the identity $\sigma^{\otimes}(\sigma'^{\otimes}(v_1 \otimes \ldots \otimes v_n)) = u_{\sigma^{-1}(1)} \otimes \ldots \otimes u_{\sigma^{-1}(n)} = v_{\sigma'^{-1}\circ\sigma^{-1}(1)} \otimes \ldots \otimes v_{\sigma'^{-1}\circ\sigma^{-1}(n)} = v_{(\sigma\circ\sigma')^{-1}(1)} \otimes \ldots \otimes v_{(\sigma\circ\sigma')^{-1}(n)} = (\sigma \circ \sigma')^{\otimes}(v_1 \otimes \ldots \otimes v_n)$. In other words

$$\sigma^{\otimes}(\sigma'^{\otimes}(v_1 \otimes \ldots \otimes v_n)) = (\sigma \circ \sigma')^{\otimes}(v_1 \otimes \ldots \otimes v_n) \,.$$

In particular, that identity holds also for a canonical basis of elements $e_{i_1} \otimes \ldots \otimes e_{i_n}$

$$\sigma^{\otimes}(\sigma'^{\otimes}(e_{i_1} \otimes \ldots \otimes e_{i_n})) = (\sigma \circ \sigma')^{\otimes}(e_{i_1} \otimes \ldots \otimes e_{i_n}) \,.$$

By linearity such an identity will hold true for all arguments in $V^{n\otimes}$ and thus we have

$$\sigma^{\otimes}\sigma'^{\otimes} = (\sigma \circ \sigma')^{\otimes} \,.$$

The maps $\sigma^{\otimes}$ are linear by constructions and are bijective because

$$\sigma^{\otimes}\sigma'^{\otimes} = (\sigma \circ \sigma')^{\otimes}$$

implies

$$\sigma^{\otimes}\sigma^{-1\otimes} = \sigma^{-1\otimes}\sigma^{\otimes} = e^{\otimes} = I\,.$$

The identity $e^{\otimes} = I$ can be proved by noticing that $e^{\otimes} - I$ is linear and vanishes when evaluated on any canonical base of $V^{n\otimes}$. We have shown that $\sigma^{\otimes} \in GL(V^{n\otimes})$ for all $\sigma \in \mathscr{P}_n$ and the map $\sigma \mapsto \sigma^{\otimes}$ is a homomorphism. This concludes the proof. $\Box$

Let us pass to consider the abstract index notation and give a representation of the action of $\sigma^{\otimes}$ within that picture.

Theorem 4.17. *If t is a tensor in $V^{n\otimes} \in \mathscr{A}_{\mathbb{K}}(V)$ with $n \geq 2$ and $\sigma \in \mathscr{P}_n$, then the components of t with respect to any canonical basis of $V^{n\otimes}$ satisfy*

$$(\sigma^{\otimes}t)^{i_1\ldots i_n} = t^{i_{\sigma(1)}\cdots i_{\sigma(n)}}\,.$$

An analogous statement holds for the tensors $s \in V^{*n\otimes}$:

$$(\sigma^{\otimes}s)_{i_1\ldots i_n} = s_{i_{\sigma(1)}\ldots i_{\sigma(n)}}\,.$$

Proof.

$$\sigma^{\otimes}t = \sigma^{\otimes}(t^{j_1\cdots j_n}e_{j_1}\otimes\ldots\otimes e_{j_n}) = t^{j_1\cdots j_n}e_{j_{\sigma^{-1}(1)}}\otimes\ldots\otimes e_{j_{\sigma^{-1}(n)}}\,.$$

Since $\sigma : I_n \to I_n$ is bijective, if we define $i_k := j_{\sigma^{-1}(k)}$, it holds $j_k = i_{\sigma(k)}$. Using this identity above we find

$$\sigma^{\otimes}t = t^{i_{\sigma(1)}\cdots i_{\sigma(n)}}e_{i_1}\otimes\ldots\otimes e_{i_n}\,.$$

That is nothing but the thesis. The final statement is an obvious consequence of the initial one just replacing the vector space V with the vector space V^* (the fact that V^* is a dual space is immaterial, the statement relies upon the only structure of vector space.) $\Box$

To conclude we introduce the notion of symmetric or anti-symmetric tensor.

Definition 4.18. **(Symmetric and antisymmetric tensors.)** Let V be a finite-dimensional vector space with field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$. Consider the space $V^{n\otimes} \in \mathscr{A}_{\mathbb{K}}(V)$ of tensors of order $(n,0)$ with $n \geq 2$.
(a) $t \in V^{n\otimes}$ is said to be **symmetric** if

$$\sigma^{\otimes}t = t\,,$$

for all of $\sigma \in \mathscr{P}_n$, or equivalently, using the abstract index notation,

$$t^{j_1\cdots j_n} = t^{j_{\sigma(1)}\cdots j_{\sigma(n)}}\,,$$

for all of $\sigma \in \mathscr{P}_n$.
(b) $t \in V^{n\otimes}$ is said to be **anti symmetric** if

$$\sigma^{\otimes}t = \epsilon_\sigma\, t\,,$$

for all of $\sigma \in \mathscr{P}_n$, or equivalently using the abstract index notation:

$$t^{j_1 \cdots j_n} = \epsilon_\sigma \, t^{j_{\sigma(1)} \cdots j_{\sigma(n)}} ,$$

for all of $\sigma \in \mathscr{P}_n$. ■

Remarks 4.19.
(1) Concerning the definition in **(b)** notice that $\epsilon_\sigma = \epsilon_{\sigma^{-1}}$.
(2) Notice that the sets of symmetric and antisymmetric tensors in $V^{n\otimes}$ are, separately, closed by linear combinations. In other words each of them is a subspaces of $V^{n\otimes}$. ■

Examples 4.20.
1. Suppose $n = 2$, then a symmetric tensor $s \in V \otimes V$ satisfies $s^{ij} = s^{ji}$ and an antisymmetric tensor $a \in V \otimes V$ satisfies $a^{ij} = -a^{ji}$.
2. Suppose $n = 3$, then it is trivially shown that $\sigma \in \mathscr{P}_3$ has parity 1 if and only if σ is a **cyclic permutation**, i.e., $(\sigma(1), \sigma(2), \sigma(3)) = (1, 2, 3)$ or $(\sigma(1), \sigma(2), \sigma(3)) = (2, 3, 1)$ or $(\sigma(1), \sigma(2), \sigma(3)) = (3, 1, 2)$.
Now consider the vector space V with $dimV = 3$. It turns out that a tensor $e \in V \otimes V \otimes V$ is anti symmetric if and only if

$$e^{ijk} = 0 ,$$

if (i, j, k) is *not* a permutation of $(1, 2, 3)$ and, otherwise,

$$e^{ijk} = \pm e^{123} ,$$

where the sign $+$ takes place if the permutation $(\sigma(1), \sigma(2), \sigma(3)) = (i, j, k)$ is cyclic and the sign $-$ takes place otherwise. That relation between parity of a permutation and cyclicity does *not* hold true for $n > 3$. ■

Remarks 4.21.
(1) Consider a generic tensor space $S \in \mathscr{A}_{\mathbb{K}}(V)$ which contains $n \geq 2$ spaces V as factors. We may suppose for sake of simplicity $S = S_1 \otimes V^{n\otimes} \otimes S_2$ where $S_1 = U_1 \otimes \ldots \otimes U_k$, $S_2 = U_{k+1} \otimes \ldots \otimes U_m$ and $U_i = V$ or $U_i = V^*$. Anyway all what we are going to say holds true also if the considered n spaces V do not define a unique block $V^{n\otimes}$. We may define the action of $\sigma \in \mathscr{P}_n$ on the whole space S starting by a multi linear map

$$\hat{\sigma} : U_1 \times \ldots \times U_k \times V^{n\times} \times U_{k+1} \times \ldots \times U_m \to U_1 \otimes \ldots \otimes U_k \otimes V^{n\otimes} \otimes U_{k+1} \otimes \ldots \otimes U_m ,$$

such that reduces to the tensor-product map on $U_1 \times \ldots \times U_k$ and $U_{k+1} \times \ldots \times U_m$:

$$\sigma : (u_1, \ldots, u_k, v_1, \ldots v_n, u_{k+1}, \ldots, u_m) \mapsto u_1 \otimes \ldots \otimes u_k \otimes v_{\sigma^{-1}(1)} \otimes \ldots \otimes v_{\sigma^{-1}(n)} \otimes u_{k+1} \otimes \ldots \otimes u_m .$$

Using the universality theorem as above, we build up a representation of $\mathscr{P}_n$ on S, $\sigma \mapsto \sigma^\otimes$ which "acts on $V^{n\otimes}$ only". Using the abstract index notation the action of $\sigma^\otimes$ is well represented:

$$\sigma^\otimes : t^{A \; i_1 \ldots i_n \; B} \mapsto t^{A \; i_{\sigma(1)} \cdots i_{\sigma(n)} \; B} .$$

This allows one to define and study the symmetry of a tensor referring to a few indices singled out among the complete set of indices of the tensors. E.g., a tensor $t^{ij}{}_k{}^r$ may be symmetric or anti symmetric, for instance, with respect to the indices i and j or j and r or ijr.
(2) Notice that no discussion on the symmetry of indices of different kind (one covariant and the other contravariant) is possible. ■

Exercises 4.22.

1. Let t be a tensor in $V^{n\otimes}$ (or $V^{*n\otimes}$). Show that t is symmetric or anti symmetric if there is a canonical basis where the components have symmetric or anti symmetric indices, i.e., $t^{i_1\cdots i_n} = t^{i_{\sigma(1)}\cdots i_{\sigma(n)}}$ or respectively $t^{i_1\cdots i_n} = \epsilon_\sigma\, t^{i_{\sigma(1)}\cdots i_{\sigma(n)}}$ for all $\sigma \in \mathscr{P}_n$.

Remarks 4.23. The result implies that, to show that a tensor is symmetric or anti symmetric, it is sufficient to verify the symmetry or anti symmetry of its components within a *single* canonical basis. ■

2. Show that the sets of symmetric tensors of order $(n,0)$ and $(0,n)$ are vector subspaces of $V^{n\otimes}$ and $V^{*n\otimes}$ respectively.

3. Show that the subspace of anti-symmetric tensors of order $(0,n)$ in $V^{*n\otimes}$ has dimension $\binom{dimV}{n}$ if $n \leq dimV$. What about $n > dimV$?

4. Consider a tensor $t^{i_1\cdots i_n}$, show that the tensor is symmetric if and only if it is symmetric with respect to each arbitrary chosen pair of indices, i.e.

$$t^{\cdots i_k \cdots i_p \cdots} = t^{\cdots i_p \cdots i_k \cdots}\,,$$

for all $p, k \in \{1, \ldots n\}$, $p \neq k$.

5. Consider a tensor $t^{i_1\cdots i_n}$, show that the tensor is anti symmetric if and only if it is anti symmetric with respect to each arbitrarily chosen pair of indices, i.e.

$$t^{\cdots i_k \cdots i_p \cdots} = -t^{\cdots i_p \cdots i_k \cdots}\,,$$

for all $p, k \in \{1, \ldots n\}$, $p \neq k$.

6. Show that $V \otimes V = A \oplus S$ where $\oplus$ denotes the direct sum and A and S are respectively the space of anti-symmetric and symmetric tensors in $V \otimes V$. Does such a direct decomposition hold if considering $V^{n\otimes}$ with $n > 2$?

4.4 Grassmann algebra, also known as Exterior algebra (and something about the space of symmetric tensors)

n-forms play a crucial role in several physical applications of tensor calculus. Let us review, very quickly, the most important features of n-forms.

Definition 4.24. **(Spaces $\Lambda^p(V)$ and $\Lambda^p(V^*)$.)** If V denotes a linear space with field $\mathbb{K}$ and finite dimension n, $\Lambda^p(V)$ and $\Lambda^p(V^*)$ indicates respectively the linear space of the *antisymmetric* $(p,0)$-order tensors and *antisymmetric* $(0,p)$-order tensors. We assume the convention that

$\Lambda^0(V) := \mathbb{K}$ and $\Lambda^0(V^*) := \mathbb{K}$ and furthermore, if $p > n$ is integer, $\Lambda^p(V) := \{0\}$, $\Lambda^p(V^*) := \{0\}$, $\{0\}$ being the trivial vector space made of the zero vector only. The elements of $\Lambda^p(V)$ are called p**-vectors**. The elements of $\Lambda^p(V^*)$ are called p**-forms**. $\blacksquare$

In the following we will be mainly concerned with the spaces $\Lambda^p(V)$ and $\Lambda^p(V^*)$, however, for some applications to physics (especially quantum mechanics), it is useful to define the spaces $S^p(V)$ and $S^p(V^*)$, respectively, the linear space of the *symmetric* $(p,0)$-order tensors and *symmetric* $(0,p)$-order tensors. As before, we assume $S^0(V) := \mathbb{K}$, $S^0(V^*) := \mathbb{K}$.

Remarks 4.25. $\Lambda^p(V) = \{0\}$ if $p > n$ is a natural definition. Indeed in the considered case, if $s \in \Lambda^p(V)$, in every component $s^{i_1 \cdots i_n}$ there must be at least two indices, say i_1 and i_2, with the same value so that $s^{i_1 i_2 \ldots i_n} = s^{i_2 i_1 \ldots i_n}$. Theorem 4.17 implies: $s^{i_1 i_2 \ldots i_n} = -s^{i_2 i_1 \ldots i_n}$, and thus $s^{i_1 i_2 \ldots i_n} = s^{i_2 i_1 \ldots i_n} = 0$. $\blacksquare$

Before going on, we remind some basic notions about *projectors*.
If U is a vector space with field $\mathbb{K} = \mathbb{R}$ of $\mathbb{C}$, a **projector** P **onto** the subspace $S \subset U$ is a linear operator $P : U \to U$ such that it is **idempotent**, i.e. $PP = P$ and $P(U) = S$.
We have the following well-known results concerning projectors.
(1) If P is a projector onto S, $Q := I - P$ is a projector onto another subspace $S' := Q(U)$. Moreover it holds: $U = S \oplus S'$, where $\oplus$ denotes the direct sum of S and S' (i.e. $S \oplus S'$ is the space generated by linear combinations of elements in S and elements in S' and it holds $S \cap S' = \{0\}$).
(2) Conversely, if S and S' are subspaces of U such that $S \oplus S' = U$ – and thus, for every $v \in U$ there are a pair of elements $v_S \in S$ and $v_{S'} \in S'$, uniquely determined by the decomposition $S \oplus S' = U$, such that $v = v_S + v_{S'}$ – the applications $P : U \ni v \mapsto v_S$ and $Q : U \ni v \mapsto v_{S'}$ are projectors onto S and S' respectively and $Q = I - P$.
(3) If (a) P is a projector onto S, (b) P' is a projector onto S' and (c) $PP' = P'P$, then PP' is a projector as well and it is a projector onto $S \cap S'$.
These results are valid also in the infinite dimensional case. However in that case, if U is a Banach space it is more convenient specializing the definition of projectors requiring their continuity too. With this restricted definition the statements above are valid anyway provided that the involved subspaces are closed (i.e. they are sub Banach spaces).

Proposition 4.26. *Consider the linear operators* $\mathcal{A} : V^{\otimes p} \to V^{\otimes p}$ *and* $\mathcal{S} : V^{\otimes p} \to V^{\otimes p}$ *where, respectively,*

$$\mathcal{A} := \frac{1}{p!} \sum_{\sigma \in \mathscr{P}_p} \epsilon_\sigma \sigma^{\otimes} , \tag{4.4}$$

and

$$\mathcal{S} := \frac{1}{p!} \sum_{\sigma \in \mathscr{P}_p} \sigma^{\otimes} . \tag{4.5}$$

$\mathcal{A}$ *is a projector onto* $\Lambda^p(V)$*, i.e.* $\mathcal{A}\mathcal{A} = \mathcal{A}$ *and* $\mathcal{A}\left(V^{\otimes p}\right) = \Lambda^p(V)$ *and, similarly,* $\mathcal{S}$ *is a projector* *onto $S^p(V)$, i.e. $\mathcal{S}\mathcal{S} = \mathcal{S}$ and $\mathcal{S}(V^{\otimes p}) = S^p(V)$. Finally one has*

$$\mathcal{A}\mathcal{S} = \mathcal{S}\mathcal{A} = 0 \tag{4.6}$$

where 0 is the zero-operator, which implies $\Lambda^p(V) \cap S^p(V) = \{0\}$.

Proof. Let us prove that $\mathcal{A}\mathcal{A} = \mathcal{A}$. (Actually it would be a trivial consequence of the two facts we prove immediately after this point, but this proof is interesting in its onw right because similar procedures will be used later in more complicated proofs.) Using theorem 4.16 and the fact that $\epsilon_\sigma \epsilon_\tau = \epsilon_{\sigma\circ\tau}$ for $\sigma, \tau \in \mathscr{P}_p$, one has

$$\mathcal{A}\mathcal{A} = \frac{1}{p!}\sum_{\sigma\in\mathscr{P}_p} \epsilon_\sigma \sigma^{\otimes} \frac{1}{p!}\sum_{\tau\in\mathscr{P}_p} \epsilon_\tau \tau^{\otimes} = \frac{1}{p!}\sum_{\sigma\in\mathscr{P}_p} \frac{1}{p!}\sum_{\tau\in\mathscr{P}_p} \epsilon_{\sigma\circ\tau}(\sigma\circ\tau)^{\otimes} = \frac{1}{p!}\sum_{\sigma\in\mathscr{P}_p} \frac{1}{p!}\sum_{\sigma\circ\tau\in\mathscr{P}_p} \epsilon_{\sigma\circ\tau}(\sigma\circ\tau)^{\otimes}.$$

In the last step, we have taken advantage of the fact that, for fixed σ, $\sigma\circ\tau$ ranges in the whole $\mathscr{P}_p$ if τ ranges in the whole $\mathscr{P}_p$. Therefore summing over τ is equivalent to summing over $\sigma\circ\tau$. In other words, if $\sigma' := \sigma\circ\tau$,

$$\mathcal{A}\mathcal{A} = \frac{1}{p!}\sum_{\sigma\in\mathscr{P}_p} \frac{1}{p!}\sum_{\sigma'\in\mathscr{P}_p} \epsilon_{\sigma'}\, \sigma'^{\otimes} = \frac{1}{p!} p!\, \mathcal{A} = \mathcal{A}\,.$$

To prove $\mathcal{A}(V^{\otimes p}) = \Lambda^p(V)$ it is sufficient to verify that, if $s \in V^{\otimes p}$, then $\sigma^{\otimes}\mathcal{A}s = \epsilon_\sigma \mathcal{A}s$ so that $\mathcal{A}(V^{\otimes p}) \subset \Lambda^p(V)$ and furthermore that $t \in \Lambda^p(V)$ entails $\mathcal{A}t = t$, so that $\mathcal{A}(V^{\otimes p}) \supset \Lambda^p(V)$. Using, in particular, the fact that $\epsilon_\sigma \epsilon_{\tau\circ\sigma} = \epsilon_\sigma\epsilon_\tau\epsilon_\sigma = (\epsilon_\sigma)^2\epsilon_\tau$, we have:

$$\sigma^{\otimes}\mathcal{A}s = \frac{1}{p!}\sum_{\tau\in\mathscr{P}_p} \epsilon_\tau\, \sigma^{\otimes}(\tau^{\otimes}s) = \frac{1}{p!}\sum_{\tau\in\mathscr{P}_p} \epsilon_\sigma\epsilon_{\tau\circ\sigma}(\sigma\circ\tau)^{\otimes}s = \epsilon_\sigma \frac{1}{p!}\sum_{\sigma\circ\tau\in\mathscr{P}_p} \epsilon_{\tau\circ\sigma}(\sigma\circ\tau)^{\otimes}s = \epsilon_\sigma\mathcal{A}s\,.$$

To conclude we prove that $t \in \Lambda^p(V)$ entails $\mathcal{A}t = t$.

$$\mathcal{A}t = \frac{1}{p!}\sum_{\sigma\in\mathscr{P}_p} \epsilon_\sigma\sigma^{\otimes}t = \frac{1}{p!}\sum_{\sigma\in\mathscr{P}_p} (\epsilon_\sigma)^2 t = \frac{1}{p!}\sum_{\sigma\in\mathscr{P}_p} t = \frac{p!}{p!}t = t\,.$$

We have used the fact that $\sigma^{\otimes}t = \epsilon_\sigma t$ since $t \in \Lambda^p(V)$.
The proof for $\mathcal{S}$ and the spaces $S^p(V)$ is essentially identical with the only difference that the coefficients ϵ_τ, ϵ_σ etc. do not take place anywhere in the proof. The last statement can be proved as follows. First notice that, if τ is a fixed transposition

$$-\sum_{\sigma\in\mathscr{P}_p}\epsilon_\sigma = \sum_{\sigma\in\mathscr{P}_p}\epsilon_\tau\epsilon_\sigma = \sum_{\sigma\in\mathscr{P}_p}\epsilon_{\tau\circ\sigma} = \sum_{\tau\circ\sigma\in\mathscr{P}_p}\epsilon_{\tau\circ\sigma} = \sum_{\sigma'\in\mathscr{P}_p}\epsilon_{\sigma'} = \sum_{\sigma\in\mathscr{P}_p}\epsilon_\sigma\,,$$

so that $\sum_{\sigma\in\mathscr{P}_p}\epsilon_\sigma = 0$. Next, notice that if $t \in V^{\otimes p}$, $\mathcal{S}t = s \in S^p(V)$ is symmetric and thus:

$$p!\mathcal{A}\mathcal{S}t = p!\mathcal{A}s = \sum_{\sigma\in\mathscr{P}_p}\epsilon_\sigma\sigma^{\otimes}s = \left(\sum_{\sigma\in\mathscr{P}_p}\epsilon_\sigma\right)s = 0$$

since $\sum_{\sigma \in \mathscr{P}_p} \epsilon_\sigma = 0$. If $t \in V^{\otimes p}$, $\mathscr{A}t = a \in \Lambda^p(V)$ is antisymmetric and thus:

$$p!\mathcal{S}\mathcal{A}t = p!\mathcal{S}a = \sum_{\sigma \in \mathscr{P}_p} \sigma^{\otimes} a = \left(\sum_{\sigma \in \mathscr{P}_p} \epsilon_\sigma \right) a = 0 \,,$$

again, since $\sum_{\sigma \in \mathscr{P}_p} \epsilon_\sigma = 0$. Using the property (3) of projector, mentioned above, it results that $\Lambda^p(V) \cap S^p(V) = \{0\}$. □

Remarks 4.27. In view of the last statement, one may wonder if the space $V^{\otimes}$ is the direct sum of $\Lambda^p(V)$ and $S^p(V)$. This is not the case if $p > 2$ and $\Lambda^p(V) \oplus S^p(V)$ is a proper subspace of $V^{\otimes p}$ for $p > 2$. The proof is left to the reader as an execise. ■

Examples 4.28. This example is very important in Quantum Mechanics. Consider a set of p identical quantum systems S, each represented on the Hilbert space $\mathcal{H}$. As we know, in principle, the states of the overall system are represented by unitary-norm vectors in the Hilbert tensor product of p copies of the space $\mathcal{H}$, $\mathcal{H} \otimes_H \cdots \otimes_H \mathcal{H}$ (see definition 2.31). This is not the whole story, because the axioms of Quantum Mechanics (but the following statement turns out to be a theorem in relativistic quantum field theory called the *spin-statistics theorem*) assume that only some vectors are permitted to describe states in the case the system S is a quantum particle. There are only two categories of quantum particles: *bosons* and *fermions*. The first ones are those with integer spin (or helicity) and the others are those with semi-integer spin (or helicity). The states of a system of p identical bonsons are allowed to be represented by unitary-norm vectors in $S^p(\mathcal{H})$ only, whereas the states of a system of p identical bonsons are allowed to be represented by unitary-norm vectors in $\Lambda^p(\mathcal{H})$. All the possible transformations of states by means of operators, representing some physical action on the system, must respect that requirement.

4.4.1 The exterior product and Grassmann algebra

We are in a position to give the central definition, that of *exterior product* of k-vectors.

Definition 4.29. **(Exterior product.)** *If $s \in \Lambda^p(V)$ and $t \in \Lambda^q(V)$ the* **exterior product** *of s and t is the $(p+q)$-vector (the 0-vector when $p + q > dimV$):*

$$s \wedge t := \mathcal{A}(s \otimes t) \,.$$

Theorem 4.30. *Consider the exterior product $\wedge$ in the spaces $\Lambda^r(V)$ where V is a linear space with field $\mathbb{K}$ and finite dimension n. It fulfills the following properties for $s \in \Lambda^p(V)$, $t \in \Lambda^q(V)$, $w \in \Lambda^r(V)$ and $a, b \in \mathbb{K}$:*
(a) *$s \wedge (t \wedge w) = (s \wedge t) \wedge w = \mathcal{A}(s \otimes t \otimes w)$;*
(b) *$s \wedge t = (-1)^{pq} t \wedge s$, in particular $u \wedge v = -v \wedge u$ if $u, v \in V$;*

(c) *$s \wedge (at + bw) = a(s \wedge t) + b(s \wedge w)$.*
Furthermore, if $u_1, \ldots, u_m \in V$, these vectors are linearly independent if and only if

$$u_1 \wedge \cdots \wedge u_m \neq 0\,,$$

where the exterior product of several vectors is well-defined (without parentheses) in accordance with the associativity property (a).

Proof. (a) Let us prove that $s\wedge(t\wedge w) = \mathcal{A}(s\otimes t\otimes w)$. Exploiting a similar procedure one proves that $(s\wedge t)\wedge w = \mathcal{A}(s\otimes t\otimes w)$ and it concludes the proof. To prove that $s\wedge(t\wedge w) = \mathcal{A}(s\otimes t\otimes w)$, we notice that, if $s \in \Lambda^p(V)$, $t \in \Lambda^q(V)$, $w \in \Lambda^r(V)$ then:

$$s \wedge (t \wedge w) = \frac{1}{(p+q+r)!} \sum_{\sigma \in \mathscr{P}_{p+q+r}} \epsilon_\sigma \sigma^{\otimes} \left(\frac{1}{(q+r)!} \sum_{\tau \in \mathscr{P}_{q+r}} \epsilon_\tau \, s \otimes \tau^{\otimes}(t \otimes w) \right) .$$

That is

$$s \wedge (t \wedge w) = \frac{1}{(q+r)!} \sum_{\tau \in \mathscr{P}_{q+r}} \left(\frac{1}{(p+q+r)!} \sum_{\sigma \in \mathscr{P}_{p+q+r}} \epsilon_\tau \epsilon_\sigma \, \sigma^{\otimes}(s \otimes \tau^{\otimes}(t \otimes w)) \right) .$$

Now we can view $\mathscr{P}_{q+r}$ as the subgroup of $\mathscr{P}_{p+q+r}$ which leaves unchanged the first p objects. With this interpretation one has

$$s \wedge (t \wedge w) = \frac{1}{(q+r)!} \sum_{\tau \in \mathscr{P}_{q+r}} \left(\frac{1}{(p+q+r)!} \sum_{\sigma \in \mathscr{P}_{p+q+r}} \epsilon_\tau \epsilon_\sigma \sigma^{\otimes}(\tau^{\otimes}(s \otimes t \otimes w)) \right) .$$

Equivalently,

$$\begin{aligned}
s \wedge (t \wedge w) &= \frac{1}{(q+r)!} \sum_{\tau \in \mathscr{P}_{q+r}} \left(\frac{1}{(p+q+r)!} \sum_{\sigma \in \mathscr{P}_{p+q+r}} \epsilon_{\sigma\circ\tau} \, (\sigma \circ \tau)^{\otimes}(s \otimes t \otimes w) \right) \\
&= \frac{1}{(q+r)!} \sum_{\tau \in \mathscr{P}_{q+r}} \left(\frac{1}{(p+q+r)!} \sum_{\sigma\circ\tau \in \mathscr{P}_{p+q+r}} \epsilon_{\sigma\circ\tau} \, (\sigma \circ \tau)^{\otimes}(s \otimes t \otimes w) \right) \\
&= \frac{1}{(q+r)!} \sum_{\tau \in \mathscr{P}_{q+r}} \left(\frac{1}{(p+q+r)!} \sum_{\sigma' \in \mathscr{P}_{p+q+r}} \epsilon_{\sigma'} \, \sigma'^{\otimes}(s \otimes t \otimes w) \right) \\
&= \frac{(q+r)!}{(q+r)!} \mathcal{A}(s \otimes t \otimes w) = \mathcal{A}(s \otimes t \otimes w)\,.
\end{aligned}$$

(b) One may pass from $u_1 \otimes \cdots \otimes u_q \otimes v_1 \otimes \cdots \otimes v_p$ to $v_1 \otimes \cdots \otimes v_p \otimes u_1 \otimes \cdots \otimes u_q$ by means of the following procedure. First, using $p+q-1$ transpositions one achieves $u_2 \otimes \cdots \otimes u_q \otimes v_1 \otimes \cdots \otimes v_p \otimes u_1$. In this case, due to the definition of $\mathcal{A}$ one has:

$$\mathcal{A}(u_1 \otimes u_2 \otimes \cdots \otimes u_q \otimes v_1 \otimes \cdots \otimes v_p) = (-1)\mathcal{A}(u_2 \otimes u_1 \otimes \cdots \otimes u_q \otimes v_1 \otimes \cdots \otimes v_p) ,$$

so that, using $p + q - 1$ iterated transpositions:

$$\mathcal{A}(u_1 \otimes \cdots \otimes u_q \otimes v_1 \otimes \cdots \otimes v_p) = (-1)^{p+q-1}\mathcal{A}(u_2 \otimes \cdots \otimes u_q \otimes v_1 \otimes \cdots \otimes v_p \otimes u_1) .$$

Using again the same procedure one gets also

$$\mathcal{A}(u_1 \otimes \cdots \otimes u_q \otimes v_1 \otimes \cdots \otimes v_p) = (-1)^{p+q-1}(-1)^{p+q-1}\mathcal{A}(u_3 \otimes \cdots \otimes u_q \otimes v_1 \otimes \cdots \otimes v_p \otimes u_1 \otimes u_2) .$$

By iteration of this prescription one finally gets:

$$\mathcal{A}(u_1 \otimes \cdots \otimes u_q \otimes v_1 \otimes \cdots \otimes v_p) = (-1)^{q(p+q-1)}\mathcal{A}(v_1 \otimes \cdots \otimes v_p \otimes u_1 \otimes \cdots \otimes u_q) .$$

That is

$$u_1 \wedge \cdots \wedge u_q \wedge v_1 \wedge \cdots \wedge v_p = (-1)^{q(q-1)}(-1)^{pq} v_1 \wedge \cdots \wedge v_p \wedge u_1 \wedge \cdots \wedge u_q .$$

Since $(-1)^{q(q-1)} = 1$, by linearity the obtained identity gives rise to the thesis.
(c) is a straightforward consequence of the definition of $\mathcal{A}$.
Let us come to the last statement. If $u_1, \ldots u_m$ are linearly dependent, there must be one of them, say u_1, which can be represented as a linear combination of the others. In this case $u_1 \wedge \cdots \wedge u_m$ can be re-written as: $\sum_{i=2}^{m} c^i u_i \wedge \cdots \wedge u_m$. Since, whenever $c_i \neq 0$, u_i appears twice in $u_i \wedge \cdots \wedge u_m = \mathcal{A}(u_i \otimes \cdots \otimes u_m)$, this term vanishes (interchanging the two copies of u^i on a hand amounts to do nothing on the other hand it amounts to a transposition which produce a sign minus) and thus $u_1 \wedge \cdots \wedge u_m = 0$. If $u_1, \ldots, u_m$ are linearly independent, one can add further vectors $u_{m+1}, \ldots u_n$ in order to obtain a basis for V, which we denote by $\{e_i\}_{i=1,\ldots m}$. In this case $u_1 \wedge \cdots \wedge u_n = \mathcal{A}(e_1 \otimes \cdots \otimes e_n) \neq 0$. Indeed, by direct inspection one finds that if $\{e^{*i}\}_{i=1,\ldots,n}$ is the dual base

$$\langle e^{*1} \otimes \cdots \otimes e^{*n}, \mathcal{A}(e_1 \otimes \cdots \otimes e_n)\rangle = \frac{1}{n!}\langle e^{*1} \otimes \cdots \otimes e^{*n}, e_1 \otimes \cdots \otimes e_n\rangle + \text{vanishing terms} = \frac{1}{n!} \neq 0$$

and so $u_1 \wedge \cdots \wedge u_n \neq 0$. As a consequence the factor $u_1 \wedge \cdots \wedge u_m$ cannot vanish as well: If not it would imply, making use of (c) in the last passage:

$$u_1 \wedge \cdots \wedge u_n = (u_1 \wedge \cdots \wedge u_m) \wedge (u_{m+1} \wedge \cdots \wedge u_n) = \mathbf{0} \wedge (u_{m+1} \wedge \cdots \wedge u_n) = 0(\mathbf{0} \wedge (u_{m+1} \wedge \cdots \wedge u_n)) = \mathbf{0}.$$

$\square$

Remarks 4.31. It is very importat to stress that there is another *inequivalent* definition of wedge product which gives rise to a naturally isomorphic structure. If $s \in \Lambda^p(V)$ and $t \in \Lambda^q(V)$ the **exterior product** of s and t is the $(p+q)$-vector (the 0-vector when $p+q > dimV$):

$$s \widetilde{\wedge} t := \frac{(p+q)!}{p!q!} \mathcal{A}(s \otimes t) = \frac{(p+q)!}{p!q!} s \wedge t \,. \tag{4.7}$$

With this different definition, *all* the statements of theorem 4.30 and the next corollay are however *true*! In particular $\widetilde{\wedge}$ is associative. The new structure is as natural as the one we have introduced and is constructed by taking suitable quotients in the tensor product space instead of exploiting the operator $\mathcal{A}$. In symplectic geometry and Hamiltonian mechanics the second definition of wedge product is preferred. This ambiguity is often source of mathematical disasters. ■

Theorem 4.30 has an immediate and important corollary:

Corollary. *Each space $\Lambda^p(V)$, with $0 \leq p \leq n = dim(V)$ has dimension $\binom{n}{p}$. Moreover, if $\{e_i\}_{i=1,\ldots,n}$ is a basis of V, the elements*

$$e_{i_1} \wedge \cdots \wedge e_{i_p}\,, \quad \textit{with } 1 \leq i_1 < i_2 < \cdots < i_p \leq n$$

form a basis of $\Lambda^p(V)$.

Proof. Define $I := \{1, 2, \ldots, n\}$. Notice that, with the given definitions, $e_{i_1} \wedge \cdots \wedge e_{i_p} = \mathcal{A}(e_{i_1} \otimes \cdots \otimes e_{i_p})$.
Since $e_{i_1} \otimes \cdots \otimes e_{i_p}$ with $i_1, \ldots, i_p \in I$ define a basis of $V^{n\otimes}$ and $\mathcal{A} : V^{n\otimes} \to \Lambda^p(V)$ is a projector, the elements $e_{i_1} \wedge \cdots \wedge e_{i_p} = \mathcal{A}(e_{i_1} \otimes \cdots \otimes e_{i_p})$ are generators of $\Lambda^p(V)$. Moreover $e_{i_1} \wedge \cdots \wedge e_{i_p} = \pm e_{j_1} \wedge \cdots \wedge e_{j_p}$ if $(i_1, \ldots, i_p)$ and $(j_1, \ldots, j_p)$ differ to each other for a permutation and $e_{i_1} \wedge \cdots \wedge e_{i_p} = 0$ if there are repeated indices. All that implies that

$$S := \{e_{i_1} \wedge \cdots \wedge e_{i_p}\,, \quad \text{with } 1 \leq i_1 < i_2 < \cdots < i_p \leq n\}$$

is still a set of generators of $\Lambda^p(V)$. Finally, the elements of S are linearly independent. Indeed, suppose that

$$c^{i_1 \cdots i_p} e_{i_1} \wedge \cdots \wedge e_{i_p} = 0\,, \quad \text{with } 1 \leq i_1 < i_2 < \cdots < i_p \leq n\,.$$

If we apply the functional $e_{i_1} \wedge \cdots \wedge e_{i_p}$ to $(e^{*j_1}, \ldots, e^{*j_n})$ with $1 \leq j_1 < j_2 < \cdots < j_p \leq n$ fixed, only the term $e_{j_1} \wedge \cdots \wedge e_{j_p}$ gives a non-vanishing contribution:

$$0 = c^{i_1 \cdots i_p} e_{i_1} \wedge \cdots \wedge e_{i_p}(e^{*j_1}, \ldots, e^{*j_p}) = c^{i_1 \cdots i_p} \left\langle \mathcal{A}(e_{i_1} \otimes \cdots \otimes e_{i_p}), e^{*j_1} \otimes \cdots \otimes e^{*j_p} \right\rangle$$

$$= c^{i_1 \cdots i_p} \frac{1}{p!} \delta^{i_1}_{j_1} \cdots \delta^{i_p}_{j_p} = c^{j_1 \cdots j_p} \frac{1}{p!}\,.$$

Thus we conclude that $c^{j_1 \cdots j_p} = 0$ for all $j_1, \ldots, j_p$ with $1 \leq j_1 < j_2 < \cdots < j_p \leq n$. Therefore the elements of S are linearly independent. As a result of combinatorial calculus those elements are exactly $n!/(p!(n-p)!) = \binom{n}{p}$. $\square$

Remarks 4.32.
(1) Let $dimV = n$ and fix a basis $\{e_i\}_{i=1,\ldots,n}$ in V. If $w \in \Lambda^p(V)$, it can be represented in two different ways:

$$w = w^{i_1\cdots i_p} e_{i_1} \otimes \cdots \otimes e_{i_p} \,,$$

or

$$w = \sum_{1\leq i_1 < \cdots < i_p \leq n} \tilde{w}^{i_1\cdots i_p} e_{i_1} \wedge \cdots \wedge e_{i_p} \,.$$

The relation between the components with $1 \leq i_1 < \cdots < i_p \leq n$ is (the reader should prove it):

$$\tilde{w}^{i_1\cdots i_p} = p! w^{i_1\cdots i_p} \,. \tag{4.8}$$

(2) Every $u^{*1} \wedge \cdots \wedge u^{*p} \in \Lambda^p(V^*) \subset V^{*n\otimes}$ acts on $V^{n\otimes}$ and thus, in particular, it acts on every tensor $v_1 \wedge \cdots \wedge v_p \in \Lambda^p(V) \subset V^{n\otimes}$. Let us focus on that action.

$$\langle v_1 \wedge \cdots \wedge v_p, u^{*1} \wedge \cdots \wedge u^{*p}\rangle = \langle \mathcal{A}(v_1 \otimes \cdots \otimes v_p), \mathcal{A}(u^{*1} \otimes \cdots \otimes u^{*p})\rangle$$

$$= \sum_{\tau \in \mathscr{P}_p} \frac{\epsilon_\tau}{p!} \langle \mathcal{A}(v_1 \otimes \cdots \otimes v_p), u^{*\tau^{-1}(1)} \otimes \cdots \otimes u^{*\tau^{-1}(p)}\rangle$$

$$= \sum_{\tau \in \mathscr{P}_p} \frac{(\epsilon_\tau)^2}{p!} \langle \mathcal{A}(v_{\tau^{-1}(1)} \otimes \cdots \otimes v_{\tau^{-1}(p)}), u^{*\tau^{-1}(1)} \otimes \cdots \otimes u^{*\tau^{-1}(p)}\rangle = \frac{p!}{p!} \langle \mathcal{A}(v_1 \otimes \cdots \otimes v_p), u^{*1} \otimes \cdots \otimes u^{*p}\rangle,$$

since, as one can prove by direct inspection:

$$\langle \mathcal{A}(v_{\tau^{-1}(1)} \otimes \cdots \otimes v_{\tau^{-1}(p)}), u^{*\tau^{-1}(1)} \otimes \cdots \otimes u^{*\tau^{-1}(p)}\rangle = \langle \mathcal{A}(v_1 \otimes \cdots \otimes v_p), u^{*1} \otimes \cdots \otimes u^{*p}\rangle \,.$$

To conclude, we have found that:

$$\langle v_1 \wedge \cdots \wedge v_p, u^{*1} \wedge \cdots \wedge u^{*p}\rangle = \frac{1}{p!} \sum_{\sigma \in \mathscr{P}_p} \epsilon_\sigma \langle v_{\sigma^{-1}(1)}, u^{*1}\rangle \cdots \langle v_{\sigma^{-1}(p)}, u^{*p}\rangle \,.$$

Since summing over $\sigma \in \mathscr{P}_p$ is completely equivalent to summing over $\sigma^{-1} \in \mathscr{P}_p$ and the right-hand side is nothing but $(1/p!) \det \left([\langle v_i, u^{*j}\rangle]_{i,j=1,\ldots,p}\right)$, we have eventually proved the following formula:

$$\langle v_1 \wedge \cdots \wedge v_p, u^{*1} \wedge \cdots \wedge u^{*p}\rangle = \frac{1}{p!} \det \left([\langle v_i, u^{*j}\rangle]_{i,j=1,\ldots,p}\right) \,. \tag{4.9}$$

To go on, we remind the reader that, if $V_1, \cdots, V_k$ are vector space on the field $\mathbb{K}$, $\bigoplus_{p=0}^k V_p$ denotes the **external direct sum** i.e. the vector space defined on $V_1 \times \cdots \times V_k$ with composition rule

$$\alpha(w_1, \ldots, w_k) + \beta(u_1, \ldots, u_k) := (\alpha w_1 + \beta u_1, \ldots, \alpha w_k + \beta u_k)$$

for all $\alpha, \beta \in \mathbb{K}$ and $(w_1, \ldots, w_k), (u_1, \ldots, u_k) \in V_1 \times \cdots \times V_k$.

Definition 4.33. **(Grassmann algebra)** Let V be a linear space on the field $\mathbb{K}$ with finite dimension n and let $\wedge$ be the exterior product defined above. The **Grassmann algebra** on V (also called **exterior algebra** on V) is the pair $(\Lambda(V), \wedge)$ where the former is the external direct sum

$$\Lambda(V) = \bigoplus_{p=0}^{n} \Lambda^p(V) \,,$$

and the product $\wedge$ denotes the obvious bi-linear extension of the standard external product to the above direct sum. ■

Let us finally focus on the space of the forms $\Lambda^p(V^*)$ where, as usual, V has finite dimension $n \geq p$. We show that it is nothing but $(\Lambda^p(V))^*$. This is stated in the following proposition.

Proposition 4.34. *If V is a linear space with field $\mathbb{K}$ and finite dimension n, and $0 \leq p \leq n$ is integer, there is a natural isomorphism F_p between $\Lambda^p(V^*)$ and $(\Lambda^p(V))^*$ obtained by the restriction of any functional in $\Lambda^p(V^*)$ to the space $\Lambda^p(V)$. In other words,*

$$F_p : \Lambda^p(V^*) \ni f \mapsto f\restriction_{\Lambda^p(V)} \in (\Lambda^p(V))^* \,.$$

Therefore $\Lambda(V^*)$ is naturally isomorphic to $(\Lambda(V))^*$.

Proof. By construction the map F is linear. Let us prove that it is injective and surjective. Concerning the first issue, it is sufficient that $f\restriction_{\Lambda^p(V)}= 0$ entails $f = 0$. Let us prove it. Fix a basis in V and generate the associated canonical bases in every relevant tensor space. Therefore, using the corollary of theorem 4.30

$$f = c_{i_1 \ldots i_p} e^{*i_1} \wedge \cdots \wedge e^{*i_p} \,,$$

where the sum is extended over the set $1 \leq i_1 < \cdots < i_p \leq n$. If $f\restriction_{\Lambda^p(V)}= 0$, one has, in particular:

$$f(e_{j_1} \wedge \cdots \wedge e_{j_p}) = 0 \,,$$

so that

$$c_{i_1 \ldots i_p} \langle e_{j_1} \wedge \cdots \wedge e_{j_p}, e^{*i_1} \wedge \cdots \wedge e^{*i_p} \rangle = 0 \,.$$

Remembering that $1 \leq i_1 < \cdots < i_p \leq n$ and $1 \leq j_1 < \cdots < j_p \leq n$ and using the definition of $\wedge$ one concludes that $\langle e_{j_1} \wedge \cdots \wedge e_{j_p}, e^{*i_1} \wedge \cdots \wedge e^{*i_p} \rangle = 0$ unless $i_1 = j_1$, $i_2 = j_2$,…, $i_p = j_p$. In this case the formula (4.9) produces $\langle e_{i_1} \wedge \cdots \wedge e_{i_p}, e^{*i_1} \wedge \cdots \wedge e^{*i_p} \rangle = (p!)^{-1} \det I = 1/p!$. We conclude that

$$\langle e_{j_1} \wedge \cdots \wedge e_{j_p}, e^{*i_1} \wedge \cdots \wedge e^{*i_p} \rangle = \frac{1}{p!} \delta^{i_1}_{j_1} \cdots \delta^{i_p}_{j_p} \,, \tag{4.10}$$

and thus

$$c_{j_1 \ldots j_p} = 0 \,.$$

Since $j_1, \ldots, j_n$ are arbitrary, this implies that $f = c_{i_1 \ldots i_p} e^{*i_1} \wedge \cdots \wedge e^{*i_p} = 0$.
Let us pass to the proof of surjectivity. Actually surjectivity follows from injectivity since $\Lambda^p(V^*)$ and $(\Lambda^p(V))^*$ have the same dimension, however we give also a direct proof. By direct inspection one verifies straightforwardly that, if $h \in (\Lambda^p(V))^*$, the functional of $\Lambda^p(V^*)$ (where, as usual, the sum is extended over the set $1 \leq j_1 < \cdots < j_p \leq n$)

$$f_h := p!\, h(e_{j_1} \wedge \cdots \wedge e_{j_p}) e^{*j_1} \wedge \cdots \wedge e^{*j_p}$$

satisfies:

$$f_h\restriction_{\Lambda^p(V)} = h\ .$$

The natural isomorphism from $\Lambda(V^*)$ to $(\Lambda(V))^*$ it is boviously defined as

$$F : \Lambda(V^*) \ni (u_0^*, u_1^*, \ldots, u_n^*) \mapsto (F_0(u_0^*), F_1(u_1^*), \ldots, F_n(u_n^*)) \in (\Lambda(V))^*\ .$$

This concludes the proof. $\square$

4.4.2 Interior product

Consider the exterior algebras $\Lambda(V)$ and $\Lambda(V^*)$ with $dimV = n$ and field $\mathbb{K}$ and the pairing: $\langle\cdot|\cdot\rangle : \Lambda(V) \times \Lambda(V^*) \to \mathbb{K}$, defined by:

$$\langle w|z^*\rangle := \sum_{p=0}^{n} p!\langle w_p, z^{*p}\rangle\ , \quad \text{for all } w = (w_0, \ldots, w_n) \in \Lambda(V) \text{ and } z^* = (z^{*0}, \ldots, z^{*n}) \in \Lambda(V^*). \tag{4.11}$$

By direct inspection one proves that the pairing $\langle\cdot|\cdot\rangle$ is bilinear. Moreover one has the following proposition.

Proposition 4.35. *Consider the exterior algebras $\Lambda(V)$ and $\Lambda(V^*)$ with $dimV = n$ and field $\mathbb{K}$ and the pairing: $\langle\cdot|\cdot\rangle : \Lambda(V) \times \Lambda(V^*) \to \mathbb{K}$ defined in (4.11).*
If $T : \Lambda(V) \to \Lambda(V)$ is a linear operator, there is a unique linear operator $T^* : \Lambda(V^*) \to \Lambda(V^*)$, called the **dual operator** *of T with respect to $\langle\cdot|\cdot\rangle$, such that:*

$$\langle Tw|z^*\rangle = \langle w|T^*z^*\rangle\ , \quad \textit{for all } w \in \Lambda(V) \textit{ and } z^* \in \Lambda(V^*)\ .$$

Proof. Fix a basis $\{e_i\}_{i=1,\ldots,n}$ in V and consider its dual one $\{e^{*i}\}_{i=1,\ldots,n}$ in V^*. These bases induce bases $\{E_{p,i_p}\}_{i_p \in I_p}$ in each space $\Lambda^p(V)$ and, taking advantage from proposition 4.34, induce associated dual bases $\{E_p^{*i_p}\}_{i_p \in I_p}$ in $\Lambda^p(V^*)$. Each element E_{p,i_p} has the form $e_{j_1} \wedge \cdots \wedge e_{j_p}$ whereas each element $E_p^{*i_p}$ has the form $e^{*j_1} \wedge \cdots \wedge e^{*j_p}$. By (4.10):

$$\langle E_{p,i_p}|E_q^{*j_q}\rangle = \delta_{pq}\delta_{i_p}^{j_q}\ .$$

Obviously $\{E_{i_p,p}\}_{i_p\in I_p, p=0,\ldots,n}$ is a basis of $\Lambda(V)$ and $\{E^{*i_p}_p\}_{i_p\in I_p,p=0,\ldots,n}$ is a basis in $\Lambda(V^*)$. Consider the operator $T^* : \Lambda(V^*) \to \Lambda(V^*)$ defined by (where we write explicitly the symbols of sum)

$$T^*z^* := \sum_{p=0}^{n} \sum_{i_p\in I_p} \langle TE_{p,i_p}|z^*\rangle E^{*i_p}_p \ .$$

With the given definitions one finds that $\langle Tw|z^*\rangle = \langle w|T^*z^*\rangle$ is fulfilled by direct inspection. Now suppose that there is another operator T'^* satisfying the requirement above. As a consequence, one finds

$$\langle E_{p,i_p}|(T^* - T'^*)z^*\rangle = 0$$

for all p and i_p. Since $(T^* - T'^*)z^* = \sum_{p=0}^{n}\sum_{i_p\in I_p} c_{p,i_p}E^{*i_p}_p$, where $c_{p,i_p} = \langle E_{p,i_p}|(T^* - T'^*)z^*\rangle$, we conclude that every c_{p,i_p} vanishes and thus $(T^* - T'^*)z^* = 0$ per all $z^* \in \Lambda(V^*)$. In other words $T^* = T'^*$. □

In view of the proved proposition we can give the following definition when, for $u \in \Lambda^p(V)$, $u\wedge : \Lambda(V) \to \Lambda(V)$ is defined by the linear extension to the whole $\Lambda(V)$ of the analogous operator $u\wedge : \Lambda^q(V) \to \Lambda^{p+q}(V)$ and the spaces $\Lambda^r(V)$ with $r > n$ integer are identified with $\{0\}$, 0 being the zero vector of $\bigoplus_{p=0}^n \Lambda^p(V)$

Definition 4.36. **(Interior product.)** Consider the exterior algebras $\Lambda(V)$ and $\Lambda(V^*)$ with $dimV = n$ and field $\mathbb{K}$. If $u \in \Lambda^p(V)$, $u\lrcorner : \Lambda(V^*) \to \Lambda(V^*)$ denotes the adjoint $(u\wedge)^*$ of the operator $u\wedge : \Lambda(V) \to \Lambda(V)$ with respect to $\langle\cdot|\cdot\rangle$. If $z^* \in \Lambda(V^*)$, $u\lrcorner\, z^*$ is called **interior product** of u and z^*. ■

The most important properties of the interior product are encompassed by the following theorem.

Theorem 4.37. *Consider the exterior algebras $\Lambda(V)$ and $\Lambda(V^*)$ with $dimV = n$ and field $\mathbb{K}$. The interior product enjoys the following properties.*
(a) *If $w \in \Lambda^p(V)$ and $z^* \in \Lambda^q(V^*)$, then $w\lrcorner\, z^* \in \Lambda^{q-p}(V^*)$ (so that $w\lrcorner\, z^* = 0$ if $p > q$).*
(b) *If $w_1 \in \Lambda^p(V), w_2 \in \Lambda^q(V)$ and $z^* \in \Lambda^r(V^*)$, then*

$$(w_1 \wedge w_2)\lrcorner\, z^* = w_2\lrcorner\,(w_1\lrcorner\, z^*)\ .$$

(c) *If $x \in V$ and $z^{*1}, \ldots, y^{*p} \in V^*$, then*

$$x\lrcorner\,(y^{*1} \wedge \cdots \wedge y^{*p}) = \sum_{k=1}^{p}(-1)^{k+1}\langle x, y^{*k}\rangle y^{*1} \wedge \ldots \wedge y^{*k-1} \wedge y^{*k+1} \wedge \ldots \wedge y^{*p}\ .$$

(d) *If $x \in V$, $x\lrcorner$ is an* **anti-derivative** *on $\Lambda(V^*)$ with respect to $\wedge$, i.e.:*

$$x\lrcorner\,(z^{*1} \wedge z^{*2}) = (x\lrcorner\, z^{*1}) \wedge z^{*2} + (-1)^p z^{*1} \wedge (x\lrcorner\, z^{*2})\ , \quad \textit{for } z^{*1} \in \Lambda^p(V^*) \textit{ and } z^{*2} \in \Lambda^q(V^*)\ .$$

Proof. (a) is valid by definition. (b) can be proved as follows:

$$\langle y|(w_1 \wedge w_2)\lrcorner\, z^*\rangle = \langle w_1 \wedge w_2 \wedge y|z^*\rangle = \langle w_2 \wedge y|w_1\lrcorner\, z^*\rangle = \langle y|w_2\lrcorner\,(w_1\lrcorner\, z^*)\rangle\ .$$

The arbitrariness of y implies the validity of the thesis.
Let us pass to prove (c).

$$\langle x_2 \wedge \cdots \wedge x_p|x_1\lrcorner\,(y^{*1} \wedge \cdots \wedge y^{*p})\rangle = \langle x_1 \wedge \cdots \wedge x_p|(y^{*1} \wedge \cdots \wedge y^{*p})\rangle = \det[\langle x_i, y^{*j}\rangle]\ ,$$

where we have used (4.9). But

$$\det[\langle x_i, y^{*j}\rangle] = \sum_{k=1}^{p}(-1)^{k+1}\langle x_1, y^{*k}\rangle \det[\langle x_r, y^{*s}\rangle]\ , \quad \text{for } r \neq 1 \text{ and } s \neq k\ .$$

The right-hand side can be re-written:

$$\sum_{k=1}^{p}(-1)^{k+1}\langle x_1, y^{*k}\rangle\langle x_2 \wedge \cdots \wedge x_p|y^{*1} \wedge \ldots \wedge y^{*k-1} \wedge y^{*k+1} \wedge \ldots \wedge z^{*p}\rangle$$

The arbitrariness of $x_2 \wedge \cdots \wedge x_p$ implies the validity of the thesis with $x = x_1$.
The proof of (d) straightforward consequence of (c). □

Remarks 4.38.
(1) The use of linearity, (b) and (c) of theorem 4.37 allow one to compute the action of $x\lrcorner$ explicitly. For instance:

$$(u^{ij}e_i \wedge e_j)\lrcorner\,(\omega_{pqr}e^{*p} \wedge e^{*q} \wedge e^{*r}) = u^{ij}\omega_{pqr}e_i\lrcorner\,(\delta^p_j e^{*q} \wedge e^{*r} - \delta^q_j e^{*p} \wedge e^{*r} + \delta^r_j e^{*p} \wedge e^{*q})$$

$$= u^{ij}\,(\omega_{ijr}e^{*r} - \omega_{iqr}e^{*q} - \omega_{jir}e^{*r} + \omega_{pij}e^{*p} + \omega_{jqi}e^{*q} - \omega_{pji}e^{*p}) = 6u^{ij}\omega_{ijr}e^{*r}\ .$$

(2) With the alternative definition of $\wedge$ (4.7), it is possible to define a corresponding notion of interior product $\widetilde{\lrcorner}$ which satisfies all properties of the theorem above with respect to $\widetilde{\wedge}$. That is the unique linear extesion of

$$x\widetilde{\lrcorner}\, z^* := \frac{1}{p}x\lrcorner\, z^* \quad \text{for } x \in V \text{ and } z^* \in \Lambda^p(V^*),\, x \in V.$$

■

Chapter 5

Scalar Products and Metric Tools

This section concerns the introduction of the notion of scalar product and several applications on tensors.

5.1 Scalar products

First of all we give the definition of a *pseudo scalar product* and *semi scalar products* which differ from the notion of *scalar product* for the positivity and the non-degenerateness requirement respectively. In fact, a pseudo scalar product is a generalization of the usual definition of scalar product which has many applications in mathematical physics, relativistic theories in particular. Semi scalar products are used in several applications of quantum field theory (for instance in the celebrated *GNS theorem* [Moretti-a]).

Definition 5.1. **(Pseudo Scalar Product.)** Let V be a vector space on the field either $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.
(a) A **pseudo scalar product** is a map $(\,|\,) : V \times V \to \mathbb{K}$ which is:
(i) **bi linear**, i.e., for all $u \in V$ both $(u|\cdot) : v \mapsto (u|v)$ and $(\cdot|u) : v \mapsto (v|u)$ are linear functionals on V;
(ii) **symmetric**, i.e., $(u|v) = (v|u)$ for all $u, v \in V$;
(iii) **non-degenerate**, i.e., $(u|v) = 0$ for all $v \in V$ implies $u = 0$.
(b) If $\mathbb{K} = \mathbb{C}$, a **Hermitian pseudo scalar product** is a map $(\,|\,) : V \times V \to \mathbb{K}$ which is:
(i) **sesquilinear**, i.e., for all $u \in V$, $(u|\cdot)$ and $(\cdot|u)$ are a linear functional and an anti-linear functional on V respectively;
(ii) **Hermitian**, i.e., $(u|v) = \overline{(v|u)}$ for all $u, v \in V$;
(iii) **non-degenerate**. ■

Definition 5.2. **(Semi Scalar Product.)** Let V be a vector space on the field either $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

(a) If $\mathbb{K} = \mathbb{R}$, a **semi scalar product** is a map $(\,|\,) : V \times V \to \mathbb{R}$ which satisfies (ai),(aii) above and is

(iv) **positive semi-definite**, i.e., $(u|u) \geq 0$ for all $u \in V$.

(b) *If* $\mathbb{K} = \mathbb{C}$, *a* **Hermitian semi scalar product** *is a map* $(\,|\,) : V \times V \to \mathbb{K}$ *which satisfies* (bi),(bii) *above and is*

(iv) **positive semi-definite.** ■

Finally we give the definition of scalar product.

Definition 5.3. **(Scalar Product.)** Let V be a vector space on the field $\mathbb{K} = \mathbb{R}$ (resp. $\mathbb{C}$) endowed with a pseudo scalar product (resp. Hermitian pseudo scalar product) $(\,|\,)$. $(\,|\,)$ is called **scalar product** (**resp. Hermitian scalar product**) if $(\,|\,)$ is also a semi scalar product, i.e., if it is **positive semi-definite**. ■

Remarks 5.4.

(1) Notice that all given definitions do not require that V is finite dimensional.

(2) If $\mathbb{K} = \mathbb{C}$ and $(\,|\,)$ is not Hermitian, in general, any requirement on positivity of $(u|u)$ does not make sense because $(u|u)$ may *not* be real. If instead Hermiticity holds, we have $(u|u) = \overline{(u|u)}$ which assures that $(u|u) \in \mathbb{R}$ and thus positivity may be investigated.

(3) Actually, a (Hermitian) scalar product is *positive definited*, i.e,

$$(u|u) > 0 \quad \text{if} \quad u \in V \setminus \{0\} ,$$

because of Cauchy-Schwarz' inequality

$$|(u|v)|^2 \leq (u|u)(v|v) ,$$

which we shall prove below for semi scalar products.

(4) A **semi norm** on a vector space V with field $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$, is a map $||\ || : V \to \mathbb{K}$ such that the following properties hold:

(i) (**semi positivity**) $||v|| \in \mathbb{R}$ and in particular $||v|| \geq 0$ for all $v \in V$;

(ii) (**homogeneity**) $||\alpha v|| = |\alpha| ||v||$ for all $\alpha \in \mathbb{K}$ and $v \in V$;

(iii) (**triangular inequality**) $||u + v|| \leq ||u|| + ||v||$ for all $u, v \in V$.

A semi norm $||\ || : V \to \mathbb{K}$ is a **norm** if

(iv) $||v|| = 0$ implies $v = 0$.

Notice that for semi norms it holds: $||0|| = 0$ because of (ii) above. With the given definitions, it quite simple to show (the reader might try to give a proof) that if V with field $\mathbb{K} = \mathbb{R}$ ($\mathbb{C}$) is equipped by a (Hermitian) semi scalar product then $||v|| := \sqrt{(v|v)}$ for all $v \in V$ defines a semi norm. Furthermore if $(\,|\,)$ is a scalar product, then the associated semi norm is a norm.

(5) If a vector space V is equipped with a norm $||\ ||$ it becomes a *metric space* by defining the *distance* $d(u, v) := ||u - v||$ for all $u, v \in V$. A **Banach space** $(V, ||\ ||)$ [Rudin] is a vector space equipped with a norm such that the associates metric space is *complete*, i.e, all Cauchy's sequences converge. A **Hilbert space** [Rudin] is a Banach space with norm given by a (Hermitian if the field is $\mathbb{C}$) scalar product as said above. Hilbert spaces are the central mathematical objects used in Quantum Mechanics. ■

Exercises 5.5.

Show that if $(\,|\,)$ is a (Hermitian) semi scalar product on V with field $\mathbb{R}$ ($\mathbb{C}$) then the map on V, $v \mapsto ||v|| := \sqrt{(v|v)}$, satisfies $||u+v|| \leq ||u|| + ||v||$ as a consequence of Cauchy-Schwarz' inequality

$$|(u|v)|^2 \leq (u|u)(v|v) \,,$$

which holds true by all (Hermitian) semi scalar product.
(*Hint*. Compute $||u+v||^2 = (u+v|u+v)$ using bi linearity or sesquilinearity property of $(\,|\,)$, then use Cauchy-Schwarz' inequality.)

Theorem 5.6. **(Cauchy-Schwarz' inequality.)** *Let V be a vector space with field $\mathbb{R}$ ($\mathbb{C}$) equipped with a (Hermitian) semi scalar product $(\,|\,)$. Then, for all $u, v \in V$, Cauchy-Schwarz' inequality holds:*

$$|(u|v)|^2 \leq (u|u)(v|v) \,.$$

Proof. Consider the complex case with a Hermitian semi scalar product. Take $u, v \in V$. For all $z \in \mathbb{C}$ it must hold $(zu+v|zu+v) \geq 0$ by definition of Hermitian semi scalar product. Using sesquilinearity and Hermiticity :

$$0 \leq \overline{z}z(u|u) + (v|v) + \overline{z}(u|v) + z(v|u) = |z|^2(u|u) + (v|v) + \overline{z}(u|v) + z\overline{(u|v)} \,,$$

which can be re-written as

$$|z|^2(u|u) + (v|v) + 2Re\{\overline{z}(u|v)\} \geq 0 \,. \tag{5.1}$$

Then we pass to the polar representation of z, $z = re^{i\,\alpha}$ with $r, \alpha \in \mathbb{R}$ arbitrarily and independently fixed. Decompose also $(u|v)$, $(u|v) = |(u|v)|e^{i\ arg(u|v)}$. Inserting above we get:

$$F(r,\alpha) := r^2(u|u) + 2r|(u|v)|Re[e^{i(arg(u|v)\,-\,\alpha)}] + (v|v) \geq 0 \,,$$

for all $r \in \mathbb{R}$ when $\alpha \in R$ is fixed arbitrarily. Since the right-hand side above is a second-order polynomial in r, the inequality implies that, for all $\alpha \in \mathbb{R}$,

$$\left\{2|(u|v)|Re[e^{i(arg(u|v)\,-\,\alpha)}]\right\}^2 - 4(v|v)(u|u) \leq 0 \,,$$

which is equivalent to

$$|(u|v)|^2 \cos(arg(u|v) \,-\, \alpha) - (u|u)(v|v) \leq 0 \,,$$

for all $\alpha \in \mathbb{R}$. Choosing $\alpha = arg(u|v)$, we get Cauchy-Schwarz' inequality:

$$|(u|v)|^2 \leq (u|u)(v|v) \,.$$

The real case can be treated similarly, replacing $z \in \mathbb{C}$ with $x \in \mathbb{R}$. By hypotheses it must hold $(xu+v|xu+v) \geq 0$, which implies the analog of inequality (5.1)

$$x^2(u|u) + 2x(u|v) + (v|v) \geq 0\,.$$

That inequality must be valid for every $x \in \mathbb{R}$. Since $2x(u|v)$ could have arbitrary sign and value fixing x opportunely, if $(u|u) = 0$, $x^2(u|u) + 2x(u|v) + (v|v) \geq 0$ is possible for every x only if $(u|v) = 0$. In that case $(u|v)^2 \leq (u|u)(v|v)$ is trivially true. If $(u|u) \neq 0$, and thus $(u|u) > 0$, the inequality $x^2(u|u) + 2x(u|v) + (v|v) \geq 0$ is valid for every $x \in \mathbb{R}$ if and only if $[2(u|u)]^2 - 4(u|u)(v|v) \leq 0$. That is $(u|v)^2 \leq (u|u)(v|v)$. $\square$.

Corollary. *A bilinear symmetric (resp. sesquilinear Hermitian) map* $(\,|\,) : V \times V \to \mathbb{R}$ *(resp.* $\mathbb{C}$*) is a scalar product (resp. Hermitian scalar product) if and only if it is* **positive definite***, that is* $(u|u) > 0$ *for all* $u \in V \setminus \{\mathbf{0}\}$.

Proof. Assume that $(\,|\,)$ is a (Hermitian) scalar product. Hence $(u|u) \geq 0$ by definition and $(\,|\,)$ is non-degenerate. Moreover it holds $|(u|v)|^2 \leq (u|u)(v|v)$. As a consequence, if $(u|u) = 0$ then $(u|v) = 0$ for all $v \in V$ and thus $u = 0$ because $(\,|\,)$ is non-degenerate. We have proved that $(u|u) > 0$ if $u \neq 0$. That is, a scalar product is a positive definite bilinear symmetric (resp. sesquilinear Hermitian) map $(|) : V \times V \to \mathbb{R}$ (resp. $\mathbb{C}$). Now assume that $(|)$ is positive definite bilinear symmetric (resp. sesquilinear Hermitian). By definition it is a semi scalar product since positive definiteness implies positive semi-definiteness. Let us prove that $(\cdot|\cdot)$ is non-degenerate and this concludes the proof. If $(u|v) = 0$ for all $v \in V$ then, choosing $v = u$, the positive definiteness implies $u = 0$. $\square$

5.2 Natural isomorphism between V and V^* and metric tensor

Let us show that if V is a finite-dimensional vector space endowed with a pseudo scalar product, V is isomorphic to V^*. That isomorphism is *natural* because it is built up using the structure of vector space with scalar product only, specifying nothing further.

Theorem 5.7. **(Natural (anti)isomorphism between V and V^*.)** *Let V be a finite-dimensional vector space with field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.*

(a) *If $\mathbb{K} = \mathbb{R}$ and V is endowed with a pseudo scalar product $(\,|\,)$*

(i) *the map defined on V, $h : u \mapsto (u|\cdot) \in V^*$ (where $(u|\cdot)$ is the linear functional $(u|\cdot) : v \mapsto (u|v)$) is an isomorphism;*

(ii) $(u|v)^* := (h^{-1}(u)|h^{-1}(v))$ *for $u, v \in V^*$ defines a pseudo scalar product on V^*.*

(b) *If $\mathbb{K} = \mathbb{C}$ and V is endowed with a Hermitian pseudo scalar product $(\,|\,)$,*

(i) *the map defined on V, $h : u \mapsto (u|\cdot) \in V^*$ (where $(u|\cdot)$ is the linear functional, $(u|\cdot) : v \mapsto (u|v)$) is an anti isomorphism;*

(ii) $(u|v)^* := \overline{(h^{-1}(u)|h^{-1}(v))}$ *$(= (h^{-1}(v)|h^{-1}(u)))$ for $u, v \in V^*$ defines a Hermitian pseudo scalar product on V^*.*

Proof. First consider (i) in the cases (a) and (b). It is obvious that $(u|\cdot) \in V^*$ in both cases. Moreover the linearity or anti-linearity of the map $u \mapsto (u|\cdot)$ is a trivial consequence of the definition of pseudo scalar product and Hermitian pseudo scalar product respectively.
Then remind the well-known theorem, $dim(Kerf) + dim f(V) = dimV$, which holds true for linear and anti-linear maps from some finite-dimensional vector space V to some vector space V'. Since $dimV = dimV^*$, it is sufficient to show that $h : V \to V^*$ defined by $u \mapsto (u|\cdot)$ has trivial kernel, i.e., is injective: this also assures the surjectivity of the map. Therefore, we have to show that $(u|\cdot) = (u'|\cdot)$ implies $u = u'$. This is equivalent to show that $(u - u'|v) = 0$ for all $v \in V$ implies $u - u' = 0$. This is nothing but the non-degenerateness property, which holds by definition of (Hermitian) scalar product.
Statements (ii) cases are obvious in both by definition of (Hermitian) pseudo scalar products using the fact that h is a (anti) isomorphism. $\square$

Remarks 5.8.
(1) Notice that $(u|v)^* = (h^{-1}(u)|h^{-1}(v))$ and, for the Hermitian case, $(u|v)^* = \overline{(h^{-1}(u|)h^{-1}(v))}$ means that h and h^{-1} *(anti)preserve* the scalar products.
(2) The theorem above holds also considering a Hilbert space and its topological dual space (i.e., the subspace of the dual space consisting of continuous linear functionals on the Hilbert space). That is the mathematical content of celebrated *Riesz' representation theorem.*
(3) From theorem 5.7 it follows immediately that, if $u, v \in V$:

$$(u|v) = \langle u, h(v) \rangle = (h(v)|h(u))^* \,,$$

either if $(\,|\,)$ is Hermitian or not. $\blacksquare$

From now on we specialize to the pseudo-scalar-product case dropping the Hermitian case. Suppose $(\,|\,)$ is a pseudo scalar product on a finite-dimensional vector space V with field $\mathbb{K}$ either $\mathbb{R}$ or $\mathbb{C}$. The map $(u, v) \mapsto (u|v)$ belongs to $\mathscr{L}(V, V) = V^* \otimes V^*$ and thus it is a tensor $\mathbf{g} \in V^* \otimes V^*$. Fixing a canonical basis in $V^* \otimes V^*$ induced by a basis $\{e_i\}_{i \in I} \subset V$, we can write:

$$\mathbf{g} = g_{ij} e^{*i} \otimes e^{*j} \,,$$

where, by theorem 2.21,

$$g_{ij} = (e_i|e_j) \,.$$

Definition 5.9. **((Pseudo) Metric Tensor.)** A pseudo scalar product $(\,|\,) = \mathbf{g} \in V^* \otimes V^*$ on a finite-dimensional vector space V with field $\mathbb{K}$ either $\mathbb{R}$ or $\mathbb{C}$ is called **pseudo-metric tensor**. If $\mathbb{K} = \mathbb{R}$, a pseudo-metric tensor is called **metric tensor** if it defines a scalar product. $\blacksquare$

Remarks 5.10.
(1) By theorem 3.8, the isomorphism $h : V \to V^*$ is represented by a tensor of $V^* \otimes V^*$ which acts on elements of V by means of a product of tensors and a contraction. The introduction

of the pseudo-metric tensor allows us to represent the isomorphism $h : V \to V^*$ by means of the abstract index notation determining the tensor representing h as we go to illustrate. Since $h : u \mapsto (u|\ ) \in V^*$ and $(u|v) = (e_i|e_j)u^i v^j = g_{ij}u^i v^j$ we trivially have:

$$(hu)_j = g_{ij}u^i \, .$$

$h(u)$ is obtained by the product $\mathbf{g} \otimes u$ followed by a contraction. Hence, in the sense of theorem 3.8, the linear map $h : V \to V^*$ is represented by the tensor g itself.
(2) Pseudo metric tensors are *symmetric* because of the symmetry of pseudo scalar products:

$$\mathbf{g}(u, v) = (u|v) = (v|u) = \mathbf{g}(v, u) \, .$$

(3) The symmetry requirement on the metric tensor is not necessary to define an isomorphism between V and V^*. In Weyl spinor theory [Wald, Streater-Wightman], the space V is a two-dimensional complex vector space whose elements are called Weyl *spinors*. V is equipped with a fixed *antisymmetric* tensor $\epsilon \in V^* \otimes V^*$ (the so-called *metric spinor*) defining a non-degenerate linear map from V to V^* by contraction. Using abstract index notation

$$V \ni \xi^A \mapsto \epsilon_{AB}\xi^B \in V^* \, .$$

Notice that, differently from the metric tensor case, now $\epsilon_{AB}\xi^B = -\epsilon_{BA}\xi^B$. Non-degenerateness of ϵ entails that the map $\xi^A \mapsto \epsilon_{AB}\xi^B$ is an isomorphism from V to V^*. ∎

Components of pseudo-metric tensors with respect to canonical basis enjoy some simple but important properties which are listed below.

Theorem 5.11. **(Properties of the metric tensor.)** *Referring to def. 5.9, the components of any pseudo-metric tensor* $\mathbf{g}$*,* $g_{ij} := \mathbf{g}(e_i, e_j)$ *with respect to the canonical basis induced in* $V^* \otimes V^*$ *by any basis* $\{e_i\}_{i=1,\ldots,n} \subset V$*, enjoy the following properties:*
(1) *define a symmetric matrix* $[g_{ij}]$*, i.e,*

$$g_{ij} = g_{ji} \, ;$$

(2) $[g_{ij}]$ *is non singular, i.e., it satisfies:*

$$\det[g_{ij}] \neq 0 \, ;$$

(3) *if* $\mathbb{K} = \mathbb{R}$ *and* g *is a scalar product, the matrix* $[g_{ij}]$ *is positive definite.*

Proof. (1) It is obvious: $g_{ij} = (e_i|e_j) = (e_j|e_i) = g_{ji}$.
(2) Suppose $\det[g_{ij}] = 0$ and define $n = dimV$. The linear map $\mathbb{K}^n \to \mathbb{K}^n$ determined by the matrix $g := [g_{ij}]$ has a non-trivial kernel. In other words, there are n reals u^j, $j = 1, \ldots, n$ defining a $\mathbb{K}^n$ vector $[u] := (u^1, \ldots, u^n)^t$ with $g[u] = 0$ and $[u] \neq 0$. In particular $[v]^t g[u] = 0$ for whatever choice of $[v] \in \mathbb{K}^n$. Defining $u := u^j e_j$, the obtained result implies that there is

$u \in V \setminus \{0\}$ with $(u|v) = (v|u) = 0$ for all $v \in V$. This is impossible because $(|)$ is non degenerate by hypothesis.
(3) The statement, $(u|u) > 0$ if $u \in V \setminus \{\mathbf{0}\}$, reads, in the considered canonical basis $[u]^t g[u] > 0$ for $[u] \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. That is one of the equivalent definitions of a positive definite matrix g. $\square$

The following theorem shows that a (pseudo) scalar product can be given by the assignment of a convenient tensor which satisfies some properties when represented in some canonical bases. The important point is that there is no need to check on these properties for *all* canonical bases, verification for a single canonical basis is sufficient.

Theorem 5.12. **(Assignment of a (pseudo) scalar product.)** *Let V be a finite-dimensional vector space with field $\mathbb{K}$ either $\mathbb{R}$ or $\mathbb{C}$. Suppose $\mathbf{g} \in V^* \otimes V^*$ is a tensor such that there is a canonical basis of $V^* \otimes V^*$ where the components g_{ij} of $\mathbf{g}$ define a symmetric matrix $g := [g_{ij}]$ with non-vanishing determinant. Then $\mathbf{g}$ is a pseudo-metric tensor, i.e. a pseudo scalar product.*
Furthermore, if $\mathbb{K} = \mathbb{R}$ and $[g_{ij}]$ is positive definite, the pseudo scalar product is a scalar product.

Proof. If $\mathbf{g}$ is represented by a symmetric matrix of components in a canonical basis then it holds in all remaining bases and the tensor is symmetric (see exercise 4.221). This implies that $(u|v) := \mathbf{g}(u, v)$ is a bi-linear symmetric functional. Suppose $(\,|\,)$ is degenerate, then there is $u \in V$ such that $u \neq \mathbf{0}$ and $(u|v) = 0$ for all $v \in V$. Using notations of the proof of the item (2) of theorem 5.11, we have in components of the considered canonical bases, $[u]^t g[v] = 0$ for all $[v] = (v^1, \ldots, v^n)^t \in \mathbb{K}^n$ where $n = dimV$. Choosing $[v] = g[u]$, it also holds $[u]^t gg[u] = 0$. Since $g = g^t$, this is equivalent to $(g[u])^t g[u] = 0$ which implies $g[u] = 0$. Since $[u] \neq \mathbf{0}$, g cannot be injective and $\det g = 0$. This is not possible by hypotheses, thus $(\,|\,)$ is non-degenerate. We conclude that $(u|v) := g(u, v)$ define a pseudo scalar product.
Finally, if $\mathbb{K} = \mathbb{R}$ and g is also positive definite, $(\,|\,)$ itself turns out to be positive definite, i.e., it is a scalar product since $(u|u) = [u]^t g[u] > 0$ if $[u] \neq \mathbf{0}$ (which is equivalent to $u \neq \mathbf{0}$). $\square$

5.2.1 Signature of pseudo-metric tensor, pseudo-orthonormal bases and pseudo-orthonormal groups

Let us introduce the notion of *signature* of a pseudo-metric tensor in a vector space with field $\mathbb{R}$ by reminding Sylvester's theorem whose proof can be found in any linear algebra textbook. The definition is interlaced with the definition of *pseudo-orthonormal basis* and *pseudo-orthonormal group.*

Theorem 5.13. **(Sylvester's theorem.)** *Let G be a real symmetric $n \times n$ matrix.*
(a) *There is a non-singular (i.e., with non vanishing determinant) real $n \times n$ matrix D such that:*

$$DGD^t = diag(0, \ldots, 0, -1, \ldots, -1, +1, \ldots, +1) ,$$

where the reals $0, -1, +1$ *appear* $v \geq 0$ *times,* $m \geq 0$ *times and* $p \geq 0$ *times respectively with* $v + m + p = n$.
(b) *the triple* (v, m, p) *does not depend on* D. *In other words, if, for some non-singular real* $n \times n$ *matrix matrix* $E \neq D$, EGE^t *is diagonal and the diagonal contains reals* $0, -1, +1$ *only (in whatever order), then* $0, -1, +1$ *respectively appear* v *times,* m *times and* p *times.*

If $\mathbf{g} \in V^* \otimes V^*$ is a pseudo-metric tensor on the finite-dimensional vector space v with field $\mathbb{R}$, the transformation rule of the components of g with respect to canonical bases (see theorem 3.6) induced by bases $\{e_i\}_{i\in I}$, $\{e'_j\}_{j\in I}$ of V are

$$g'_{pq} = B_p{}^i B_q{}^j g_{ij} \ .$$

Defining $g' := [g_{pq}]$, $g := [g_{ij}]$, $B := [B_h{}^k]$, they can be re-written as

$$g' = BgB^t \ .$$

We remind (see theorem 3.6) that the non-singular matrices B are defined by $B = A^{-1t}$, where $A = [A^i{}_j]$ and $e_m = A^l{}_m e'_l$. Notice that the specification of B is completely equivalent to the specification of A because $A = B^{-1t}$.
Hence, since g is real and symmetric by theorem 5.11, Sylvester's theorem implies that, starting from any basis $\{e_i\}_{i\in I} \subset V$ one can find another basis $\{e'_j\}_{j\in I}$, which induces a canonical basis in $V^* \otimes V^*$ where the pseudo-metric tensor is represented by a diagonal matrix. It is sufficient to pick out a transformation matrix B as specified in **(a)** of theorem 5.13. In particular, one can find B such that each element on the diagonal of g' turns out to be either -1 or $+1$ only. The value 0 is not allowed because it would imply that the matrix has vanishing determinant and this is not possible because of theorem 5.11. Moreover the pair (m, p), where (m, p) are defined in theorem 5.13, does not depend on the basis $\{e'_j\}_{j\in I}$. In other words, it is an *intrinsic* property of the pseudo-metric tensor: that is the *signature* of the pseudo-metric tensor.

Definition 5.14. **(Pseudo Orthonormal Bases and Signature).** Let $\mathbf{g} \in V^* \otimes V^*$ be a pseudo-metric tensor on the finite-dimensional vector space V with field $\mathbb{R}$.
(a) A basis $\{e_i\}_{i\in I} \subset V$ is called **pseudo orthonormal** with respect to $\mathbf{g}$ if the components of $\mathbf{g}$ with respect to the canonical basis induced in $V^* \otimes V^*$ form a diagonal matrix with eigenvalues in $\{-1, +1\}$. In other words, $\{e_i\}_{i\in I}$ is pseudo orthonormal if

$$(e_i, e_j) = \pm\delta_{ij} \ .$$

If the pseudo-metric tensor is a metric tensor the pseudo-orthonormal bases are called orthonormal bases.
(b) The pair (m, p), where m is the number of eigenvalues -1 and p is the number of eigenvalues $+1$ of a matrix representing the components of $\mathbf{g}$ in an orthonormal basis is called **signature** of $\mathbf{g}$.
(c) $\mathbf{g}$ and its signature are said **elliptic** or **Euclidean** or **Riemannian** if $m = 0$, **hyperbolic**

if $m > 0$ and $p \neq 0$, **Lorentzian** or **normally hyperbolic** if $m = 1$ and $p \neq 0$.
(d) If $\mathbf{g}$ is hyperbolic, an orthonormal basis $\{e_i\}_{i \in I}$ is said to be **canonical** if the matrix of the components of $\mathbf{g}$ takes the form:

$$diag(-1, ..., -1, +1, ..., +1) \, .$$

■

Exercises 5.15.
Show that a *pseudo-metric* tensor $\mathbf{g}$ is a *metric* tensor if and only if its signature is elliptic.

Remarks 5.16. If $\{e_i\}_{i \in I}$ is an orthonormal basis with respect to a hyperbolic pseudo-metric tensor $\mathbf{g}$, one can trivially re-order the vectors of the basis giving rise to a canonical orthonormal basis. ■

Let us consider a pseudo-metric tensor $\mathbf{g}$ in V with field $\mathbb{R}$. Let (m, p) be the signature of $\mathbf{g}$ and let $\mathcal{N}_g$ be the class of all of the canonical pseudo-orthonormal bases in V with respect to $\mathbf{g}$. In the following we shall indicate by η the matrix $diag(-1, ..., -1, +1, ..., +1)$ which represents the components of $\mathbf{g}$ with respect to each basis of $\mathcal{N}_\mathbf{g}$. If A is a matrix corresponding to a change of basis in $\mathcal{N}_\mathbf{g}$, and $B := A^{-1t}$ is the associated matrix concerning change of basis in V^*, it has to hold

$$\eta = B \eta B^t \, .$$

Conversely, each real $n \times n$ matrix B which satisfies the identity above determines $A := B^{-1t}$ which represents a change of basis in $\mathcal{N}_\mathbf{g}$. That A is well defined because every B satisfying $\eta = B\eta B^t$ is *non singular* and, more precisely, $\det B = \pm 1$. Indeed, taking the determinant of both sides in $\eta = B\eta B^t$, using the fact that $\det \eta = (-1)^m$ and $\det B = \det B^t$, we conclude that $(\det B)^2 = 1$ and thus $\det B = \pm 1$. As a consequence, since $\det A = (\det B)^{-1}$, one also has $\det A = \pm 1$.
The identity $\eta = B\eta B^t$ can be equivalently re-written in terms of the matrix A. Since $A = B^{-1t}$, one has $B^{-1} = A^t$. Thus, we applying A^t on the left and A on the right of $\eta = B\eta B^t$, we get the equivalent relation in terms of the matrix A:

$$\eta = A^t \eta A \, .$$

That equation completely determines the set $O(m, p) \subset GL(n, \mathbb{R})$ $(n = m + p)$ of all real non-singular $n \times n$ matrices which correspond to changes of bases in $\mathcal{N}_\mathbf{g}$.
It is simply proved that $O(m, p)$ is a subgroup of $GL(n, \mathbb{R})$. We can state the following definition.

Definition 5.17. If η is the matrix $diag(-1, ..., -1, +1, ..., +1)$ where -1 occurs m times and $+1$ occurs p times, the subgroup of $GL(m + p, \mathbb{R})$:

$$O(m, p) := \{A \in GL(m + p, \mathbb{R}) \mid \eta = A^t \eta A\}$$

is called the **pseudo orthogonal** group of order (m, p). ■

Notice that, if $m = 0$, $O(0,p) = O(n)$ reduces to the usual orthogonal group of order n. $O(1,3)$ is the celebrated **Lorentz group** which is the central mathematical object in relativistic theories. We shall come back on those issues in the last two chapters, focusing on the Lorentz group in particular.

Exercises 5.18.

1. Show that if $A \in O(m,p)$ then A^{-1} exists and

$$A^{-1} = \eta A^t \eta \,.$$

2. Show that $O(m,p)$ is a group with respect to the usual multiplication of matrices.

Remarks 5.19. This implies that $O(m,p)$ is a subgroup of $GL(n,\mathbb{R})$ with $n = p + m$. (*Hint*. You have to prove that, (1) the identity matrix I belongs to $O(m,p)$, (2) if A and A' belong to $O(m,p)$, AA' belongs to $O(m,p)$, (3) if A belongs to $O(m,p)$, then A^{-1} exists and belongs to $O(m,p)$.)

3. Show that $SO(m,p) := \{A \in O(m,p) \mid \det A = 1\}$ is not the empty set and is a subgroup of $O(m,p)$. $SO(m,p)$ is called the *special pseudo orthogonal group* of order (m,p).

4. Consider the special Lorentz group $SO(1,3)$ and show that the set (called the *special orthochronous Lorentz group*)

$$SO(1,3)\Uparrow := \{A \in SO(1,3) \mid A^1{}_1 > 0\}$$

is a not empty subgroup. ■

5.2.2 Raising and lowering indices of tensors

Consider a finite dimensional vector space V with field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ endowed with a pseudo-metric tensor $\mathbf{g}$. As we said above, there is a natural isomorphism $h : V \to V^*$ defined by $h : u \mapsto (u|\cdot) = \mathbf{g}(u,\cdot)$. This isomorphism may be extended to the whole tensor algebra $\mathscr{A}_{\mathbb{K}}(V)$ using the universality theorem and def. 4.11.
Indeed, consider a space $S \in \mathscr{A}_{\mathbb{K}}(V)$ of the form $A \otimes V \otimes B$, where A and B are tensor spaces of the form $U_1 \otimes \ldots \otimes U_k$, and $U_{k+1} \otimes \ldots \otimes U_m$ respectively, U_i being either V or V^*. We may define the operators:

$$h^{\otimes} := I_1 \otimes \ldots \otimes I_k \otimes h \otimes I_{k+1} \otimes \ldots \otimes I_m : A \otimes V \otimes B \to A \otimes V^* \otimes B \,,$$

and

$$(h^{-1})^{\otimes} := I_1 \otimes \ldots \otimes I_k \otimes h^{-1} \otimes I_{k+1} \otimes \ldots \otimes I_m : A \otimes V^* \otimes B \to A \otimes V \otimes B \,,$$

where $I_j : U_j \to U_j$ is the identity operator. Using the remark after def. 4.11, one finds

$$(h^{-1})^{\otimes} h^{\otimes} = I_1 \otimes \ldots \otimes I_k \otimes (h^{-1}h) \otimes I_{k+1} \otimes \ldots \otimes I_m = Id_{A \otimes V \otimes B} \,,$$

and

$$h^{\otimes}(h^{-1})^{\otimes} = I_1 \otimes \ldots \otimes I_k \otimes (hh^{-1}) \otimes I_{k+1} \otimes \ldots \otimes I_m = Id_{A\otimes V^* \otimes B} .$$

Therefore $h^{\otimes}$ is an isomorphism with inverse $(h^{-1})^{\otimes}$.

The action of $h^{\otimes}$ and $(h^{-1})^{\otimes}$ is that of **lowering** and **raising indices** respectively. In fact, in abstract index notation, one has:

$$h^{\otimes} : t^{AiB} \mapsto t^{A}{}_{j}{}^{B} := t^{AiB} g_{ij} ,$$

and

$$(h^{-1})^{\otimes} : u^{A}{}_{i}{}^{B} \mapsto u^{AjB} := t^{A}{}_{i}{}^{B} \tilde{g}^{ij} .$$

Above g_{ij} represents the pseudo-metric tensor as specified in the remark 1 after def. 5.9. What about the tensor $\tilde{g} \in V \otimes V$ representing h^{-1} via theorem 3.8?

Theorem 5.20. *Let $h : V \to V^*$ be the isomorphism determined by a pseudo scalar product, i.e. a pseudo-metric tensor $\mathbf{g}$ on the finite-dimensional vector space V with field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.*
(a) *The inverse map $h^{-1} : V^* \to V$ is represented via theorem 3.8 by a* **symmetric** *tensor $\tilde{\mathbf{g}} \in V \otimes V$ such that, if $\{e_i\}_{i\in I}$ is a basis of V, $\tilde{g}^{rs} := \tilde{g}(e^{*r}, e^{*s})$ and $g_{ij} := \mathbf{g}(e_i, e_j)$, then the matrix $[\tilde{g}^{ij}]$ is the* inverse matrix *of $[g_{ij}]$.*
(b) *The tensor $\tilde{\mathbf{g}}$ coincides with the pseudo-metric tensor* **with both indices raised**.

Proof. **(a)** By theorem 3.8, h^{-1} determines a tensor $\tilde{\mathbf{g}} \in V \otimes V$ with $h^{-1}(u^*) = \tilde{\mathbf{g}}(u^*, \cdot)$. In components $(h^{-1}u^*)^i = u^*_k \tilde{g}^{ki}$. On the other hand it must be

$$h(h^{-1}u^*) = u^*$$

or,

$$u^*_k \tilde{g}^{ki} g_{ir} = u^*_r ,$$

for all $u^* \in V^*$. This is can be re-written

$$[u^*]^t(\tilde{g}g - I) = 0 ,$$

for all $\mathbb{K}^n$ vectors $[u^*] = (u^*_1, \ldots, u^*_n)$. Then the matrix $(\tilde{g}g - I)^t$ is the null matrix. This implies that

$$\tilde{g}g = I ,$$

which is the thesis. $\tilde{g}$ is symmetric because is the inverse of a symmetric matrix and thus also the tensor $\tilde{\mathbf{g}}$ is symmetric.
(b) Let g^{ij} be the pseudo-metric tensor with both indices raised, i.e.,

$$g^{ij} := g_{rk}\ \tilde{g}^{kj}\tilde{g}^{ri} .$$

By **(a)**, the right-hand side is equal to:

$$\delta^j_r \tilde{g}^{ri} = \tilde{g}^{ji} = \tilde{g}^{ij} .$$

That is the thesis. $\square$

Remarks 5.21.
(1) The isomorphisms $h : V \to V^*$ and $h^{-1} : V^* \to V$ are often respectively indicated by $\flat : V \to V^*$ and $\sharp : V^* \to V$ and are called **musical isomorphisms**. In particular, the following notations are used

$$v^\sharp := \sharp(v) \quad \text{and} \quad u^\flat := \flat(u)$$

respectively corresponding to the process of raising and lowering an index.
(2) Another result which arises from the proof of the second part of the theorem is that

$$g_i^j = \delta_i^j .$$

(3) When a vector space is endowed with a pseudo scalar product, tensors can be viewed as abstract objects which may be represented either as covariant or contravariant concrete tensors using the procedure of raising and lowering indices. For instance, a tensor t^{ij} of $V \otimes V$ may be viewed as a covariant tensor when "represented" in its covariant form $t_{pq} := g_{pi}\, g_{qj}\, t^{ij}$. Also, it can be viewed as a mixed tensor $t_p{}^j := g_{pi}\, t^{ij}$ or $t^i{}_q := g_{qj}\, t^{ij}$.
(4) Now consider a finite-dimensional vector space on $\mathbb{R}$, V, endowed with a metric tensor $\mathbf{g}$, i.e., with *elliptic signature*. In orthonormal bases the contravariant and covariant components numerically coincides because $g_{ij} = \delta_{ij} = g^{ij}$. This is the reason because, using the usual scalar product of vector spaces isomorphic to $\mathbb{R}^n$ and working in orthonormal bases, the difference between covariant and contravariant vectors does not arise.
Conversely, in relativistic theories where a Lorentzian scalar product is necessary, the difference between covariant and contravariant vectors turns out to be evident also in orthonormal bases, since the diagonal matrix $[g_{ij}]$ takes an eigenvalue -1. $\blacksquare$

Chapter 6

Pseudo tensors and tensor densities

This section is devoted to introduce very important tools either in theoretical/mathematical physics and in pure mathematics: pseudo tensors and tensor densities.

6.1 Orientation and pseudo tensors

The first example of "pseudo" object we go to discuss is a *orientation* of a *real* vector space.

Consider a finite-dimensional vector space V with field $\mathbb{R}$. In the following $\mathcal{B}$ indicates the set of all the vector bases of V. Consider two bases $\{e_i\}_{i\in I}$ and $\{e'_j\}_{j\in I}$ in $\mathcal{B}$. Concerning the determinant of the transformation matrix $A := [A^r{}_s]$, with $e_i = A^j{}_i e'_j$, there are two possibilities only: $\det A > 0$ or $\det A < 0$. It is a trivial task to show that the relation in $\mathcal{B}$:

$$\{e_i\}_{i\in I} \sim \{e'_j\}_{j\in I} \quad \text{iff} \quad \det A > 0$$

where A indicates the transformation matrix as above, is an *equivalence relation*. Since there are the only two possibilities above, the partition of $\mathcal{B}$ induced by $\sim$ is made of two *equivalence classes* $\mathcal{B}_1$ and $\mathcal{B}_2$. Hence if a basis belongs to $\mathcal{B}_1$ or $\mathcal{B}_2$ any other basis belongs to the same set if and only if the transformation matrix has positive determinant.

Definition 6.1. **(Orientation of a vector space.)** Consider a finite-dimensional vector space V with field $\mathbb{R}$, an **orientation** of V is a bijective map $\mathcal{O} : \{\mathcal{B}_1, \mathcal{B}_2\} \to \{-1, +1\}$. If V has an orientation $\mathcal{O}$, is said to be **oriented** and a basis $\{e_i\}_{i\in I} \in \mathcal{B}_k$ is said to be **positive oriented** if $\mathcal{O}(\mathcal{B}_k) = +1$ or **negative oriented** if $\mathcal{O}(\mathcal{B}_k) = -1$. ■

Remarks 6.2. The usual physical vector space can be oriented "by hand" using the natural basis given by our own right hand. When we use the right hand to give an orientation we determine $\mathcal{O}^{-1}(+1)$ by the exhibition of a basis contained therein. ■

The given definition can be, in some sense, generalized with the introduction of the notion of *pseudo tensor*.

Definition 6.3. **(Pseudotensors.)** Let V be a finite-dimensional vector space with field $\mathbb{R}$. Let S be a tensor space of $\mathscr{A}_{\mathbb{R}}(V)$. A **pseudo tensor of** S is a bijective map $t_s : \{\mathcal{B}_1, \mathcal{B}_2\} \to \{s, -s\}$, where $s \in S$. Moreover:
(a) the various tensorial properties enjoyed by both s and $-s$ are attributed to t_s. (So, for instance, if s, and thus $-s$, is symmetric, t_s is said to be symmetric);
(b) if $\{e_i\}_{i\in I} \in \mathcal{B}_i$, the **components of** t_s with respect to the canonical bases induced by that basis are the components of $t_s(\mathcal{B}_i)$. ■

Remarks 6.4.
(1) The given definition encompasses the definition of *pseudo scalar*.
(2) It is obvious that the assignment of a pseudo tensor t_s of, for instance, $V^{n\otimes} \otimes V^{*m\otimes}$, is equivalent to the assignment of components

$$t^{i_1 \ldots i_n}{}_{j_1 \ldots j_m}$$

for each canonical basis

$$\{e_{i_1} \otimes \ldots e_{i_n} \otimes e^{*j_1} \otimes \cdots \otimes e^{*j_m}\}_{i_1,\ldots,i_n,j_1,\ldots,j_m \in I}$$

such that the transformation rules passing to the basis

$$\{e'_{r_1} \otimes \ldots e'_{r_n} \otimes e'^{*l_1} \otimes \cdots \otimes e'^{*l_m}\}_{r_1,\ldots,r_n,l_1,\ldots,l_m \in I} \,,$$

are given by:

$$t'^{k_1 \ldots k_n}{}_{h_1 \ldots h_m} = \frac{\det A}{|\det A|} \; A^{k_1}_{\;\;i_1} \cdots A^{k_n}_{\;\;i_n} \, B_{h_1}^{\;\;j_1} \cdots B_{h_m}^{\;\;j_m} \; t^{i_1 \ldots i_n}{}_{j_1 \ldots j_n} \,,$$

where $e_l = A^m_{\;\;l} e'_m$ and $B = A^{-1t}$ with $B := [B_k^{\;\;j}]$ and $A := [A^p_{\;\;q}]$.
In fact, $t^{i_1 \ldots i_n}{}_{j_1 \ldots j_m} = s^{i_1 \ldots i_n}{}_{j_1 \ldots j_m}$ if the considered base is in $t_s^{-1}(+1)$ or $t^{i_1 \ldots i_n}{}_{j_1 \ldots j_m} = (-s)^{i_1 \ldots i_n}{}_{j_1 \ldots j_m}$ if the considered base is in $t_s^{-1}(-1)$.
(3) Evidently the pseudotensors of a fixed type (e.g. (p,q)) form a vector space. ■

Examples 6.5.
1. Consider the **magnetic field** $B = B^i e_i$ where e_1, e_2, e_3 is a right-hand orthonormal basis of the space V_3 of the vectors with origin in a point of the physical space E_3. Actually, as every physicist knows, changing basis, the components of B changes as usual only if the new basis is a right-hand basis, otherwise a sign $-$ appears in front of each component. That is a physical requirement due to the Lorentz law. This means that the magnetic field has to be represented in terms of **pseudo vectors**.

6.2 Ricci's pseudo tensor

A particular pseudo tensor is Ricci's one which is very important in physical applications. The definition of this pseudo tensor requires a preliminary discussion.
Consider an n-dimensional vector space V with a pseudo-metric tensor $\mathbf{g}$. We know that, changing basis $\{e_i\}_{i\in I} \to \{e'_j\}_{j\in I}$, the components of the pseudo-metric tensor referred to the corresponding canonical bases, transform as:

$$g' = BgB^t ,$$

where $g = [g_{ij}]$, $g' = [g'_{pq}]$ and $B = A^{-1t}$, $A := [A^p{}_q]$, $e_l = A^m{}_l e'_m$. This implies that

$$\det g' = (\det B)^2 \ \det g , \tag{6.1}$$

which is equivalent to

$$\sqrt{|\det g'|} = |\det A|^{-1} \ \sqrt{|\det g|} \tag{6.2}$$

Now fix a basis $E := \{e_i\}_{i\in I}$ of V in the set of bases $\mathcal{B}$, consider the canonical basis induced in $V^{*n\otimes}$ where $n = dimV$, $\{e^{*i_1} \otimes \ldots \otimes e^{*i_n}\}_{i_1,\ldots,i_n\in I}$. Finally define the n-form associated with that basis:

$$\mathcal{E}^{(E)} := n!\sqrt{|\det g|}e^{*1} \wedge \cdots \wedge e^{*n} \tag{6.3}$$

We want to prove that the map

$$\mathcal{B} \ni E \mapsto \mathcal{E}^{(E)} \in \Lambda^n(V)$$

defines an antisymmetric pseudotensor of maximal order. In other words we intend to prove that, if $E, E' \in \mathcal{B}$ with $E = \{e_i\}_{i=1,\ldots,n}$, $E' = \{e'_j\}_{i=1,\ldots,n}$ with $e_i = A^j{}_i e'_j$ then:

$$\mathcal{E}^{(E')} = \frac{\det A}{|\det A|}\mathcal{E}^{(E)} . \tag{6.4}$$

To prove (6.4), we expand $\mathcal{E}^{(E)}$ in terms of the correspondings objects associated with the basis E'

$$\mathcal{E}^{(E)} = n!\sqrt{|\det g|}e^{*1} \wedge \cdots \wedge e^{*n} = n!\sqrt{|\det g'|}\frac{1}{|\det B|}B_{j_1}{}^1 \cdots B_{j_n}{}^n e'^{*j_1} \wedge \cdots \wedge e'^{*j_n}$$

$$= n!\frac{\det B}{|\det B|}\frac{\sqrt{|\det g'|}}{\det B}B_{j_1}{}^1 \cdots B_{j_n}{}^n e'^{*j_1} \wedge \cdots \wedge e'^{*j_n} .$$

We have found that:

$$\mathcal{E}^{(E)} = n!\frac{\det A}{|\det A|}\frac{\sqrt{|\det g'|}}{\det B}B_{j_1}{}^1 \cdots B_{j_n}{}^n e'^{*j_1} \wedge \cdots \wedge e'^{*j_n} \tag{6.5}$$

Now focus attention on the last term in the right-hand side:

$$B_{j_1}{}^1 \cdots B_{j_n}{}^n e'^{*j_1} \wedge \cdots \wedge e'^{*j_n} .$$

In view of the properties of the wedge product, if there are some repeated indices among the values $j_1, \ldots, j_n$, we have $e'^{*j_1} \wedge \cdots \wedge e'^{*j_n} = 0$. So, the only possibility is that $(j_1, \ldots, j_n)$ is a permutation of $(1, 2, \ldots, n)$, that is $j_k = \sigma(k)$ for some $\sigma \in \mathcal{P}_n$. Since: $e'^{*\sigma(1)} \wedge \cdots \wedge e'^{*\sigma(n)} = \epsilon_\sigma e'^{*1} \wedge \cdots \wedge e'^{*n}$, we finally have:

$$B_{j_1}{}^1 \cdots B_{j_n}{}^n e'^{*j_1} \wedge \cdots \wedge e'^{*j_n} = \sum_{\sigma \in \mathcal{P}_n} \epsilon_\sigma B_{\sigma(1)}{}^1 \cdots B_{\sigma(n)}{}^n e'^{*1} \wedge \cdots \wedge e'^{*n} = (\det B) e'^{*1} \wedge \cdots \wedge e'^{*n} .$$

Inserting this result in (6.5), we conclude that:

$$\mathcal{E}^{(E)} = n! \frac{\det A}{|\det A|} \frac{\sqrt{|\det g'|}}{\det B} (\det B) e'^{*1} \wedge \cdots \wedge e'^{*n} = \frac{\det A}{|\det A|} \mathcal{E}^{(E')} ,$$

namely

$$\mathcal{E}^{(E')} = \frac{|\det A|}{\det A} \mathcal{E}^{(E)}$$

so that (6.4) turns out to be proved because $\frac{|\det A|}{\det A} = \frac{\det A}{|\det A|}$.

Definition 6.6. **(Ricci's Pseudo tensor).** Let V be a vector space with field $\mathbb{R}$ and dimension $n < +\infty$, endowed with a pseudo-metric tensor $\mathbf{g}$. **Ricci's pseudo tensor** is the anti-symmetric $(0, n)$ pseudo tensor $\mathcal{E}$ represented in each canonical basis by

$$\mathcal{E} = n! \sqrt{|\det g|}\, e^{*1} \wedge \cdots \wedge e^{*n} , \tag{6.6}$$

where $g = [g_{ij}]$, g_{ij} being the components of g in the considered basis. ■

Remarks 6.7.
(1) $\mathcal{E}$ can also be written in standard components:

$$\mathcal{E} = \mathcal{E}_{i_1 \cdots i_n} e^{*i_1} \otimes \cdots \otimes e^{*i_n} ,$$

where $i_k \in \{1, 2, \ldots, , n\}$ for $k = 1, 2 \ldots, n$ and, as usual, we are using the summation convention over the repeated indices. It is easy to prove that:

$$\mathcal{E}_{i_1 \cdots i_n} = \sqrt{|\det g|} \eta_{i_1 \ldots i_n} \tag{6.7}$$

where

$$\eta_{i_1 \ldots i_n} = \epsilon_{\sigma_{i_1 \ldots i_n}} ,$$

whenerver $(i_1, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$, otherwise

$$\eta_{i_1 \ldots i_n} = 0 ,$$

above $\epsilon_{\sigma_{i_1\ldots i_n}}$ is just the sign of the permutation $\sigma_{i_1\ldots i_n} \in \mathscr{P}_n$ defined by:

$$(\sigma(1),\ldots,\sigma(n)) = (i_1,\ldots,i_n)\,.$$

(2) If V is oriented, it is possible to define $\mathcal{E}$ as a *proper tensor* instead a pseudo tensor. In this case one defines the components in a canonical basis associated with a positive-oriented basis of V as

$$\mathcal{E}_{i_1\ldots i_n} := \sqrt{|\det g|}\;\eta_{i_1\ldots i_n}\,, \text{that is } \;\mathcal{E} = n!\,\sqrt{|\det g|}\,e^{*1}\wedge\cdots\wedge e^{*n}$$

and

$$\mathcal{E}_{i_1\ldots i_n} := -\sqrt{|\det g|}\;\eta_{i_1\ldots i_n}\,, \text{that is } \;\mathcal{E} = -n!\,\sqrt{|\det g|}\,e^{*1}\wedge\cdots\wedge e^{*n}$$

if the used basis is associated with a basis of V which is negative-oriented. One can prove straightforwardly that the defined components give rise to a tensor of $V^{*n\otimes}$ called **Ricci tensor**. This alternative point of view is equivalent, in the practice, to the other point of view corresponding to definition 6.6. ■

Ricci's pseudo tensor has various applications in mathematical physics in particular when it is used as a linear operator which produces pseudo tensors when it acts on tensors. In fact, consider $t \in V^{m\otimes}$ with $m \leq n$. Fix a basis in V and, in the canonical bases induced in the space of tensors by that basis, consider the action of ε on t:

$$t^{i_1\ldots i_m} \mapsto \tilde{t}_{j_1\ldots j_{n-m}} := \mathcal{E}_{j_1\ldots j_{n-m}i_1\ldots i_m}t^{i_1\ldots i_m}\,.$$

We leave to the reader the proof of the fact that the components $\tilde{t}_{j_1,\ldots,j_{n-m}}$ define a antisymmetric pseudo tensor of order $(0, n-m)$ which is called the **conjugate** pseudo tensor of t. Let us restrict to antisymmetric tensors and pseudotensores. The correspondence above from antisymmetric *tensors* of order $(m,0)$ and antisymmetric *pseudo*tensors of order $(0, n-m)$ is surjective, due to the formula $(p = n-m)$

$$t_{i_1\ldots i_p} = \frac{|\det g|}{\det g}\;\frac{(-1)^{p(n-p)}}{p!(n-p)!}\;\mathcal{E}_{i_1\ldots i_p j_1\ldots j_m}\mathcal{E}^{j_1\ldots j_m r_1\ldots r_p}\;\,t_{r_1\ldots t_p}\,.$$

However since the space of antisymmetric tensors of order $(m,0)$ and that of antisymmetric pseudotensors of order $(0, n-m)$ is identical:

$$\binom{n}{m} = \binom{n}{n-m}\,,$$

we conclude that the linear map associating $(m,0)$ antisymmetric tensors with their $(0, n-m)$ (antisymmetric) dual pseudotensors is an isomorphism of vector spaces. This implies that, if the tensor t is anti symmetric, then its conjugate pseudo tensor $\tilde{t}$ takes the same information than t itself.

Examples 6.8. **1**. As a trivial but very important example consider the **vector product** in a three-dimensional vector space V on the field $\mathbb{R}$ endowed with a metric tensor g. If $u, v \in V$ we may define the pseudo vector of order $(1, 0)$:

$$(u \times v)^r := g^{ri} \mathcal{E}_{ijk} u^j v^k .$$

If $\{e_i\}_{i=1,2,3}$ is an orthonormal basis in V, everything strongly simplifies. In fact, the Ricci tensor is represented by components

$$\mathcal{E}_{ijk} := 0$$

if (i, j, k) is not a permutation of $(1, 2, 3)$ and, otherwise,

$$\mathcal{E}_{ijk} := \pm 1 ,$$

where $+1$ corresponds to *cyclic* permutations of $1, 2, 3$ and -1 to *non-cyclic* permutations (see **Examples 3.1.2**). In such a basis:

$$(u \times v)_i = \mathcal{E}_{ijk}\, u^j v^k ,$$

because $g^{ij} = \delta^{ij}$ in each orthonormal bases. We have obtained that

$$(u \times v)_i = \mathcal{E}_{ijk} \frac{1}{2}(u^j v^k - u^k v^j) = \mathcal{E}_{ijk}(u \wedge v)^{jk} ,$$

where we have used

$$u \wedge v = u^j e_j \wedge v^k e_k = u^j v^k e_j \wedge e_k = u^j v^k \frac{1}{2!}(e_j \otimes e_k - e_k \otimes e_j) = \frac{1}{2}(u^j v^k - u^k v^j) e_j \otimes e_k .$$

The identity

$$(u \times v)_i = \mathcal{E}_{ijk}(u \wedge v)^{jk} ,$$

proves that *the vector product of two vectors is nothing but the conjugate pseudotensor of the wedge product of those vectors. It also proves that the vector product is a pseudo (covariant) vector.*

Exercises 6.9.

1. Often, the definition of vector product in $\mathbb{R}^3$ is given, in orthonormal basis, as

$$(u \wedge v)_i = \mathcal{E}_{ijk}\, u^i v^k ,$$

where it is assumed that the basis is *right oriented*. Show that it defines a proper vector (and not a pseudo vector) if a convenient definition of $\times$ is given in *left oriented* basis.

2. Is it possible to define a sort of vector product (which maps pair of vectors in vectors) in $\mathbb{R}^4$ generalizing the vector product in $\mathbb{R}^3$?

3. In physics literature one may find the statement "Differently from the impulse $\vec{p}$ which is a **polar vector**, the angular momentum $\vec{l}$ is an **axial vector**". What does it mean? (Solution. Polar vector = vector, Axial vector = pseudo vector.)

4. Consider the **parity inversion**, $P \in O(3)$, as the active transformation of vectors of physical space defined by $P := -I$ when acting in components of vectors in any orthonormal basis. What do physicists mean when saying "Axial vectors transform differently from polar vectors under parity inversion"?
(*Hint*. interpret P as a passive transformation, i.e. a changing of basis and extend the result to the active interpretation.)

5. Can the formula defining the conjugate pseudo tensor of t:

$$t^{i_1\ldots i_m} \mapsto \tilde{t}_{j_1\ldots j_{n-m}} := \mathcal{E}_{j_1\ldots j_{n-m} i_1\ldots i_m} t^{i_1\ldots i_m} ,$$

be generalized to the case where t is a pseudo tensor? If yes, what sort of geometric object is $\tilde{t}$?

6. Consider a vector product $u \times v$ in $\mathbb{R}^3$ using an orthonormal basis. In that basis there is an anti-symmetric matrix which takes the same information as $u \times v$ and can be written down using the components of the vectors u and v. Determine that matrix and explain the tensorial meaning of the matrix.

7. Prove the formula introduced above:

$$t_{i_1\ldots i_p} = \frac{|\det g|}{\det g} \frac{(-1)^{p(n-p)}}{p!(n-p)!} \mathcal{E}_{i_1\ldots i_p j_1\ldots j_m} \mathcal{E}^{j_1\ldots j_m r_1\ldots r_p}\; t_{r_1\ldots t_p} ,$$

for anti-symmetric tensors $t \in V^{*p\otimes}$, $0 \leq p \leq n = dimV$, $m := n - p$.

6.3 Tensor densities

In section 5.2 we have seen that the determinant of the matrix representing a pseudo-metric tensor g transforms, under change of basis with the rule

$$\det g' = |\det A|^{-2} \det g$$

where the pseudo-metric tensor is $g'_{ij} e'^{*i} \otimes e'^{*i} = g_{pq} e^{*p} \otimes e^{*q}$ and $A = [A^i{}_j]$ is the matrix used in the change of basis for contravariant vectors $t^i e_i = t'^p e'_p$, that is $t^i = A^i{}_p t^p$. Thus the assignment of the numbers $\det g$ for each basis in $\mathcal{B}$ does not define a scalar because of the presence of the factor $|\det A|^{-1}$. Similar mathematical objects plays a relevant role in mathematical/theoretical physics and thus deserve a precise definition.

Definition 6.10. **(Tensor densities.)** Let V be a finite-dimensional vector space with field $\mathbb{R}$ and $\mathcal{B}$ the class of all bases of V.
If $S \in \mathscr{A}_{\mathbb{R}}(V)$, a **tensor density of S with weight** $w \in \mathbb{Z} \setminus \{0\}$ is a map $d : \mathcal{B} \to S$ such that, if $B = \{e_i\}_{i\in I}$ and $B' = \{e'_j\}_{j\in I}$ are two bases in $\mathcal{B}$ with $e_k = A^i{}_k e'_i$ then

$$d(B') = |\det A|^w d(B) .$$

where $A = [A^i{}_k]$. Furthermore:
(a) the various tensorial properties enjoyed by all $d(B)$ are attributed to d. (So, for instance, if

a $d(B)$ is symmetric (and thus all $d(B)$ with $B \in \mathcal{B}$ are symmetric), d is said to be symmetric);
(b) if $B \in \mathcal{B}$, the **components of** d with respect to the canonical bases induced by B are the components of $d(B)$ in those bases. ■

If $\mathbf{g}$ is a pseudo-metric tensor on V a trivial example of a density with weight w in, for instance $S = V \otimes V^* \otimes V$, can be built up as follows. Take $t \in V \otimes V^* \otimes V$ and define

$$d_t(\{e_i\}_{i\in I}) := (\sqrt{|\det g|})^{-w} t \,,$$

where g is the matrix of the coefficients of $\mathbf{g}$ in the canonical basis associated with $\{e_i\}_{i\in I}$. In components, in the sense of (b) of the definition above:

$$(d_t)^i{}_j{}^k = (\sqrt{|\det g|})^{-w} t^i{}_j{}^k \,.$$

To conclude we give the definition of pseudo tensor density which is the straightforward extension of the definition given above.

Definition 6.11. **(Pseudo-tensor densities.)** Let V be a finite-dimensional vector space with field $\mathbb{R}$ and $\mathcal{B}$ the class of all bases of V.
If $S \in \mathscr{A}_{\mathbb{R}}(V)$, a **pseudo-tensor density of** S **with weight** $w \in \mathbb{Z} \setminus \{0\}$ is a map $d : \mathcal{B} \to S$ such that, if $B = \{e_i\}_{i\in I}$ and $B' = \{e'_j\}_{j\in I}$ are two bases in $\mathcal{B}$ with $e_k = A^i{}_k e'_i$ then

$$d(B') = \frac{\det A}{|\det A|} |\det A|^w d(B) \,.$$

where $A = [A^i{}_k]$. Furthermore:
(a) the various tensorial properties enjoyed by all $d(B)$ are attributed to d. (So, for instance, if a $d(B)$ is symmetric (and thus all $d(B)$ with $B \in \mathcal{B}$ are symmetric), d is said to be symmetric);
(b) if $B \in \mathcal{B}$, the **components of** d with respect to the canonical bases induced by B are the components of $d(B)$ in those bases. ■

Remarks 6.12. It is obvious that the sets of tensor densities and pseudo-tensor densities of a fixed space S and with fixed weight form linear spaces with composition rule which reduces to usual linear composition rule of components. ■

There is an important property of densities with weight -1 which is very useful in integration theory on manifolds. The property is stated in the following theorem.

Theorem 6.13. *Let V be a vector space on $\mathbb{R}$ with dimension $n < +\infty$. There is a natural isomorphism G from the space of scalar densities of weight -1 and the space of antisymmetric covariant tensors of order n. In components, using notation as in Definition 5.3,*

$$G : \alpha \mapsto \alpha \eta_{i_1 \ldots i_n} \,.$$

Proof. Fix a canonical basis of $V^{*n\otimes}$ associated with a basis of V. Any non-vanishing tensor t in space, $\Lambda^n(V^*)$, of antisymmetric covariant tensors of order $n = dimV$ must have the form $t_{i_1 \ldots i_n} = \alpha \eta_{i_1 \ldots i_n}$ in components because different non-vanishing components can be differ only for the sign due to antisymmetry properties of t. Therefore the dimension of $\Lambda_n(V)$ is 1 which is also the dimension of the space of scalar densities of weight -1. The application G defined above in components from $\mathbb{R}$ to $\Lambda^n(V^*)$ is linear and surjective and thus is injective. Finally, re-adapting straightforwardly the relevant part of the discussion used to define ϵ, one finds that the coefficient α in $t_{i_1 \ldots i_n} = \alpha \eta_{i_1 \ldots i_n}$ transforms as a scalar densities of weight -1 under change of basis. $\square$

Chapter 7

Polar Decomposition Theorem in the finite-dimensional case

The goal of this chapter is to introduce the polar decomposition theorem of operators in finite dimensional spaces which has many applications in mathematics and physics. We consider here the finite dimensinal case only (see, e.g., [Rudin, Moretti-a] for the case of an infinite dimensional Hilbert space)

7.1 Operators in spaces with scalar product

We remind here some basic definitions and results which should be known by the reader from elementary courses of linear algebra [Lang, Sernesi].

If $A \in \mathscr{L}(V|V)$, that is $A : V \to V$ is a (linear) operator on any finite-dimensional vector space V with field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, an **eigenvalue** of A is an element $\lambda \in \mathbb{K}$ such that

$$(A - \lambda I)u = 0$$

for some $u \in V \setminus \{0\}$. In that case u is called **eigenvector** associated with λ. The set $\sigma(A)$ containing all of the eigenvalues of A is called the **spectrum** of A. The **eigenspace** E_λ associated with $\lambda \in \sigma(A)$ is the subspace of V spanned by the eigenvectors associated with λ.

Proposition 7.1. *Let V be a real (complex) finite-dimensional vector space equipped with a (resp. Hermitian) scalar product $(\cdot|\cdot)$. For every operator $A : V \to V$ there exists exactly one of operator $A^\dagger : V \to V$, called the* **adjoint operator** *of A, such that*

$$(A^\dagger u|v) = (u|Av) ,$$

for all $u, v \in V$.

Proof. Fix $u \in V$, the map $v \mapsto (u|Av)$ is a linear functional and thus an element of V^*. By theorem 5.7 there is a unique element $w_{u,A} \in V$ such that $(u|Av) = (w_{u,A}|v)$ for all $v \in V$. Consider the map $u \mapsto w_{u,A}$. It holds, if a, b are scalars in the field of V and $u, u' \in V$

$$(w_{au+bu',A}|v) = (au+bu'|Av) = \overline{a}(u|Av)+\overline{b}(u'|Av) = \overline{a}(w_{u,A}|v)+\overline{b}(w_{u',A}|v) = (aw_{u,A}+bw_{u',A}|v).$$

Hence, for all $v \in V$:

$$(w_{au+bu',A} - aw_{u,A} - bw_{u',A}|v) = 0\,,$$

The scalar product is non-degenerate by definition and this implies

$$w_{au+bu',A} = aw_{u,A} + bw_{u',A}\,.$$

We have obtained that the map $A^\dagger : u \mapsto w_{u,A}$ is linear, in other words it is an operator. The uniqueness is trivially proved: if the operator B satisfies $(Bu|v) = (u|Av)$ for all $u, v \in V$, it must holds $((B - A^\dagger)u|v) = 0$ for all $u, v \in V$ which, exactly as we obtained above, entails $(B - A^\dagger)u = 0$ for all $u \in V$. In other words $B = A^\dagger$. □

Remarks 7.2. There is an interesting interplay between this notion of adjoint operator and that given in definition 2.14. Let $h : V \to V^*$ be the natural (anti)isomorphism induced by the scalar product as seen in theorem 5.7. If $T \in \mathscr{L}(V|V)$ and $T^*, T^\dagger$ denotes, respectively the two adjoint operators of T defined in definition 2.14 and in proposition 7.1, one has:

$$h \circ T^\dagger = T^* \circ h\,. \tag{7.1}$$

The proof of this fact is an immediate consequence of (3) in remarks 5.5. ■

There are a few simple properties of the adjoint operator whose proofs are straightforward. Below A, B are operators in a complex (resp. real) finite-dimensional vector space V equipped with a Hermitian (resp. real) scalar product $(\cdot|\cdot)$ and a, b belong to the field of V.
(1) $(A^\dagger)^\dagger = A$,
(2) $(aA + bB)^\dagger = \overline{a}A^\dagger + \overline{b}B^\dagger$ (resp. $(aA + bB)^\dagger = aA^\dagger + bB^\dagger$),
(3) $(AB)^\dagger = B^\dagger A^\dagger$,
(4) $(A^\dagger)^{-1} = (A^{-1})^\dagger$ (if A^{-1} exists).

Take a finite-dimensional vector space V equipped with a scalar product $(\cdot|\cdot)$ and consider $A \in \mathscr{L}(V|V)$. We consider two cases: the *real case* where the field of V and the scalar product are real, and the *complex case*, where the field of V is $\mathbb{C}$ and the scalar product is Hermitian. We have the following definitions.
(1) In the real case, A is said to be **symmetric** if $(Au|v) = (u|Av)$ for all $u, v \in V$. In the complex case, A is said to be **Hermitian** if $(Au|v) = (u|Av)$ for all $u, v \in V$. It is simply proved that, in the respective cases, A is symmetric or Hermitian if and only if $A = A^\dagger$. In both cases $\sigma(A) \subset \mathbb{R}$.

(2) In the real case, A is said to be **orthogonal** if $(Au|Av) = (u|v)$ for all $u, v \in V$. In the complex case, A is said to be **unitary** if $(Au|Av) = (u|v)$ for all $u, v \in V$. It is simply proved that, in the respective cases, A is orthogonal or unitary if and only if A is bijective and $A^{-1} = A^{\dagger}$. In both cases if $\lambda \in \sigma(A)$ then $|\lambda| = 1$.
(3) In both cases A is said to be **normal** if $AA^{\dagger} = A^{\dagger}A$. It is obvious that symmetric, Hermitian, unitary, orthogonal operators are normal.
An important and straightforwardly-proved result is that, if $A \in \mathscr{L}(V|V)$ is normal, then:
(a) $u \in V$ is eigenvector of A with eigenvalue λ if and only if the same u is eigenvector of $A^{\dagger}$ with eigenvalue $\overline{\lambda}$ (which coincides with λ in the real case);
(b) If $\lambda, \mu \in \sigma(A)$ and $\lambda \neq \mu$ then $(u_\lambda|u_\mu) = 0$ if u_λ and u_μ are eigenvectors with eigenvalue λ and μ respectively for A.

In either the real or complex case, if $U \subset V$ is a subspace, the **orthogonal** of U, $U^{\perp}$, is the subspace of V made of all the vectors which are orthogonal to U, i.e., $v \in U^{\perp}$ if and only if $(u|v) = 0$ for all $u \in U$. As is well-known (see [Lang, Sernesi]), if $w \in V$, the decomposition $w = u + v$ with $u \in U$ and $v \in U^{\perp}$ is uniquely determined, and the map $P_U : w \mapsto u$ is linear and it is called **orthogonal projector** onto U.
If V is a finite-dimensional vector space (either real or complex), and $U \subset V$ is a subspace, the subspace $U^*_{\perp} \subset V^*$ defined as

$$U^*_{\perp} := \{v \in V^* \;\; | \;\; \langle u, v\rangle = 0\,, \quad \text{for all } u \in U\}$$

gives another and more general notion of "orthogonal space", in the absence of a scalar product in V. It is simply proved that, in the presence od a (pseudo)scalar product and where $h : V \to V^*$ is the natural (anti)isomorphism induced by the scalar product as seen in theorem 5.7, one has

$$h^{-1}(U^*_{\perp}) = U^{\perp}\,.$$

We leave to the reader the simple proof of the fact that an operator $P : V \to V$ is an orthogonal projector onto some subspace $U \subset V$ if and only if both the conditions below hold
(1) $PP = P$,
(2) $P = P^{\dagger}$.
In that case P is the orthogonal projector onto $U = \{Pv \mid v \in V\}$.
Another pair of useful results concerning orthogonal projectors is the following. Let V be a space as said above, let U, U' be subspaces of V, with P, P' are the corresponding orthogonal projectors P, P'.
(a) U and U' are **orthogonal** to each other, i.e., $U' \subset U^{\perp}$ (which is equivalent to $U \subset U'^{\perp}$) if and only if $PP' = P'P = 0$.
(b) $U \subset U'$ if and only if $PP' = P'P = P$.

If V is as above and it has finite dimension n and $A : V \to V$ is normal, there exist a well-known spectral decomposition theorem [Lang, Sernesi]: (the finite-dimensional version of the "spectral theorem").

Proposition 7.3. (**Spectral decomposition for normal operators in complex spaces**.) *Let V be a complex finite-dimensional vector space equipped with a Hermitian scalar product $(\cdot|\cdot)$. If $A \in \mathscr{L}(V|V)$ is normal (i.e., $A^\dagger A = AA^\dagger$), the following expansion holds:*

$$A = \sum_{\lambda \in \sigma(A)} \lambda P_\lambda \,,$$

where P_λ is the orthogonal projector onto the eigenspace associated with λ. Moreover the map $\sigma(A) \ni \lambda \mapsto P_\lambda$ satisfies the following two properties:
(1) *$I = \sum_\lambda P_\lambda$,*
(2) *$P_\lambda P_\mu = P_\mu P_\lambda = 0$ for $\mu \neq \lambda$.*
A **spectral measure***, i.e. a map $B \ni \mu \mapsto P'_\mu$ with $B \subset \mathbb{C}$ finite, P'_μ non-vanishing orthogonal projectors and:*
(1)' *$I = \sum_{\mu \in B} P'_\mu$,*
(2)' *$P'_\lambda P'_\mu = P'_\mu P'_\lambda = 0$ for $\mu \neq \lambda$,*
coincides with $\sigma(A) \ni \lambda \mapsto P_\lambda$ if
(3)' *$A = \sum_{\mu \in B} \mu P'_\mu$.*

Proof. The equation $\det(A - \lambda I) = 0$ must have a (generally complex) solution λ due to fundamental theorem of algebra. As a consequence, A admits an eigenspace $E_{\lambda_1} \subset V$. From the properties (a) and (b) of normal operators one obtains that $A(E_\lambda^\perp) \subset E_{\lambda_1}^\perp$. Moreover $A \restriction_{E_{\lambda_1}^\perp}$ is normal. Therefore the procedure can be iterated obtaining a new eigenspace $E_{\lambda_2} \subset E_{\lambda_1}^\perp$. Notice that $\lambda_2 \neq \lambda_1$ because every eigenvector of $A \restriction_{E_{\lambda_1}^\perp}$ is also eigenvector for A with the same eigenvalue. Since eigenspaces with different eigenvalues are orthogonal, one finally gets a sequence of pairwise orthogonal eigenspaces $E_{\lambda_1}, E_{\lambda_2}, \ldots, E_{\lambda_k}$, with $E_{\lambda_l} \subset E_{\lambda'_l}^\perp$ if $l' < l$. This sequence must be finite because the dimension of V is finite and orthogonal (eigen)vectors are linearly independent. Therefore $E_{\lambda_k}^\perp = \{0\}$ and so $\oplus_{l=1}^k E_{\lambda_l} = V$. On the other hand, the set $\{\lambda_1, \lambda_2, \ldots \lambda_k\}$ must coincide with the whole $\sigma(A)$ because eigenspaces with different eigenvalues are orthogonal (and thus any other eigenspace E_{λ_0} with $\lambda_0 \notin \{\lambda_1, \lambda_2, \ldots \lambda_k\}$ would be included in $(\oplus_{l=1}^k E_{\lambda_l})^\perp = \{0\}$). If $\{e_j^{(l)}\}_{j \in I_l}$ is an orthonormal basis for E_{λ_l}, the orthogonal projector on E_{λ_l} is

$$P_\lambda := \sum_{j \in I_l} (e_j^{(l)} | \quad) e_j^{(l)} \,.$$

Property (1) is nothing but the fact that $\{e_j^{(l)}\}_{j \in I_l, l=1,\ldots,k}$ is an orthonormal basis for V. Property (2) is another way to say that eigenspaces with different eigenvalues are orthogonal. Finally, since $e_j^{(l)}$ is an eigenvector of A with eigenvalue λ_l, it holds

$$A = \sum_{j \in I_l, l=1,\ldots,k} \lambda_l (e_j^{(l)} | \quad) e_j^{(l)} \,,$$

which can be re-written as:

$$A = \sum_{\lambda \in \sigma(A)} \lambda P_\lambda \,.$$

The uniqueness property of spectral measures is proved as follows. It holds

$$\sum_{\mu \in B} \mu P'_\mu = \sum_{\lambda \in \sigma(A)} \lambda P_\lambda \,.$$

Applying P'_ν on the left and P_τ on the right, and using properties (1)' and (1) one finds

$$\nu P'_\nu P_\tau = \tau P'_\nu P_\tau \,,$$

so that

$$(\nu - \tau) P'_\nu P_\tau = 0 \,.$$

This is possible if $\nu = \tau$ or $P'_\nu P_\tau = 0$. For a fixed P'_ν it is not possible that $P'_\nu P_\tau = 0$ for all P_τ, because, using (1), it would imply

$$0 = P'_\nu \sum_{\tau \in \sigma(A)} P_\tau = P'_\nu I = P'_\nu$$

but $P'_\nu \neq 0$ hypotheses. Therefore $\nu = \tau$ for some $\tau \in \sigma(A)$ and hence $B \subset \sigma(A)$. By means of the analogous procedure one sees that $\sigma(A) \subset B$. We conclude that $B = \sigma(A)$. Finally, the decomposition:

$$A = \sum_{\lambda \in \sigma(A)} \lambda P'_\lambda$$

implies that each P'_λ is an orthogonal projector on a subspace of E_λ. As a consequence, if u_λ is an eigenvector of A with eigenvalue λ, it must be $P_{\lambda'} u_\lambda = 0$ for $\lambda' \neq \lambda$. Since (1)' is valid, one concludes that $P'_\lambda u_\lambda = u_\lambda$ and thus not only $P'_\lambda(V) \subset E_\lambda$, but, more strongly, $P'_\lambda(V) = E_\lambda$. Since the orthogonal projector onto a subspace is biunivocally determined by the subspace itself, we have proved that $P'_\lambda = P_\lambda$ for every $\lambda \in \sigma(A)$. $\square$

7.1.1 Positive operators

Definition 7.4. If V is a real (complex) vector space equipped with a (resp. Hermitian) scalar product $(\cdot|\cdot)$, an operator $A : V \to V$ is said to be **positive** (or *positive semidefinite*) if

$$(u|Au) \geq 0 \quad \text{for all } u \in V \,.$$

A positive operator A is said to be **strictly positive** (or *positive definite*) if

$$(u|Au) = 0 \quad \text{entails } u = 0 \,.$$

$\blacksquare$

A straightforward consequence of the given definition is the following lemma.

Lemma 7.5. *Let V be a complex vector space equipped with a Hermitian scalar product $(\cdot|\cdot)$. Any positive operator $A : V \to V$, is Hermitian.*
Moreover, if $dimV < \infty$, a normal operator $A : V \mapsto V$
(a) *is positive if and only if $\sigma(A) \subset [0, +\infty)$;*
(b) *is strictly positive if and only if $\sigma(A) \subset (0, +\infty)$.*

Proof. As $(v|Av) \geq 0$, by complex conjugation $(Av|v) = \overline{(v|Av)} = (v|Av)$ and thus

$$((A^\dagger - A)v|v) = 0$$

for all $v \in V$. In general we have:

$$2(Bu|w) = (B(u+w)|(u+w)) + i(B(w+iu)|(w+iu)) - (1+i)(Bw|w) - (1+i)(Bu|u) \,.$$

So that, taking $B = A^\dagger - A$ we get $(Bu|w) = 0$ for all $u, w \in V$ because $(Bv|v) = 0$ for all $v \in V$. $(Bu|w) = 0$ for all $u, w \in V$ entails $B = 0$ or $A^\dagger = A$.
Let us prove (a). Suppose A is positive. We know that $\sigma(A) \subset \mathbb{R}$. Suppose there is $\lambda < 0$ in $\sigma(A)$. Let u be an eigenvector associated with λ. $(u|Au) = \lambda(u|u) < 0$ because $(u|u) > 0$ since $u \neq 0$. This is impossible.
Now assume that A is normal with $\sigma(A) \subset [0, +\infty)$. By proposition 7.3:

$$(u|Au) = \left(\sum_\mu P_\mu u \middle| \sum_\lambda \lambda P_\lambda \sum_\nu P_\nu u \right) = \sum_{\mu,\lambda,\nu} \lambda \left(P_\nu^\dagger P_\lambda^\dagger P_\mu u \middle| u \right) = \sum_{\mu,\lambda,\nu} \lambda \left(P_\nu P_\lambda P_\mu u \middle| u \right)$$

because, if P is an orthogonal projector, $P = P^\dagger$. Using Proposition A.2 once again, $P_\nu P_\lambda P_\mu = \delta_{\nu\mu}\delta_{\mu\lambda}P_\lambda$ and thus

$$(u|Au) = \sum_\lambda \lambda \left(P_\lambda u \middle| u \right) = \sum_\lambda \lambda \left(P_\lambda u \middle| P_\lambda u \right) \,,$$

where we have used the property of orthogonal projectors $PP = P$. $\lambda \left(P_\lambda u \middle| P_\lambda u \right) \geq 0$ if $\lambda \geq 0$ and thus $(u|Au) \geq 0$ for all $u \in V$.
Concerning (b), assume that A is strictly positive (so it is positive and Hermitian). If $0 \in \sigma(A)$ there must exist $u \neq 0$ with $Au = 0u = 0$. That entails $(u, Au) = (u, 0) = 0$ which is not allowed. Therefore $\sigma(A) \subset [0, +\infty)$. Conversely if A is normal with $\sigma(A) \subset (0, +\infty)$, A is positive by (a). If A is not strictly positive, there is $u \neq 0$ such that $(u|Au) = 0$ and thus, using the same procedure as above,

$$(u|Au) = \sum_\lambda \lambda \left(P_\lambda u \middle| P_\lambda u \right) = 0 \,.$$

Since $\lambda > 0$ and $\left(P_\lambda u \middle| P_\lambda u \right) \geq 0$, it must be $\left(P_\lambda u \middle| P_\lambda u \right) = 0$ for all $\lambda \in \sigma(A)$. This means $P_\lambda u = 0$ for all $\lambda \in \sigma(A)$. This is not possible because, using (1) in proposition 7.3, $0 \neq u = Iu = \sum_\lambda P_\lambda u = 0$. □

7.1.2 Complexification procedure.

There is a standard and useful way to associate a complex vector space endowed with Hermitian scalar product to a given real vector space equipped with a scalar product. This correspondence allows one to take advantage of some results valid for complex vector spaces also in the case of real vector spaces. We shall employ that possibility shortly.

Definition 7.6. If V is a real vector space equipped with the scalar product $(\cdot|\cdot)$, the **complexification** of V, $V \oplus iV$, is the complex vector space with Hermitian scalar product $(\cdot|\cdot)_{\mathbb{C}}$ defined as follows. The elements of $V \oplus iV$, are the corresponding pairs $(u, v) \in V \times V$ denotes with $u + iv := (u, v)$ moreover,
(1) the product of scalar in $\mathbb{C}$ and vectors in $V \oplus iV$ is defined as:

$$(a + ib)(u + iv) := au - bv + i(bu + av) \quad \text{for } a + ib \in \mathbb{C} \text{ and } u + iv \in V \times V ,$$

(2) the sum of vectors is defined as:

$$(u + iv) + (x + iy) := (u + x) + i(v + y) , \quad \text{for } u + iv, x + iy \in V \times V ,$$

(3) the Hermitian scalar product is defined as:

$$(u + iv|w + ix)_{\mathbb{C}} := (u|v) + (v|x) + i(u|x) - i(v|w) , \quad \text{for } u + iv, w + ix \in V \times V .$$

■

Let us introduce a pair of useful operators from $V \oplus iV$ to $V \oplus iV$.
The **complex conjugation** is defined as

$$J : u + iv \mapsto u - iv .$$

It is anti linear and satisfies $(J(u + iv)|J(w + ix))_{\mathbb{C}} = (w + ix|u + iv)_{\mathbb{C}}$ and $JJ = I$.
An important property of J is the following: $s \in V \oplus iV$ satisfies $s \in V$ (i.e. $s = (u, 0) \in V \times V$) if and only if $Js = s$, whereas it satisfies $s \in iV$ (i.e. $s = (0, v) \in V \times V$) if and only if $Js = -s$.
The proofs of these features follows trivially by the definition.
The second interesting operator is the **flip operator**, defined as

$$C : u + iv \mapsto v - iu .$$

It satisfies $CC = I$ and $C^\dagger = C$, where the adjoint is refereed to the Hermitian scalar product.
Also in this case the proof is straightforward using the given definition.
A linear operator $A : V \oplus iV \to V \oplus iV$ is said to be **real** if $JA = AJ$.

Proposition 7.7. *Referring to definition 7.6 let $A \in \mathscr{L}(V \oplus iV|V \oplus iV)$. The following facts hold.*

(a) *A is real if and only if there is a uniquely-determined pair of $\mathbb{R}$-linear operators $A_j : V \to V$, $j = 1, 2$, such that*

$$A(u + iv) = A_1 u + iA_2 v\ , \quad \textit{for all } u + iv \in V \oplus iV\ .$$

(b) *A is real and $AC = CA$, if and only if there is a uniquely-determined $\mathbb{R}$-linear operator $A_0 : V \to V$, such that*

$$A(u + iv) = A_0 u + iA_0 v\ , \quad \textit{for all } u + iv \in V \oplus iV\ .$$

Proof. The proof of uniqueness is trivial in both cases and it is based on linearity and on the following fact. If $T, T' \in \mathscr{L}(V|V)$, $Tu + iT'v = 0$ for all $u, v \in V$ entails $T = T' = 0$. Moreover, by direct inspection one sees that A is real if $A(u + iv) = A_1 u + iA_2 v$ for all $u + iv \in V \oplus iV$ and A is real and commute with C if $A(u+iv) = A_0 u + iA_0 v$ for all $u + iv \in V \oplus iV$, where the operators A_j are $\mathbb{R}$-linear operators from V to V.
Let us conclude the proof of (a) proving that if A is real, there must be A_1 and A_2 as in (a). Take $s \in V$, as a consequence of $AJ = JA$ one has: $JAs = AJs = As$, where we used the fact that $Js = s$ if and only if $s \in V$. Since $JAs = s$ we conclude that $As \in V$ if $s \in V$ and so $A \restriction_V$ is well defined as $\mathbb{R}$-linear operator from V to V. With an analogous procedure one sees that $A \restriction_{iV}$ is well defined as $\mathbb{R}$-linear operator from iV to iV. Defining $A_1 := A \restriction_V : V \to V$ and $A_2 := -iA \restriction_{iV} i : V \to V$ one has:

$$A(u + iv) = A \restriction_V u + A \restriction_{iV} iv = A_1 u + iA_2 v\ .$$

We have proved that (a) is true.
Let us conclude the proof of (b). If A is real and commute with C, by (a) there must be A_1 and A_2 satisfying the condition in (a) and, furthermore, the condition $AC = CA$ reads $A_1 v = A_2 v$ and $A_2 u = A_1 v$ for all $u, v \in V$. Therefore $A_0 := A_1 = A_2$ verifies (b). $\Box$

7.1.3 Square roots

Definition 7.8. If V is a complex (real) vector space equipped with a Hermitian (respectively real) scalar product $(\cdot|\cdot)$, Let $A : V \to V$ a (symmetric in the real case) positive operator. If $B : V \to V$ is another (symmetric in the real case) positive operator such that $B^2 = A$, B is called a **square root of** A. $\blacksquare$

Notice that square roots, if they exist, are Hermitian by lemma 7.5. The next theorem proves that the square root of a positive operator exist and is uniquely determined.

Theorem 7.9. *Let V be a finite-dimensional either complex or real vector space equipped with a Hermitian or, respectively, real scalar product $(\cdot|\cdot)$. Every positive operator $A : V \to V$ admits a unique square root indicated by $\sqrt{A}$. $\sqrt{A}$ is Hermitian or, respectively, symmetric and $\sqrt{A}$ is bijective if and only if A is bijective.*

Proof. First consider the complex case. A is Hermitian by lemma 7.5. Using proposition 7.3,

$$A = \sum_{\lambda \in \sigma(A)} \lambda P_\lambda \,.$$

Since $\lambda \geq 0$ we can define the linear operator

$$\sqrt{A} := \sum_{\lambda \in \sigma(A)} \sqrt{\lambda} P_\lambda \,.$$

By proposition 7.3 we have

$$\sqrt{A}\sqrt{A} = \sum_{\lambda \in \sigma(A)} \sqrt{\lambda} P_\lambda \sum_{\mu \in \sigma(A)} \sqrt{\mu} P_\mu = \sum_{\lambda\mu} \sqrt{\lambda\mu} P_\lambda P_\mu \,.$$

Using property (2)

$$\sqrt{A}\sqrt{A} = \sum_{\lambda\mu} \sqrt{\lambda\mu}\delta_{\mu\nu} P_\lambda = \sum_{\lambda} (\sqrt{\lambda})^2 P_\lambda = \sum_{\lambda} \lambda P_\lambda = A \,.$$

Notice that $\sqrt{A}$ is Hermitian by construction:

$$\sqrt{A}^\dagger := \sum_{\lambda \in \sigma(A)} \sqrt{\lambda} P_\lambda^\dagger = \sum_{\lambda \in \sigma(A)} \sqrt{\lambda} P_\lambda = \sqrt{A} \,.$$

Moreover, if $B = \{\mu = \sqrt{\lambda} \ |\lambda \in \sigma(A)\}$ and $P'_\mu := P_\lambda$ with $\mu = \sqrt{\lambda}$, it holds

$$\sqrt{A} := \sum_{\mu \in B} \mu P'_\mu \,,$$

and $B \ni \mu \mapsto P'_\mu$ satisfy the properties (1)',(2)',(3)' in proposition 7.3 As a consequence it coincides with the spectral measure associated with $\sqrt{A}$,

$$\sqrt{A} := \sum_{\lambda \in \sigma(A)} \sqrt{\lambda} P_\lambda$$

is the unique spectral decomposition of A,

$$\sigma(\sqrt{A}) = \{\mu = \sqrt{\lambda} \ |\lambda \in \sigma(A)\} \,,$$

and thus $\sqrt{A}$ is positive by lemma 7.5, $\sqrt{A}$ is a Hermitian square root of A. Notice that, by construction $0 \in \sigma(\sqrt{A})$ if and only if $0 \in \sigma(A)$ so that $\sqrt{A}$ is bijective if and only if A is bijective.

Let us pass to prove the uniqueness property. Suppose there is another square root S of A. As S is positive, it is Hermitian with $\sigma(S) \subset [0, +\infty)$ and it admits a (unique) spectral decomposition

$$S = \sum_{\nu \in \sigma(S)} \nu P'_\mu \,.$$

Define $B := \{\nu^2 \mid \nu \in \sigma(S)\}$. It is simply proved that the map $B \ni \lambda \mapsto P'_{\sqrt{\lambda}}$ satisfy:
(1)' $I = \sum_{\lambda \in B} P'_{\sqrt{\lambda}}$,
(2)' $P'_{\sqrt{\lambda}} P'_{\sqrt{\mu}} = P'_{\sqrt{\mu}} P'_{\sqrt{\lambda}} = 0$ for $\mu \neq \lambda$,
(3)' $A = S^2 = \sum_{\lambda \in B} \lambda P'_{\sqrt{\lambda}}$.
proposition 7.3 entails that the spectral measure of A and $B \ni \lambda \mapsto P'_{\sqrt{\lambda}}$ coincides: $P'_{\sqrt{\lambda}} = P_\lambda$ for all $\lambda \in \sigma(A) = B$. In other words

$$S = \sum_{\nu \in \sigma(S)} \nu P'_\nu = \sum_{\lambda \in \sigma(A)} \lambda P_\lambda = \sqrt{A}\,.$$

Let us finally consider the case of a real space V. If $A \in \mathscr{L}(V|V)$ is positive and symmetric, the operator on $V \oplus iV$, $S : u + iv \mapsto Au + iAv$ is positive and Hermitian. By the first part of this proof, there is only one Hermitian positive operator $B \in \mathscr{L}(V \oplus iV|V \oplus iV)$ with and $B^2 = S$, that is the square root of S which we indicate by $\sqrt{S}$. Since S commutes with both J and C and $CC = JJ = I$, one has

$$\sum_{\lambda \in \sigma(S)} \lambda J P^{(S)}_\lambda J = S\,, \quad \text{and} \sum_{\lambda \in \sigma(S)} \lambda C P^{(S)}_\lambda C = S\,.$$

Since J and C (anti-) preserve the scalar product and $JJ = CC = I$, one straightforwardly proves that $\sigma(S) \ni \lambda \mapsto JP^{(S)}_\lambda J$ and $\sigma(S) \ni \lambda \mapsto JP^{(S)}_\lambda J$ are spectral measures. By the uniqueness property in proposition 7.3 one concludes that these spectral measures coincide with $\sigma(S) \ni \lambda \mapsto P^{(S)}_\lambda$ and thus, in particular, each projector of the spectral measure commutes with both J and C. Hence $\sqrt{S} = \sum_{\lambda \in \sigma(S)} \sqrt{\lambda} P^{(S)}_\lambda$ commutes with both J and C. We conclude that $\sqrt{S}$ is real with the form $\sqrt{S} : u + iv \mapsto Ru + iRv$. The operator $\sqrt{A} := R$ fulfills all of requirements of a square root it being symmetric and positive because $\sqrt{S}$ is Hermitian and positive, and $R^2 = A$ since $(\sqrt{S})^2 = S : u + iv \mapsto Au + iAv$. If A is bijective, S is such by construction and thus its kernel is trivial. Since $\sqrt{S} = \sum_{\lambda \in \sigma(S)} \sqrt{\lambda} P^{(S)}_\lambda$, its kernel is trivial too and $\sqrt{S}$ is bijective. In turn, R is bijective by construction. If A is not bijective, and $0 \neq u \in KerA$, $u + iu \in KerR$ so that R is not bijective.
Let us consider the uniqueness of the found square root. If R' is another positive symmetric square root of A, $B : u + iv \to R'u + iR'v$ is a Hermitian positive square root of S and thus it must coincide with $\sqrt{S}$. This implies that $R = R'$. □

7.2 Polar Decomposition

The notions and the results obtained above allow us to state and prove the polar decomposition theorem for operators in finite dimensional vector spaces equipped with scalar product.

Theorem 7.10. (**Polar Decomposition of operators**.) *If $T \in \mathscr{L}(V|V)$ is a bijective operator where V is a real (resp. complex), finite-dimensional vector space equipped with a real*

(resp. Hermitian) scalar product space:
(a) *there is a unique decomposition $T = UP$, where U is orthogonal (resp. unitary) and P is bijective, symmetric (resp. Hermitian), and positive. In particular $P = \sqrt{T^\dagger T}$ and $U = T(\sqrt{T^\dagger T})^{-1}$;*
(b) *there is a unique decomposition $T = P'U'$, where U' is orthogonal (resp. unitary) and P' is bijective, symmetric (resp. Hermitian), and positive. In particular $U' = U$ and $P' = UPU^\dagger$.*

Proof. (a) Consider $T \in \mathscr{L}(V|V)$ bijective. $T^\dagger T$ is symmetric/self-adjoint, positive and bijective by construction. Define $P := \sqrt{T^\dagger T}$, which exists and is symmetric/self-adjoint, positive and bijective by theorem 7.9, and $U := TP^{-1}$. U is orthogonal/unitary because

$$U^\dagger U = P^{-1}T^\dagger TP^{-1} = P^{-1}P^2P^{-1} = I \,,$$

where we have used $P^\dagger = P$. This proves that a polar decomposition of T exists because $UP = T$ by construction. Let us pass to prove the uniqueness of the decomposition. If $T = U_1P_1$ is a other polar decomposition, $T^\dagger T = P_1U_1^\dagger U_1P_1 = PU^\dagger UP$. That is $P_1^2 = P^2$. Theorem 7.9 implies that $P = P_1$ and $U = T^{-1}P = T^{-1}P_1 = U_1$.
(b) $P' := UPU\dagger$ is symmetric/self-adjoint, positive and bijective since $U^\dagger$ is orthogonal/unitary and $P'U' = UPU^\dagger U = UP = T$. The uniqueness of the decomposition in (b) is equivalent to the uniqueness of the polar decomposition $U'^\dagger P'^\dagger = T^\dagger$ of $T^\dagger$ which holds true by (a) replacing T by $T^\dagger$. $\square$

Chapter 8

Special Relativity: a Geometric Presentation

This chapter is a pedagogical review on basic geometric notions of Special Relativity theory. We employ definitions and results presented in Section 3.4. The adopted approach is completely assiomatic and no physical discussion to justify the used geometrical apparatus form experimental facts is provided. A physically constructive approach can be found in Chapter 15 of [Moretti-c] (and in [Moretti-d] in Italian).

8.1 Minkowski spacetime

We henceforth adopt again the notation ∂_{x^a} to indicate the vector e_a of a basis of the space of translations V of an affine space $\mathbb{A}^n$ used to construct a Cartesian coordinate system $x^1, \ldots, x^n$ over $\mathbb{A}^n$ upon the choice of the origin $O \in \mathbb{A}^n$. Moreover, as before, dx^a denotes the corresponding element of the dual basis in V^*. Finally

$$T_p\mathbb{A}^n := \{(u,p)\}_{u\in V} \quad \text{and} \quad T^*_p\mathbb{A}^n := \{(v^*,p)\}_{v^*\in V^*}$$

respectively indicate the vector spaces of vectors *applied to* p, and covectors *applied to* p in the sense of Section 3.4. With these conventions $\partial_{x^i}|_p := (\partial_{x^i}, p)$ and $dx^i|_p := (dx^i, p)$. We shall also use the short $v_p := (v,p) \in T_p\mathbb{A}^n$.
A *vector field* X on $\mathbb{A}^n$ is an assigment $\mathbb{A}^n \ni p \mapsto X_p \in T_p\mathbb{A}^n$. X is said to be *constant* if $X_p = X_q$ for all $p, q \in \mathbb{A}^n$. The vector field is said to be *of class* C^k if given a basis of V (and thus of $T_p\mathbb{A}^n$) the components of the applied vector X_p are C^k functions ($k = 0, 1, \ldots, \infty$) of a Cartesian coordinates system constructed out of that basis. If this fact is true for a Cartesian coordinate system, it is true for every other Cartesian coordinate system, since the transformations between Cartesian coordinates systems are C^∞ with C^∞ inverse.

8.1.1 General Minkowski spacetime structure

Let us remind some features of the relativistic spacetime from a mathematical point of view. We adopt in the following notations and conventions for tensors on affine spaces as in Section 3.4.

Definition 8.1. Minkowski spacetime $\mathbb{M}^4$ is a four-dimensional affine space whose real vector space of translation T^4 (identified to $\mathbb{R}^4$) is equipped with a pseudo-scalar product defined by a metric tensor $\mathbf{g}$, with signature $(1,3)$ (i.e. $(-1,+1,+1,+1)$). The following further definitions hold.

(i) The points of Minkowski spacetime are called **events**.

(ii) The Cartesian coordinate systems x^0, x^1, x^2, x^3 induced from the affine structure by arbitrarily fixing any pseudo-orthonormal canonical basis $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ in T^4 (with $\mathbf{g}(\mathbf{e}_0, \mathbf{e}_0) = -1$, $\mathbf{g}(\mathbf{e}_1, \mathbf{e}_1) = 1$, $\mathbf{g}(\mathbf{e}_2, \mathbf{e}_2) = 1$, $\mathbf{g}(\mathbf{e}_3, \mathbf{e}_3) = 1$) and any origin $O \in \mathbb{M}^4$, are called **Minkowskian coordinate systems** or **Minkowskian coordinate systems frames**.■

Remarks 8.2. The elements of T^4 (as well as the vectors applied at every event) are often called *four-vectors* in the literature. ■

In practice, exploiting the affine structure and using standard notation for affine spaces, Minkowskian coordinates x^0, x^1, x^2, x^3 are defined in order that the map $\mathbb{M}^4 \ni p \mapsto (x^0(p), x^1(p), x^2(p), x^3(p)) \in \mathbb{R}^4$ satisfies (where we are using the convention of summation over repeated indices):

$$T^4 \ni p - O = x^a(p)\mathbf{e}_a$$

for every event $p \in \mathbb{M}^4$.
The pseudo-scalar product $(\cdot|\cdot)$, that is the metric tensor $\mathbf{g}$ in T^4, has form constant and diagonal in Minkowskian coordinates:

$$\mathbf{g} = -dx^0 \otimes dx^0 + dx^1 \otimes dx^1 + dx^2 \otimes dx^2 + dx^3 \otimes dx^3 \,. \tag{8.1}$$

Physically speaking, the events are the minimal space-temporal characterization of everything occurs in the universe at any time. Modulo technicalities we shall examine shortly, Minkowskian coordinate frames defines (not biunivocally) the the class of coordinate system of all *inertial observers* or *inertial reference systems*. Referring to the decomposition (8.1) of $\mathbf{g}$, the coordinates x^1, x^2, x^3, are thought as "spatial coordinates", whereas the coordinate $x^0 = ct$ is a temporal coordinate. The number $c > 0$ is a constant with the dimension of a velocity whose value is that of the *speed of light* (around 3×10^5 km/s). In the following we assume a system of units such that $c = 1$.
The metric g is used to perform measurements either in time and in space as we clarify in the rest of this section. If $X_p \neq 0$ is a vector in $T_p\mathbb{M}^4$, it may represent either infinitesimal temporal displacements if $\mathbf{g}_p(X_p, X_p) \leq 0$ or infinitesimal spatial displacements if $\mathbf{g}_p(X_p, X_p) > 0$. In both cases $|\mathbf{g}_p(X_p, X_p)|$ has the physical meaning of the length (duration) of V_p. Actually the

distinguishable case $\mathbf{g}(X_p, X_p) = 0$ (but $X_p \neq 0$) deserves a particular comment. These vectors represent an infinitesimal part of the history of a particle of light (we could call it "photon", even if a photon is a quantum object).

Definition 8.3. For every event $p \in \mathbb{M}^4$, $T_p\mathbb{M}^4 \setminus \{0\}$ is decomposed in three pair-wise disjoint subsets:

(i) the set of **spacelike** vectors which satisfy: $\mathbf{g}_p(X_p, X_p) > 0$,

(ii) the set of **timelike** vectors which satisfy: $\mathbf{g}_p(X_p, X_p) < 0$,

(iii) the set of **lightlike**, also called **null**, vectors which satisfy: $\mathbf{g}_p(X_p, X_p) = 0$.

The following further definitions hold.

(a) The union of the sets $\{0\}$ and timelike and lightlike vectors is a closed cone, $\overline{V_p}$, called **closed light cone** at p. Its non-vanishing elements are called **causal** vectors.

(b) The interior V_p of $\overline{V_p}$ is called **open light cone** at p (it therefore does *not* contains 0).

(c) The boundary ∂V_p is called **light cone** at p (it therefore *contains* 0). ■

Exercises 8.4.

1. Prove that, in Minkowski spacetime, if $\mathbf{g}(X, Y) = 0$ and X is timelike, then Y is spacelike. Show that, $\mathbf{g}(X, Y) = 0$ may hold for X spacelike and Y of any type.

(*Solution*. Fix a pseudo orthonormal basis to decompose X and Y. Assume X timelike. We have, $0 = \mathbf{g}(X, Y) = -X^0Y^0 + \sum_{k=1}^3 X^kY^k$. However $|X^0| > \sqrt{\sum_k (X^k)^2}$ and $|X^0Y^0| = |\sum_{k=1}^3 X^kY^k| \leq \sqrt{\sum_k (X^k)^2}\sqrt{\sum_k (Y^k)^2}$. The only possibility is $|Y^0| < \sqrt{\sum_k (Y^k)^2}$. To prove th second statement, fix a pseudo orthonormal basis and consider, e.g., the spacelike vector $X := \mathbf{e}_3 \equiv (0, 0, 0, 1)$. In that case $Y \equiv (1, 0, 0, 0)$ is timelike, $Y \equiv (0, 1, 0, 0)$ is spacelike and $Y \equiv (1, 0, -1, 0)$ is lightlike and $\mathbf{g}(X, Y) = 0$ is true in the three cases .)

2. Prove that, in Minkowski spacetime, if X is timelike, then there is a pseudo orthonormal basis $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ such that $\mathbf{e}_0$ is parallel to X.

(*Hint*. Re-adapt, with a suitable choice of signs, the Gramm-Schmidt orthogonalization procedure to a basis that completes X also taking the previous exercise into account.)

8.1.2 Time orientation

Definition 8.5. Two continuous timelike vector fields T, T' on $\mathbb{M}^4$ are said to **have the same time orientation** if $\mathbf{g}_p(T_p, T'_p) < 0$ for every $p \in \mathbb{M}^4$. ■

To explain that fundamental defiinition we start by noticing the following fact. Using a given system of Minkowskian coordinates and referring to the base of $T_p\mathbb{M}^4$ associated with these coordinates, one sees that the open light cone, in every Minkowskian coordinate system is always pictured as the set

$$V_p = \{(X^0, X^1, X^2, X^3) \in \mathbb{R}^4 \setminus \{0\} \mid (X^0)^2 > (X^1)^2 + (X^2)^2 + (X^3)^2\} .$$

Thus, in those Minkowskian coordinates, V_p is made of two disjoint halves

$$V_p^{(>)} := \{(X^0, X^1, X^2, X^3) \in V_p \mid X^0 > 0\} , \quad V_p^{(<)} := \{(X^0, X^1, X^2, X^3) \in V_p \mid X^0 < 0\} . \tag{8.2}$$

Remarks 8.6. It could seem that this decomposition of V_p depends on the used coordinate system: The set $V_p^{(>)}$, for instance could intersect both the sects $V_p^{(>)}$ and $V_p^{(<)}$ defined with respect to *another* Minkowskian coordinate system. Actually this is not the case for the following topological reason. $V_p^{(>)}$ and $V_p^{(<)}$ are open sets in $\mathbb{R}^4$ and thus in T^4, since the topology of the latter is homeomorphic to the one of the former through the coordinate system itself (more precisely, the map associating the components of a vector to the vector itself). Furthermore they are disjoint connectd sets. Hence they are the two connected components of the open light cone V_p. This fact does not depend on the used Minkowskian coordinate system. At most, changing Minkowskian coordinate system one can swap $V_p^{(>)}$ and $V_p^{(<)}$. ■

Proposition 8.7. *The class $\mathscr{T}(\mathbb{M}^4)$ of continuous timelike vector fields on $\mathbb{M}^4$ satisfies following.*
(a) *$\mathscr{T}(\mathbb{M}^4)$ is not empty.*
(b) *If $T, T' \in \mathscr{T}(\mathbb{M}^4)$, it holds either $\mathbf{g}(T_p, T'_p) < 0$ for all $p \in \mathbb{M}^4$, or $\mathbf{g}(T_p, T'_p) > 0$ for all $p \in \mathbb{M}^4$. More precisely, decomposing every V_p into the two disjoint halves (8.2),*
$\mathbf{g}(T_p, T'_p) < 0$ happens if and only if both T_p, T'_p belong to the same half of V_p,
$\mathbf{g}(T_p, T'_p) > 0$ holds if and only if T_p, T'_p belong to different halves.
(c) *"to have the same time orientation" is a equivalence relation in $\mathscr{T}(\mathbb{M}^4)$ and it admits two equivalence classes only.*

Proof. (a) $\mathscr{T}(\mathbb{M}^4)$ is not empty since it includes the vector field ∂_{x^0} associated to any Minkowskian coordinate frame on $\mathbb{M}^4$. Let us show (b). Consider a continuous timelike vector field S. Using Minkowskian coordinates x^0, x^1, x^2, x^3 and the associated orthonormal bases of elements $\mathbf{e}_{k,p} = \partial_{x^k}|_p \in T_p\mathbb{M}^4$, $k = 0, 1, 2, 3$, one has:

$$(S_p^0)^2 > \sum_{i=1}^{3} (S_p^i)^2 \, . \tag{8.3}$$

Consider the two halves $V_p^{(>)}$ and $V_p^{(<)}$ said above. S_p^0 cannot change it sign varying $p \in \mathbb{M}$ because it would imply that $S_p^0 = 0$ which is not allowed. Therefore it holds $S_p \in V_p^{(>)}$ constantly in $p \in \mathbb{M}^4$, that is

$$S_p^0 > \sqrt{\sum_{i=1}^{3} (S_p^i)^2} \, , \quad \text{for all } p \in \mathbb{M}^4 \, , \tag{8.4}$$

or $S_p \in V_p^{(<)}$ constantly in $p \in \mathbb{M}^4$, that is

$$S_p^0 < -\sqrt{\sum_{i=1}^{3} (S_p^i)^2} \, , \quad \text{for all } p \in \mathbb{M}^4 \, . \tag{8.5}$$

Now consider two timelike continuous vector fields T and T'. One has

$$\mathbf{g}_p(T, T') = -T^0 T'^0 + \sum_{\alpha=1}^{3} T'^\alpha T^\alpha \,. \tag{8.6}$$

On the other hand, by Cauchy-Schwartz inequality

$$\left|\sum_{\alpha=1}^{3} T^\alpha T'^\alpha\right| \leq \sqrt{\sum_{\alpha=1}^{3} (T_p^\alpha)^2} \sqrt{\sum_{\alpha=1}^{3} (T_p'^\alpha)^2} \,.$$

Now, taking into account that it must hold either (8.4) or (8.5) for T and T' in place of S, we conclude that the sign of the right-hand side of (8.6) is that of $-T^0 T'^0$. In other words, it holds $\mathbf{g}_p(T_p, T'_p) < 0$ constantly in $p \in \mathbb{M}^4$ or $\mathbf{g}_p(T_p, T'_p) > 0$ constantly in $p \in \mathbb{M}^4$ and the first case happens if and only if both T, T' belong to the same half of V_p, whereas the second case arises if and only if T, T' belong to different halves. The proof of (b) is concluded.
Using (b) result we can discuss the situation in a single tangent space fixing $p \in \mathbb{M}^4$ arbitrarily and we can prove (c), that "to have the same time orientation" is a equivalence relation. By definition of timelike vector $\mathbf{g}(T_p, T_p) < 0$, so T has the same time orientation as T itself. If $\mathbf{g}(T_p, T'_p) < 0$ then $\mathbf{g}(T'_p, T_p) = \mathbf{g}(T_p, T'_p) < 0$ so that the symmetric property is satisfied. Let us come to transitivity. Suppose that $\mathbf{g}(T_p, T'_p) < 0$ so that T_p and T'_p belong to the same half of V_p, and $\mathbf{g}(T'_p, S_p) < 0$ so that T'_p and S_p belong to the same half of V_p. We conclude that T_p and S_p belong to the same half of V_p and thus $\mathbf{g}(T_p, S_p) < 0$. This proves transitivity.
To conclude, notice that, if T is a continuous timelike vector field on $\mathbb{M}^4$, T and $-T$ belong to different equivalence classes and if $\mathbf{g}(T, S) > 0$ then $\mathbf{g}(-T, S) < 0$ so that, every other timelike continuous vector field S belongs to the equivalence class of T or to the equivalence class of $-T$. $\Box$

Definition 8.8. A pair $(\mathbb{M}^4, \mathcal{O})$ where $\mathcal{O}$ is one of the two equivalence classes of the relation "to have the same time orientation" in the set of continuous timelike vector fields on $\mathbb{M}^4$, is called **time oriented** Minkowski spacetime, and $\mathcal{O}$ is called **time orientation** of $\mathbb{M}^4$. $\blacksquare$

There is an alternative way to fix a time orientation: form proposition 8.7 and its corollary we also conclude that:

Proposition 8.9. *The assignment of a time orientation $\mathcal{O}$ is equivalent to continuously select one of the two disjoint halves of V_p, for every $p \in \mathbb{M}^4$: that containing time-like vectors which are restriction to p of a continuous vector field in $\mathcal{O}$.*

Definition 8.10. Considering a time oriented Minkowski spacetime $(\mathbb{M}^4, \mathcal{O})$ and an event $p \in \mathbb{M}^4$, the following definitions are valid

(i) the half open cone at p containing vectors which are restrictions to p of vector fields in $\mathcal{O}$ is denoted by V_p^+ and is called **future open light cone** at p. Its elements are said **future-directed timelike vectors** at p.

(ii) ∂V_p^+ is called **future light cone** and the vectors in $\partial V_p^+ \setminus \{0\}$ are said **future-directed lightlike vectors** at p.

(iii) $\overline{V^+} := V^+ \cup \partial V_p^+$ is called **closed future light cone** and the vectors in $\overline{V^+} \setminus \{0\}$ are said **future-directed causal vectors**. ■

Remarks 8.11. We henceforth suppose to deal with a time oriented Minkowski spacetime and we indicate it by means of $\mathbb{M}^4$ simply. ■

As a final technical proposition we have the following elementary though interesting result.

Proposition 8.12. *If A, B are causal vectors, then the* **inverse Cauchy-Schwartz inequality** *holds*

$$|\mathbf{g}(A,B)| \geq \sqrt{-\mathbf{g}(A,A)}\sqrt{-\mathbf{g}(B,B)}\,,$$

where right-hand side is $-\mathbf{g}(A,B)$ if the vectors are both future-directed or both past-directed.

Proof. It is sufficient to prove the thesis for A, B causal and future-directed (if they are simply causal the sign of $\mathbf{g}(A,B)$ may change, but the final result does not change in view of the absolute value). So we assume that A and B are future-directed.

$$-A_\mu B^\mu = A^0B^0 - \vec{A}\cdot\vec{B} \geq A^0B^0 - |\vec{A}\cdot\vec{B}| \geq A^0B^0 - ||\vec{A}||||\vec{B}|| \geq \sqrt{((A^0)^2 - ||\vec{A}||^2)((B^0)^2 - ||\vec{B}||^2)}$$

$$= \sqrt{-A_\mu A^\mu}\sqrt{-B_\mu B^\mu}\,.$$

The last inequality is immediate using the fact that $A^0, B^0 > 0$ and proving it into the equivalent version

$$(A^0B^0 - ||\vec{A}||||\vec{B}||)^2 \geq ((A^0)^2 - ||\vec{A}||^2)(B^0)^2 - ||\vec{B}||^2)$$

whose proof is immediate, since it boils down to

$$(A^0B^0)^2 + (||\vec{A}||||\vec{B}||)^2 - 2A^0B^0||\vec{A}||||\vec{B}|| \geq (A^0B^0)^2 + (||\vec{A}||||\vec{B}||)^2 - (A^0||\vec{B}||)^2 - (B^0||\vec{A}||)^2\,,$$

namely

$$(A^0||\vec{B}|| - B^0||\vec{A}||)^2 \geq 0\,.$$

□

8.2 Kinematics I: basic notions

We pass now to discuss the physical intepretation of some geometrical structures of Minkowski spacetime in terms of kinematical notions.

8.2.1 Curves as histories or world lines

If $I \subset \mathbb{R}$ is an open interval, a C^1 curve (def. 3.13) $\gamma : I \to \mathbb{M}^4$ may represent the history of a particle of matter, also called its **world line**, provided its tangent vector (def. 3.15) is causal and future directed. A particle of light, in particular, has a world line represented by a curve with future-directed lightlike tangent vector.

Definition 8.13. A C^1 curve $\gamma : I \to \mathbb{M}^4$ is said to be **timelike**, **spacelike**, **causal**, **lightlike** if all of its tangent vectors γ'_p, $p \in \gamma$, are respectively timelike, spacelike, causal, lightlike. A causal, timelike, lightlike curve γ with future oriented tangent vectors is said to be **future-directed (resp. causal, timelike, lightlike) curve**. ■

Notice that $\gamma'(\xi) \neq 0$ necessarily for $\xi \in I$ when γ is causal. However, in that case, $|g(\gamma'(\xi), \gamma'(\xi))| \neq 0$ only if $\gamma'(\xi)$ is timelike

Definition 8.14. If $\gamma = \gamma(\xi)$, $\xi \in I$ with $I \subset \mathbb{R}$ any interval of $\mathbb{R}$, is any C^1 future-directed timelike curve in $\mathbb{M}^4$ and $\xi_0 \in I$, the length function (that is C^1)

$$\tau(\xi) := \int_{\xi_0}^{\xi} \sqrt{|g(\gamma'(l), \gamma'(l))|}dl \tag{8.7}$$

is called **proper time** of γ. The tangent vector obtained by re-parameterizing the curve with the proper time which is denoted by $\dot{\gamma}$ (instead of γ') is called **four-velocity** of the curve. ■

Notice that proper time is independent from the used parametrization of γ but it is defined up to the choice of the origin: the point on γ where $\tau = 0$.
From the point of view of physics, when one considers timelike curves, proper time is *time measured by an ideal clock at rest with the particle whose world line is γ.*
For timelike curves, proper time can be used as natural parameter to describe the history γ of the particle because, in that case, directly from (8.7), $d\tau/d\xi > 0$. Notice also that, if $\dot{\gamma}$ is a four-velocity (8.7) implies that:

$$\mathbf{g}(\dot{\gamma}, \dot{\gamma}) = -1 \,, \tag{8.8}$$

(where -1 is replaced by $-c^2$ if restoring the value of the speed of light c), that is a four-velocity is *unitary*. Indeed,

$$\frac{d\gamma}{d\tau} = \frac{d\gamma}{d\xi}\left(\frac{d\tau}{d\xi}\right)^{-1} = \frac{d\gamma}{d\xi}\left(\sqrt{|\mathbf{g}(\gamma'(\xi), \gamma'(\xi))|}\right)^{-1}$$

which produces (8.8) by noticing that $\mathbf{g}(\gamma'(\xi), \gamma'(\xi)) \leq 0$.

Remarks 8.15. An important fact is that future-directed causal curves can be parametrized by the x^0 coordinate of any Minkowskian system of coordinates. This is because, if the curve is represented as $I \ni \xi \mapsto (x^0(\xi), x^1(\xi), x^2(\xi), x^3(\xi))$ in those coordinates, then

$$0 \geq \mathbf{g}(\gamma'(\xi), \gamma'(\xi)) = -\left(\frac{dx^0}{d\xi}\right)^2 + \sum_{k=1}^{3}\left(\frac{dx^k}{d\xi}\right)^2$$

implies $\frac{dx^0}{d\xi} \neq 0$ (the only other possibility is that $\gamma'(\xi) = 0$ somewhere but it is not permitted because γ' is causal), so that we can re-parametrize the curve using x^0 instead of ξ. More precisely, if γ' and ∂_{x^0} have the same temporal orientation, i.e., if also ∂_{x^0} is future-directed, then

$$-\mathbf{g}(\gamma', \partial_{x^0}) = \frac{dx^0}{d\xi} > 0$$

so that if x^0 increases then $\gamma(\xi(x^0))$ evolves forward in time. ■

From a physical point of view, future-directed causal curves which are straight lines (segments) with respect to the affine structure of $\mathbb{M}^4$ are thought to be the histories of particles not subjected to forces. *These are particles evolving with inertial motion.*

8.2.2 Minkowskian reference systems and some kinematical effect

If x^0, x^1, x^2, x^3 are Minkowskian coordinates, the curves tangent to x^0 determine a constant vector field ∂_{x^0} which is unitary, that is it satisfies $\mathbf{g}(\partial_{x^0}, \partial_{x^0}) = -1$. It is constant in the sense that, by definition of Cartesian coordinate system, it does not depend on the point in the affine space. A unitary constant timelike vector field can always be viewed as the tangent vector to the coordinate x^0 of a Minkowskian coordinate frame. There are anyway *several* Minkowskian coordinate frames associated this way with a given unitary constant timelike vector field. They deffer to aech other just for the choice of the origin.

Definition 8.16. A constant timelike vector field $\mathscr{R}$ on $\mathbb{M}^4$ is said a **(Minkowskian) reference system** provided it is future-directed and unitary, i.e. $\mathbf{g}(\mathscr{R}, \mathscr{R}) = -1$. Furthermore the following definitions hold.
(a) Any Minkowskian coordinate system such that $\partial_{x^0} = \mathscr{R}$, is said to be **co-moving with** (or equivalently **adapted to**) $\mathscr{R}$.
(a) If $\gamma = \gamma(\xi)$, where $\xi \in \mathbb{R}$, is an integral curve of $\mathscr{R}$ (an affine straight line), the length function

$$t_{\mathscr{R}}(\gamma(\xi)) := \int_0^{\xi} \sqrt{|\mathbf{g}(\mathscr{R}_{\gamma(l)}, \mathscr{R}_{\gamma(l)})|}dl$$

defines a **time coordinate for $\mathscr{R}$ (along γ)**. ■

For a fixed $\mathscr{R}$, an adapted Minkowskian coordinate frame can be obtained by fixing a three-dimensional affine plane Σ orthogonal to γ in O, and fixing an orthonormal basis in $T_p\mathbb{M}$ such that $\mathbf{e}_0 = \mathscr{R}$ and $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ belong (are parallel) to Σ. Now one proves straightforwardly that the

Minkowskian coordinate frame associated with $O, \mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is co-moving with $\mathscr{R}$ and, along γ, x^0 coincides with $t_{\mathscr{R}}$ up an additive constant. Notice that x^0 defines in this way a **global temporal coordinate** associated with $\mathscr{R}$, since the Cartesian coordinate x^0 associates a time valure coordinate $t_{\mathscr{R}}(p) := x^0(p)$ to every event of $p \in \mathbb{M}^4$, not only to those along the initially chosen curve γ tangent to $\mathscr{R}$. Moreover, if a global time coordinate x^0 is defined for $\mathscr{R}$, every three-dimensional affine plane orthogonal to $\mathscr{R}$ contains only points with constant value of that time coordinate.

Definition 8.17. Let $\mathscr{R}$ be a reference system in $\mathbb{M}^4$ and t an associated global time coordinate (the coordinate x^0 of a Minkowskian coordinate system co-moving with $\mathscr{R}$). Any three-dimensional affine plane $\Sigma_t^{(\mathscr{R})}$ orthogonal to $\mathscr{R}$, $t \in \mathbb{R}$ being the value of the time coordinate of the points in the plane, is called **rest space of $\mathscr{R}$ at time** t. ■

From the point of view of differential geometry, by construction, every rest space $\Sigma_t^{(\mathscr{R})}$ of $\mathscr{R}$ is an embedded three-dimensional submanifold of $\mathbb{M}^4$. Furthermore, the *metric induced by* $\mathbf{g}$ *on* $\Sigma_t^{(\mathscr{R})}$ – i.e. the *scalar product of vectors parallel to* $\Sigma_t^{(\mathscr{R})}$ – is positive definite. This is trivially true because that scalar product works on the span of three orthonormal vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ which are separately orthonormal to $\mathbf{e}_0 = \mathscr{R}$, and we know that $\mathbf{g}(\mathbf{e}_\alpha, \mathbf{e}_\beta) = \delta_{\alpha\beta}$ if $\alpha, \beta = 1, 2, 3$.
From the point of view of physics, a Minkowski frame $\mathscr{R}$ is nothing but an *inertial reference system*. Furthermore:

(a) any global time coordinate of $\mathscr{R}$ constructed out of $\mathbf{g}$ is *time measured by ideal clocks at rest with the space of the reference system*;

(b) the metric induced by $\mathbf{g}$ on the rest spaces of $\mathscr{R}$ is the mathematical tool corresponding to *physical spatial measurements performed by rigid rulers at rest with the space of the reference system.*

This mathematical picture is sufficiently developed to allows one to define the notion of *velocity of a particle* represented by a C^1 future-directed causal curve γ, *with respect to a reference system* $\mathscr{R}$ *at time* t. The procedure is straightforward. Consider the event $e := \Sigma_t^{(\mathscr{R})} \cap \gamma$ where the curve intersect the rest space $\Sigma_t^{(\mathscr{R})}$ (the reader should prove that each rest space intersects γ exactly in a point). $T_e\mathbb{M}^4$ is decomposed to the orthogonal direct sum

$$T_e\mathbb{M}^4 = \mathrm{Span}(\mathscr{R}_e) \oplus \Sigma_t^{(\mathscr{R})} , \tag{8.9}$$

$\mathrm{Span}(\mathscr{R}_e)$ being the linear space spanned by the vector $\mathscr{R}_e$ and $\Sigma_t^{(\mathscr{R})}$ the subspace of vectors tangent (parallel) to $\Sigma_t^{(\mathscr{R})}$. As a consequence, the tangent vector γ'_e turns out to be uniquely decomposed as

$$\gamma'_e = \delta t \mathscr{R}_e + \delta X , \tag{8.10}$$

where $\delta X \in T_e\Sigma_t^{(\mathscr{R})}$ and $\delta t \in \mathbb{R}$. The fact that $\dot{\gamma}_e$ is causal prevents δt from vanishing and the fact that γ is future-directed implies that $\delta t > 0$ (the reader should prove it), so that it makes sense to give the following definition of velocity.

Definition 8.18. Let γ be a C^1 future-directed causal curve and $\mathscr{R}$ a reference system in $\mathbb{M}^4$. The **velocity of γ with respect to $\mathscr{R}$ at time t** is the vector $\mathbf{v}_{\gamma,t,\mathscr{R}} \in T_{\Sigma_t^{(\mathscr{R})}\cap\gamma}\Sigma_t^{(\mathscr{R})}$, given by

$$\mathbf{v}_{\gamma,t,\mathscr{R}} := \frac{\delta X}{\delta t} \,.$$

referring to (8.9) and (8.10) with $e := \Sigma_t \cap \gamma$. ■

Notice that the ratio above does not depend on the parametrization of γ because, changing parametrization form ξ to ξ', the components of γ' whould be multiplied by the same factor $\frac{d\xi}{d\xi'}$, and these factors would cancel when computing the ration defining the velocity. Within our framework one has the following physically well-known result (where we explicitely write the value c of the speed of light).

Proposition 8.19. *Consider a future-directed causal curve γ and fix a reference system $\mathscr{R}$,*
(a) *The absolute value of $\mathbf{v}_{\gamma,t,\mathscr{R}}$ (the* speed *of γ at t) is bounded above by c and that value is attained only at those events along γ where the tangent vector of γ is lightlike.*
(b) *If γ is timelike and $\dot\gamma(t)$ is the four-velocity of γ at time t of $\mathscr{R}$:*

$$\dot\gamma(t) = \frac{c}{\sqrt{1-\mathbf{v}_{\gamma,t,\mathscr{R}}{}^2/c^2}}\mathscr{R} + \frac{\mathbf{v}_{\gamma,t,\mathscr{R}}}{\sqrt{1-\mathbf{v}_{\gamma,t,\mathscr{R}}{}^2/c^2}} \,. \tag{8.11}$$

(c) *If γ intersects the rest spaces $\Sigma_{t_1}^{(\mathscr{R})}$ and $\Sigma_{t_2}^{(\mathscr{R})}$ with $t_2 > t_1$ at proper time τ_1 and $\tau_2 > \tau_1$ respectively, the corresponding intervals of time and proper time satisfy*

$$\Delta\tau = \int_{t_1}^{t_2} \sqrt{1-\mathbf{v}_{\gamma,t,\mathscr{R}}{}^2/c^2}\,dt \tag{8.12}$$

and so $\Delta\tau < \Delta t$ unless $\mathbf{v}_{\gamma,t,\mathscr{R}} = 0$ in the considered interval of time and, in that case, $\Delta\tau = \Delta t$.

Remarks 8.20.
(1) The absolute value $||\mathbf{v}_{\gamma,t,\mathscr{R}}||^2 = (\mathbf{v}_{\gamma,t,\mathscr{R}})^2$ is referred to the scalar product induced by g in the rest spaces of $\mathscr{R}$. As previously said this is the physical metric tool which corresponds to perform measurements.
(2) The value of $||\mathbf{v}_{\gamma,t,\mathscr{R}}||$ evaluated along curves with $\gamma'_e = \delta t\mathscr{R}_e + \delta X$ such that $\mathbf{g}(\gamma'_e, \gamma'_e) = 0$, is the same for *every* reference system. This value is denoted by c as already said (and sometime put $= 1$ by the sake of semplicity). It coincides, in particular with the universal value of the speed of particles of light. ■

Proof of proposition 8.19. Using definition 8.18, decomposition (8.10), the orthogonality of δX and $\mathscr{R}$ and the fact that $\mathscr{R}$ is unitary, one has

$$0 \geq \mathbf{g}(\gamma', \gamma') = -c^2\delta t^2 + g(\delta X, \delta X) = -c^2\delta t^2 + ||\delta X||^2 \,.$$

The the sign $=$ occurs only if γ' is lightlike. This implies the thesis (a) immediately. (b) We write v in place of $\mathbf{v}_{\gamma,t,\mathscr{R}}$. Now we explicitly assume that the curve is timelike and parametrized with the proper time. By definition, as $\partial_{x^0} = \mathscr{R}$:

$$\dot{\gamma} = \dot{\gamma}^0 \partial_{x^0} + \sum_{\alpha=1}^{3} \dot{\gamma}^\alpha \partial_{x^\alpha}$$

where

$$-(\dot{\gamma}^0)^2 + \sum_{\alpha=1}^{3} (\dot{\gamma}^\alpha)^2 = -c^2 ,$$

and $v^i = c\dot{\gamma}^i/\dot{\gamma}^0$, so that

$$-1 + \sum_{\alpha=1}^{3} (v^\alpha/c)^2 = -c^2/(\dot{\gamma}^0)^2 ,$$

and so:

$$\dot{\gamma}^0 = c/\sqrt{1 - v^2/c^2} \quad \text{and} \quad \dot{\gamma}^\alpha = \dot{\gamma}^0 v^\alpha = v^\alpha/\sqrt{1 - v^2/c^2}$$

Finally (c) is a straightforward consequence of (b) since the the factor in front of $\mathscr{R}$ in (8.11) is $cdt/d\tau$. That is $cdt/d\tau = \frac{c}{\sqrt{1 - \mathbf{v}_{\gamma,t,\mathscr{R}}{}^2/c^2}}$. $\square$

We have a very interesting geometrical consequence.

Proposition 8.21. *The following facts are true in Minkowski spacetime.*
(a) *A timelike segment joining a pair of events in Minkwski spacetime is the unique* longest *causal curve joining these events.*
(b) *If a pair of events can be joined by a spacelike segment, then the infimum of the lenght of all C^1 curves joining those events is* 0*.*

Proof. To prove (a) it is sufficient to adopt a Minkowskian coordinate system with ∂_t parallel to $\mathscr{R}$. With this choice $\Delta\tau$ computed in (8.12) of Proposition 8.19 reduces to $\Delta t = t_2 - t_1$ for a timelike segment joining the said events (which, by construction, stay on the same coordinate line tangent to ∂_t). If, keeping those endpoints, we change the causal curve γ joining them, we see that since

$$\sqrt{1 - \mathbf{v}_{\gamma,t,\mathscr{R}}{}^2/c^2} \leq 1$$

it necessarily holds

$$\Delta\tau \leq \Delta t ,$$

proving the assertion. The segment is the unique longest curve because, if $\Delta t = \Delta\tau$ then $\int_{t_1}^{t_2} (1 - \sqrt{1 - \mathbf{v}_{\gamma,t,\mathscr{R}}{}^2/c^2}) dt = 0$. Since the integrand is continuous and non-negative, this is possible only if the integrand is the zero function. In other words $\mathbf{v}_{\gamma,t,\mathscr{R}} = 0$ along all the curve. This implies that the curve is constant and coincides with the temporal segment joining the two endpoints.

The proof of (b) is elementary: Let A and B the said segments. Fix a Minkowskian coordinate system such that ∂_{x^0} is orthogonal to the spacelike segment joining the two events and arrange the origin of the coordinates in the middle point of the segment AB in order that AB stays along the axis x^1 with $-x^1(A) = -x^1(B) = \ell$. Consider a class of C^1 curves joining A and B arbitrarily approximating a continuous curve made of two lighlike AT and TB, where (assuming $c = 1$) $T \equiv (\ell, 0, 0, 0)$. The length of this broken segments is 0 proving the thesis. $\Box$.

Examples 8.22.
1. One recognizes in (c) of Proposition 8.19 the mathematical formulation of the celebrated relativistic phenomenon called *dilatation of time*. In particular when the absolute value of $\mathbf{v}_{\gamma,t,\mathscr{R}}$ is constantly v, (c) specializes to:

$$\Delta t = \frac{\Delta\tau}{\sqrt{1-\frac{v^2}{c^2}}} \tag{8.13}$$

where $\Delta\tau$ is an interval of time measured – between the rest spaces $\Sigma^{(\mathscr{R})}_{t_1}$ and $\Sigma^{(\mathscr{R})}_{t_2}$ – by a clock at rest with the particle with speed v in the reference system $\mathscr{R}$ and $t_1 - t_1 = \Delta t = \Delta x^0/c$ is the corresponding interval of time masured by a clock at rest in the reference system.
This phenomenon is well known in particle physics. There are instable particles called *pions* produced in Earth's atmosphere by the interaction of atmospheric particles and consmic rays. If produced in a laboratory, these particle prove to have a certain known lifetime $\Delta\tau$. This lifetime is not enough to reach the ground from the regions of high atmospehere where they are produced, even if they moved with the speed of light. However these particles are detected at the ground level of Earth! The explanation realies upon (c) of Proposition 8.19. On the one hand, $\Delta\tau$ is a feature of the particle measured at rest with the particle. On the other hand the associated and larger interval of time Δt necessary to reach the ground, computed in (8.13) and referred to the rest frame of Earth, is larger that $\Delta\tau$ pretty enough to allow the pion with speed $v < c$ to reach the ground before decaying[1].
2. The so-called *twin paradox* arises when one (wrongly) applies the time dilatation phenomenon to the case of a couple of twin brothers, one of whom makes a journey into space in a high-speed rocket and returns home to find that the twin who remained at reast has aged more. The paradox consists of noticing that the point of view can be reversed by arguing that if staying in the reference frame of the moving brother, actually the other borther made a journey at high speed. Hence the younger twin should be the other one! The paradox is only apparent, because it is impossible that the twins meet twice both describing a straight timelike wordline in spacetime. At most one of them remained at reast with an inertial frame and he is the youngest guy on their second meeting according to (b) in Proposition 8.21. If none of them remained at rest with an inertial reference frame, the difference of age observed on their next meeteng can be computend by integrating the lenght of their worldlines as in (8.12) and the result is just matter of computation.
3. The *Hafele–Keating* experiment was a famous test of special relativity. In October 1971, J.C.

[1]Our planet cannot be considered an inertial reference system, but the corrections due to this fact can be proved to be negligible in this case.

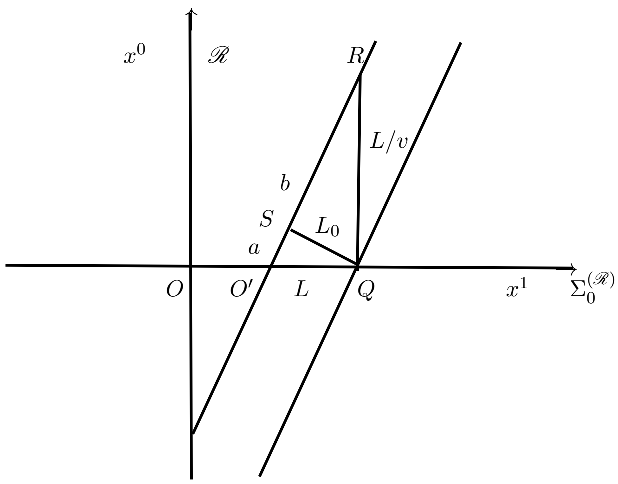

Figure 8.1: Phenomenon of length contraction

Hafele and R.E. Keating, took four cesium-beam atomic clocks aboard commercial airliners. They flew twice around the world, first eastward, then westward, and compared the clocks with others that remained at the United States Naval Observatory. When reunited, the three sets of clocks were found to disagree with one another, and their differences were consistent with the predictions of special (and general) relativity.

4. Another interesting relativistic phenomenon is the so-called *length contraction*. Consider a rigid ruler with length L_0 measured in its rest frame, moving with constant velocity $\mathbf{v} = v\mathbf{e}_1$ with respect to the inertial reference system $\mathscr{R}$ (assuming that $\mathbf{e}_1$ is one of the spacelike vectors used to define a Minkowkian coordinate system co-moving with $\mathscr{R}$). What is the length L of the ruler measured in $\mathscr{R}$? To anwswer we have to stress that the spatial length is however obtained using the metric $\mathbf{g}$, simply referring to spacelike vectors.

Figure 8.22 illustrates a way to compute it. Let us fix a Minkowkian reference frame at rest with the ruler with origin O' and let $\mathscr{R}'$ be the reference frame at rest with the ruler and $\mathbf{e}'_1$ a unit spatial vector normal to $\mathscr{R}'$ and parallel to the ruler. Consider the rectangular triangle of events $O'QR$ in $\mathbb{M}^4$. The lenght of the segment (vector) QR amout to the time lapse L/v necessary for the the left end point to cover the (unknown) length L of the ruler. Therefore as $R - O' = R - Q + Q - O'$ and $Q - O' \perp R - Q$, the length of $O'R$ can be computed using:

$$g(R - O', R - O') = g(R - Q, R - Q) + g(Q - O', Q - O')$$

that is, decomposing also $R - O'$ with respect to S as $R - O' = b\mathscr{R}' + a\mathscr{R}'$ where a and b are evalueated along $\mathscr{R}'$ starting from O',

$$-(a+b)^2 = L^2 - \frac{L^2}{v^2} \,. \tag{8.14}$$

Since both triangles PQS and RSQ are rectangular, and respectively $-a\mathscr{R}' + L_0\mathbf{e}'_1 = L\mathbf{e}_1$,

$b\mathscr{R}' - L_0\mathbf{e}_1' = \frac{L}{v}\mathscr{R}$ taking the square of both sides similar relations hold true:

$$-a^2 + L_0^2 = L^2 \,, \quad -b^2 + L_0^2 = -\frac{L^2}{v^2} \,. \tag{8.15}$$

(8.14) and (8.15) imply $-ab = L_0^2$ (see below for the apparent weird sign) and thus

$$a^2 b^2 = L_0^4 \,.$$

That is

$$(L_0^2 - L^2)\left(L_0^2 + \frac{L^2}{v^2}\right) = L_0^4 \,.$$

Some trivial simplification leads to

$$L = L_0\sqrt{1 - \frac{v^2}{c^2}} \,, \tag{8.16}$$

where we have explicitly restored the value of c. It is evident from (8.14) that $L < L_0$ as soon as $0 < v < c$.

The assertion $L < L_0$ seems wrong if looking at the picture 8.22, but this is nothing but a consequence of the fact that that we automaticaly adopt an Euclidean geometry intepretation when looking that sort of diagrams: Lorentzian lenghts are not Euclidean lengths. A related point is the fact that the identity $-ab = L_0^2$ says that a and b cannot have the same sign as instead our figure suggests (beacuse we are using the Euclidean notion of orthogonality!). Once more, the figure is not a faithful image of what happens in relativity though it is useful in computations.

■

8.3 Kinematics II: Lorentz and Poincaré groups

The following natural question arises: if $\mathscr{R}$ and $\mathscr{R}'$ are two Minkowskian reference systems equipped with co-moving Minkowskian coordinate systems x^0, x^1, x^2, x^3 and x'^0, x'^1, x'^2, x'^3 respectively, what about the most general relation between these different systems of coordinates? From now on we exploit again the convention of summation over repeated indices. Since both coordinate frame are Minkowskian which, in turn, are Cartesian coordinate frames, the relation must be linear:

$$x'^a = \Lambda^a{}_b x^b + T^a \,, \tag{8.17}$$

the requirement of non singularity of Jacobian determinant is obviously equivalent to non singularity of the matrix Λ of coefficient $\Lambda^\mu{}_\nu$. Finally, the requirement that in both coordinate system $\mathbf{g}$ must have the form (8.1), i.e.

$$\mathbf{g} = \eta_{ab} dx^a \otimes dx^b = \eta_{ab} dx'^a \otimes dx'^b \tag{8.18}$$

where we have introduced the matrix $\eta = diag(-1, +1, +1, +1)$ of coefficients η_{ab}, leads to the requirement

$$\Lambda^t \eta \Lambda = \eta \,. \tag{8.19}$$

Notice that this relation implies the non singularity of matrices Λ because, taking the determinant of both sides one gets:

$$(\det \Lambda)^2 \det \eta = \det \eta \,,$$

where we exploited $\det(\Lambda^t \eta \Lambda) = \det(\Lambda^t) \det(\eta) \det(\Lambda) = \det(\Lambda) \det(\eta) \det(\Lambda)$. Since $\det \eta = -1$, it must be $\det \Lambda = \pm 1$. Proceeding backwardly one sees that if x'^0, x'^1, x'^2, x'^3 is a Minkowskian coordinate frame and (8.17) hold with Λ satisfying (8.19), then x^0, x^1, x^2, x^3 is a Minkowskian coordinate frame too. Summarizing one has the following straightforward result.

Proposition 8.23. *If x^0, x^1, x^2, x^3 and x'^0, x'^1, x'^2, x'^3 are Minkowskian coordinate systems on $\mathbb{M}^4$, the transformation rule between these coordinates has the form (8.17) where T^a, $a = 0, 1, 2, 3$ are suitable reals and the matrix Λ satisfies (8.19).*
Conversely, if x'^0, x'^1, x'^2, x'^3 is a Minkowskian coordinate system and (8.17) hold for arbitrary real constants T^a, $a = 0, 1, 2, 3$ and an arbitrary matrix Λ satisfying (8.19), then x^0, x^1, x^2, x^3 is a another Minkowskian coordinate system.

We stress that up to now nothing has been said about time orientation.

8.3.1 The Lorentz group

The result proved above allows one to introduce the celebrated Lorentz group. In the following, $M(4, \mathbb{R})$ will denote the algebra of real 4×4 matrices and $GL(4, \mathbb{R})$ (see section 4.1) indicates the group of real 4×4 matrices with non-vanishing determinant. Consider the set of matrices

$$O(1,3) \quad := \{\Lambda \in M(4, \mathbb{R}) \mid \Lambda^t \eta \Lambda = \eta\} \,. \tag{8.20}$$

It is is a subgroup of $GL(4, \mathbb{R})$. To establish it it is sufficient to verify that it is closed with respect to the multiplication of matrices, and this is trivial from (8.20) using the fact that $\eta\eta = I$, and that if $\Lambda \in O(1,3)$ also $\Lambda^{-1} \in O(1,3)$. The proof this last fact is consequence of (a), (b) and (c) in proposition 8.25 whose proofs is completely based on (8.20). We are now in a position to give the following defintion.

Definition 8.24. (**Lorentz Group**.) The **Lorentz group** is the group of matrices, with group strucuture induced by that of $GL(4, \mathbb{R})$,

$$O(1,3) \quad := \{\Lambda \in M(4, \mathbb{R}) \mid \Lambda^t \eta \Lambda = \eta\} \,.$$

■

The next technical proposition will allow us to introduce some physically relevant subgroups of $O(1,3)$ later.

Proposition 8.25. *The Lorentz group enjoys the following properties.*
(a) $\eta, -I, -\eta, \in O(1,3)$.
(b) $\Lambda \in O(1,3)$ *if and only if* $\Lambda^t \in O(1,3)$.
(c) *If* $\Lambda \in O(1,3)$ *then* $\Lambda^{-1} = \eta\Lambda^t\eta$.
(d) *If* $\Lambda \in O(1,3)$ *then* $\det\Lambda = \pm 1$. *In particular, if* $\Lambda, \Lambda' \in O(1,3)$ *and* $\det\Lambda = \det\Lambda' = 1$ *then* $\det(\Lambda\Lambda') = 1$ *and* $\det\Lambda^{-1} = 1$. *As* $\det I = 1$ *as well, the set of matrices* $\Lambda \in O(1,3)$ *with* $\det\Lambda = 1$ *form a subgroup of* $O(1,3)$ *denoted by* $SO(1,3)$ *and called* **special Lorentz group**.
(e) *If* $\Lambda \in O(1,3)$ *then either* ${\Lambda^0}_0 \geq 1$ *or* ${\Lambda^0}_0 \leq -1$. *In particular, if* $\Lambda, \Lambda' \in O(1,3)$ *and* ${\Lambda^0}_0 \geq 1$ *and* ${\Lambda'^0}_0 \geq 1$ *then* ${(\Lambda\Lambda')^0}_0 \geq 1$ *and* ${(\Lambda^{-1})^0}_0 \geq 1$. *As* ${I^0}_0 = 1$ *as well, the set of matrices* $\Lambda \in O(1,3)$ *with* ${\Lambda^0}_0 \geq 1$ *form a subgroup of* $O(1,3)$ *indicated by* $O(1,3)\uparrow$ *and called* **orthochronous Lorentz group**.

Proof. The proof of (a) is immediate from (8.20) also using $\eta\eta = I$. To prove (b) we start from $\Lambda^t\eta\Lambda = \eta$. Since Λ is not singular, Λ^{-1} exists and one has $\Lambda^t\eta\Lambda\Lambda^{-1} = \eta\Lambda^{-1}$, namely $\Lambda^t\eta = \eta\Lambda^{-1}$. Therefore, applying η on the right $\Lambda^t = \eta\Lambda^{-1}\eta$. Finally applying $\Lambda\eta$ on the left one gets
$$\Lambda\eta\Lambda^t = \eta\,,$$
so $\Lambda^t \in O(1,3)$ if $\Lambda \in O(1,3)$.
To show (c) we notice that $\eta\Lambda^t\eta \in O(1,3)$ because this set is closed with respect to composition of matrices and $\eta, \Lambda^t \in O(1,3)$ for (a) and (b). Finally: $\Lambda(\eta\Lambda^t\eta) = (\Lambda\eta\Lambda^t)\eta = \eta\eta = I$. Since every $\Lambda \in SO(1,3)$ is non singular as noticed below (8.19) we can conclude that $\eta\Lambda^t\eta = \Lambda^{-1}$ but also that $\Lambda^{-1} \in O(1,3)$ if $\Lambda \in O(1,3)$.
The first part of (d) has been proved previously. The remaining part is straightforward: $\det(\Lambda\Lambda') = (\det\Lambda)\cdot(\det\Lambda') = 1\cdot 1 = 1$ and $\det(\Lambda^{-1}) = (\det\Lambda)^{-1} = (1)^{-1} = 1$.
Let us conclude the proof by demonstrating (e) whose proof is quite involved. The constraint $(A^t\eta A)_{00} = \eta_{00}$ gives :
$$({A^0}_0)^2 = 1 + \sum_{\alpha=1}^{3} {A^\alpha}_0 {A^\alpha}_0\,, \tag{8.21}$$
so that ${A^0}_0 \geq 1$ or ${A^0}_0 \leq -1$ if $A \in O(1,3)$. This proves the first statement of (e) if $\Lambda = A$. Let us pass to the second part. Suppose that $\Lambda, \Lambda' \in O(1,3)$ and ${\Lambda^0}_0 \geq 1$ and ${\Lambda'^0}_0 \geq 1$, we want to show that ${(\Lambda\Lambda')^0}_0 \geq 1$. Actually it is sufficient to show that ${(\Lambda\Lambda')^0}_0 > 0$ because of the first statement. We shall prove it.
We start from the identity
$${(\Lambda\Lambda')^0}_0 = {\Lambda^0}_0 {\Lambda'^0}_0 + \sum_{\alpha=1}^{3} {\Lambda^0}_\alpha {\Lambda'^\alpha}_0\,.$$

It can be re-written down as

$$(\Lambda\Lambda')^0{}_0 = (\Lambda^t)^0{}_0\Lambda'^0{}_0 + \sum_{\alpha=1}^{3}(\Lambda^t)^\alpha{}_0\Lambda'^\alpha{}_0 \tag{8.22}$$

Using Cauchy-Schwarz' inequality:

$$\left|\sum_{\alpha=1}^{3}(\Lambda^t)^\alpha{}_0\Lambda'^\alpha{}_0\right|^2 \leq \left(\sum_{\alpha=1}^{3}(\Lambda^t)^\alpha{}_0(\Lambda^t)^\alpha{}_0\right)\left(\sum_{\beta=1}^{3}\Lambda'^\beta{}_0\Lambda'^\beta{}_0\right) \leq \left((\Lambda^t)^0{}_0\Lambda'^0{}_0\right)^2 ,$$

where, in the last passage we have used (8.21) which is valid both for Λ and Λ^t (since the Lorentz group is closed under tansposition as alerady established) and furthermore $(\Lambda^t)^0{}_0 = \Lambda^0{}_0 > 0$ and $\Lambda'^0{}_0 > 0$ by hypothesis. We conclude from (8.22) that the sign of $(\Lambda\Lambda')^0{}_0$ must be the sign of $(\Lambda^t)^0{}_0\Lambda'^0{}_0$ i.e., strictly positive as wanted.
To prove the last statement, notice that, if $\Lambda^0{}_0 \geq 1$, from (c), $(\Lambda^{-1})^0{}_0 = (\eta\Lambda^t\eta)^0{}_0 = \Lambda^0{}_0 \geq 1$. $\Box$

8.3.2 The Poincaré group

Considering the complete transformation (8.17) we can introduce the celebrated Poincaré group, also called inhomogeneous Lorentz group. To do it, we notice that the set

$$IO(1,3) := O(1,3) \times \mathbb{R}^4 , \tag{8.23}$$

is a group when equipped with the composition rule

$$(\Lambda, T) \circ (\Lambda', T') := (\Lambda\Lambda', T + \Lambda T') . \tag{8.24}$$

The proof of it is immediate and it is left to the reader.

Definition 8.26. The **Poincaré group** or **inhomogeneous Lorentz group** is the set of matrices

$$IO(1,3) := O(1,3) \times \mathbb{R}^4 ,$$

equipped with the composition rule

$$(\Lambda, T) \circ (\Lambda', T') := (\Lambda\Lambda', T + \Lambda T') .$$

■

We make only a few remarks.

Remarks 8.27. The composition rule (8.24) is nothing but that obtained by composing the two transformations of Minkowskian coordinates:

$$x_1^a = \Lambda^a{}_b x_2^b + T^a \quad \text{and} \quad x_2^b = \Lambda'^b{}_c x_3^c + T'^b$$

obtaining

$$x_1^a = (\Lambda\Lambda')^a{}_c x_3^c + T^a + (\Lambda T')^a \,.$$

It is also evident that $IO(1,3)$ it is nothing but the semi-direct product (see subsection 4.1.2) of $O(1,3)$ and the group of spacetime displacements $\mathbb{R}^4$, $IO(1,3) \times_\psi \mathbb{R}^4$. In this case, for every $\Lambda \in O(1,3)$, the group isomorphism $\psi_\Lambda : \mathbb{R}^4 \to \mathbb{R}^4$ is nothing but $\psi_T := \Lambda T$. ■

The subgroup $IO(1,3)\!\uparrow$ of $IO(1,3)$ whose elements (Λ, T) satisfy $\Lambda^0{}_0 \geq 0$ (equivalently $\Lambda^0{}_0 \geq 1$) is called the **orthochronous Poincaré group**. Its physical meaning should be clear it is the group of Poicaré transformations connecting pairs of Minkowski coordinate systems associated with corresponding pairs of *inertial reference systems*. This is due to the reason that the vectors $\mathscr{R}$ and $\mathscr{R}'$ of two inertial reference systems are both future oriented and thus $\mathbf{g}(\mathscr{R}, \mathscr{R}') < 0$. If x^0, x^1, x^2, x^3 and x'^0, x'^1, x'^2, x'^3 are Minkowskian coordinates respectively associated with these reference system we have:

$$0 > \mathbf{g}(\mathscr{R}', \mathscr{R}) = \mathbf{g}(\partial_{x'^0}, \partial_{x^0}) = \mathbf{g}(\partial_{x'^0}, \Lambda^a{}_0 \partial_{x'^a}) = \mathbf{g}(\partial_{x'^0}, \Lambda^0{}_0 \partial_{x'^0}) = -\Lambda^0{}_0 \,,$$

so that $\Lambda^0{}_0 \geq 0$ and, taking (e) of proposition 8.25 into account $\Lambda^0{}_0 \geq 1$ which is equivalent to say that (Λ, T) belongs to the orthochronous subgroup. If also imposing the constraint $\det \Lambda = 1$, noticing that the intersection of subgroups is a subgroup, gives rise to the so called **orthochronous proper Poincaré group** and it is indicated by $ISO(1,3)\!\uparrow$

8.3.3 Pure and special transformations

From a kinematic point of view it makes sense to define the *four velocity* of $\mathscr{R}$ (which does *not* depend on $\mathscr{R}'$) and the *velocity* of $\mathscr{R}$ *with respect to* $\mathscr{R}'$ (which *does* depend on $\mathscr{R}'$) when they are connected by Poincaré transformation (Λ, T). The four velocity and the velocity turn out to be constant in space and time. Indeed, let $\mathscr{R}$ and $\mathscr{R}'$ be Minkowskian reference systems with associated co-moving Minkowskian coordinate frames x^0, x^1, x^2, x^3 and x'^0, x'^1, x'^2, x'^3 respectively and suppose that (8.17) hold. Let $\gamma_{\mathscr{R}}$ represent the world line of a material point *at rest* with $\mathscr{R}$, that is, $\gamma_{\mathscr{R}}$ projected to the rest space of $\mathscr{R}$ admits the trivial parametrization $x^\alpha(x^0) = x_0^\alpha$ constant for $\alpha = 1, 2, 3$. Since x^0 is the proper time of the history of any point at rest with $\mathscr{R}$, the four velocity of that point is nothing but $\dot{\gamma}_{\mathscr{R}} = \mathscr{R}$ itself. However it is interesting to write down the components of the four-velocity with respect to the other reference system in order to see how these components enters the coefficients of the Poincaré transform connecting the two reference system. From (8.17)

$$x'^a = \Lambda^a{}_b x^b + T^a$$

which implies

$$\partial_{x^b} = \Lambda^a{}_b \partial_{x'^a}$$

we obtain that, *referring to the coordinates* x'^0, x'^1, x'^2, x'^3, the components of the four-velocity

$$\dot{\gamma}_{\mathscr{R}} = \partial_{x^0}$$

are

$$\dot{\gamma}^0_{\mathscr{R}} = \Lambda^0{}_0\,, \quad \dot{\gamma}^\alpha_{\mathscr{R}} := \Lambda^\alpha{}_0\,, \quad \alpha = 1,2,3\,. \tag{8.25}$$

The result, due to definition 8.18, immediately implies that also the velocity $\mathbf{v}^{(\mathscr{R}')}_{\mathscr{R}}$ of $\gamma_{\mathscr{R}}$ with respect to $\mathscr{R}'$ does not depend on the chosen point at rest with $\mathscr{R}$ and it is constant in $\mathscr{R}'$-time. Its components in coordinates x'^1, x'^2, x'^3 ($c = 1$) turn out to be:

$$\mathbf{v}^{(\mathscr{R}')\alpha}_{\mathscr{R}} = \frac{\dot{\gamma}^\alpha_{\mathscr{R}}}{\dot{\gamma}^0_{\mathscr{R}}} = \frac{\Lambda^\alpha{}_0}{\Lambda^0{}_0}\,, \quad \alpha = 1,2,3\,. \tag{8.26}$$

Reversing the issue, it is interesting to wonder if the information included in the components – with respect to some $\mathscr{R}'$ – of a four velocity $\dot{\gamma}_{\mathscr{R}}$ is enough to construct a Lorentz transformation connecting $\mathscr{R}'$ with an inertial system $\mathscr{R}$ with that four velocity. Here a "four velocity" is just understood as future-oriented timelike vector normalized to -1 (or $-c^2$ depending on the units). The symbol $\dot{\gamma}_{\mathscr{R}}$ is just heuristic for the moment, since we do not know what $\mathscr{R}$ is.

However the answer is positive. Consider a unit timelike future-oriented vector $\dot{\gamma}$ whose componets are $\dot{\gamma}^0$ and $\vec{\dot{\gamma}}$ in Minkowskian coordinates x'^0, x'^1, x'^2, x'^3 of an Minkowski reference frame $\mathscr{R}'$, then (where we assume $c \neq 1$)

$$\Lambda_{\dot{\gamma}} = \left[\begin{array}{c|c} \dot{\gamma}^0/c & \vec{\dot{\gamma}}^t/c \\ \hline \vec{\dot{\gamma}}/c & I + \frac{\vec{\dot{\gamma}}\vec{\dot{\gamma}}^t}{c^2(1+\dot{\gamma}^0/c)} \end{array}\right], \tag{8.27}$$

this is a *Lorentz transformation* connecting a Minkowskian coordinate frame x^0, x^1, x^2, x^3 with the Minkowskian coordinate frame x'^0, x'^1, x'^2, x'^3:

$$x'^a = (\Lambda_\gamma)^a{}_b x^b + T^a\,. \tag{8.28}$$

Indeed, by direct inspection, one immediately proves that (exercise)

(i) $(\Lambda_{\dot{\gamma}})^t \eta \Lambda_{\dot{\gamma}} = \eta$,

(ii) $(\Lambda_{\dot{\gamma}})^0{}_0 \geq 0$ and $\det \Lambda_{\dot{\gamma}} > 0$,

(iii) $\partial_{x^0} = \dot{\gamma} =: \mathscr{R}$.

As a consequence of (iii), the Minkowskian coordinates x^0, x^1, x^2, x^3 are really co-moving with the initial reference system $\mathscr{R} = \dot{\gamma}$ as expected. We observe that the components T^a play no role in the discussion, in fact, they can always fixed to be $T^a = 0$ by changing the origin of the axis in $\mathscr{R}'$.

Lorentz transformations with the form (8.27), where $\dot{\gamma}$ is a unit future-oriented timelike vector represented in Minkowskian coordinates of a reference frame $\mathscr{R}'$ are said **pure Lorentz transformations**. When adding a spacetime translation as in (8.28), we obtain a **pure Poincaré transformation**.

A particular case is when $\vec{\gamma}_{\mathscr{R}}$ has only one components. If $\vec{\gamma}_{\mathscr{R}}$ has only one components along x'^3 and v is the only component of $\mathbf{v}_{\mathscr{R}}^{\mathscr{R}'}$, one immediately sees that (8.27) simplifies to the standard **special Lorentz transformation** along x^3:

$$\Lambda^{(3)} : \begin{cases} t' & = \frac{1}{\sqrt{1-v^2/c^2}} \left(t + \frac{vx^3}{c} \right) \ , \\ x'^1 & = x^1 \ , \\ x'^2 & = x^2 \ , \\ x'^3 & = \frac{1}{\sqrt{1-v^2/c^2}} \left(x^3 + vt \right) \ . \end{cases}$$

The product of two pure transformations is not a pure transformation as is easy to check. However in the special case of special transformations along an axis we have a subgroup of $SO(1,3)\!\uparrow$. The proof immediately arises from the fact that the above transforamtion can be re-written

$$\Lambda^{(3)}_\chi : \begin{cases} ct' = & ct\cosh\chi + x^3 \sinh\chi \ , \\ x'^1 = & x^1 \ , \\ x'^2 = & x^2 \ , \\ x'^3 = & x^3 \cosh\chi \, x^3 + ct\sinh\chi \ , \end{cases}$$

where we have introduced the so-called **rapidity** χ:

$$\mathbb{R} \ni \chi := \tanh^{-1}\frac{v}{c} \ , \quad v \in (-c, c) \ . \tag{8.29}$$

Taking advantage of the addition forms $\cosh(\chi'+\chi) = \cosh\chi\cosh\chi' + \sinh\chi\sinh\chi'$ and $\sinh(\chi'+\chi) = \sinh\chi\cosh\chi' + \cosh\chi\sinh\chi'$ one immediately proves that

$$\Lambda^{(3)}_\chi \Lambda^{(3)}_{\chi'} = \Lambda^{(3)}_{\chi+\chi'} \quad \chi, \chi' \in \mathbb{R}$$

which immediately proves that the set of special transformations along x^3 is a commutative subgroup of $SO(1,3)\!\uparrow$.

8.3.4 Are pure/special transformations and spatial rotation sufficient to renconstruct the whole Lorentz group?

From physical arguments we expect that, starting from a pure transformation $\Lambda_{\vec{\gamma}}$ connecting two reference frames, it is always possible to reduce to this case by suitably orienting the spatial axes at rest with the two reference systems. Actually this fact can be proved mathematically (and we will come back again to these sort of result in the next section). To this end, we observe that $SO(1,3)\!\uparrow := SO(1,3) \cap O(1,3)\uparrow$ includes $SO(3)$ as a subgroup simply represented by the subgroup of matrices

$$\Omega_R = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & R \end{array} \right] \ , \quad R \in SO(3) \tag{8.30}$$

(if $R \in O(3)$ the corresponding matrix Ω_R is an element of $O(1,3)\!\uparrow$ which therefore contains $O(3)$ as a subgroup.) The composition $\Omega_R \Lambda_{\dot\gamma} \Omega_R^t$ yields (setting again $c = 1$)

$$
\Omega_R \Lambda_{\dot\gamma} \Omega_R^t = \left[\begin{array}{c|c} \dot\gamma^0/c & (R\vec{\dot\gamma})^t/c \\ \hline R\vec{\dot\gamma}/c & I + \frac{R\vec{\dot\gamma}(R\vec{\dot\gamma})^t}{c^2(1+\dot\gamma^0/c)} \end{array}\right] , \tag{8.31}
$$

It is therefore obvious that, starting form a pure matrix of the form (8.31) and rotating the spatial axis of the reference system $\mathscr{R}'$ of a suitable rotation $R \in SO(3)$ we can always achieve a special Lorentz transform along x^3 proving the assertion above.
A final issue is whether or not the product $\Omega_R \Lambda_{\dot\gamma}$ exhausts all possibles elements of $SO(1,3)\!\uparrow$. The answer is positive and will be discussed in Chapter 9. We only state a minor version of the decomposition theorem which will be established in full generality in chapter 9 also taking the matrix Lie group structure into account.

Theorem 8.28. *If $\Lambda \in SO(1,3)\!\uparrow$ there exists a unique pair (Ω_R, P), where Ω_R is a spatial rotation of the form (8.30) and P a pure Lorentz transformation of the form (8.27), such that*

$$\Lambda = \Omega_R P \,.$$

A nice immediate consequence of this theorem and of (8.31) is that, whatever is the initial Poincaré transformation between a pair of Minkowskian reference frames $\mathscr{R}$ and $\mathscr{R}'$, we can always rearrange the used co-moving Minkowskian coordinate systems changing the space-time origin and rotation the spatial axes in order to have a system of co-moving Minkowskian coordinate systems connected by a special Lorentz transformation along the axis x^3.

8.3.5 Contraction of volumes

As an application of the explicit form of pure transformations provided in (8.27), we prove the formula of *contraction of volumes*.
Suppose that an extended body is at rest in a Minkowskian reference system $\mathscr{R}$ where it is defined by a region B of the rest space of $\mathscr{R}$. This region can be intepreted as a (Lebesgue) measurable set of $\mathbb{R}^3$ with respect to Minkowskian coordinates x^0, x^1, x^2, x^3 co-moving with $\mathscr{R}$. Let $Vol_{\mathscr{R}}(B)$ be the measure of that spatial region in the rest space of $\mathscr{R}$. Let us pass to the new reference system $\mathscr{R}'$. The points of $B'_{x'^0}$ are the intersections with the rest space of $\mathscr{R}'$ at time x'^0 of the world lines of the ponts in B evolving parallely to $\mathscr{R}$. We are interested in the voume of $B_{x'^0}$, now denoted by $Vol_{\mathscr{R}'}(B'_{x'^0})$, measured the other Minkowskian reference system $\mathscr{R}'$ equipped with co-moving Minkowskian coordinates x'^0, x'^1, x'^2, x'^3.

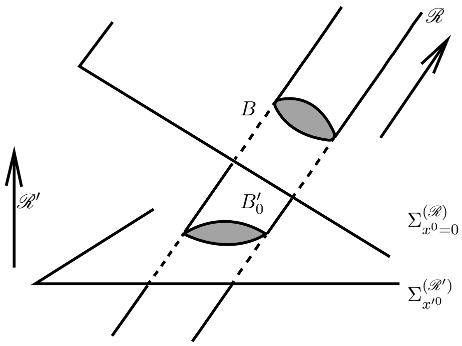

Figure 8.2: Contraction of volumes

Remarks 8.29. It is important stressing that these volumes are independent of the choice of Minkowskian coordinate systems respectively co-moving with $\mathscr{R}'$ and $\mathscr{R}'$ since the internal changes of Minkowskian coordinates just amount to spatial roto-translations which are isometries and therefore they preserve the Lebesgue measure. Similarly, a translation of the origin of the time axis just amounts to a corresponding translation in the index which labels the volumes B' without changing the values of them. ■

The relation between two different coordinate system respectively co-moving with $\mathscr{R}$ and $\mathscr{R}'$ is implemented by some $(\Lambda, T) \in IO(1,3)\uparrow$. As discussed above, we can always rotate the axis in $\mathscr{R}'$ in order to have Λ of the form (8.27). Furthermore, changing the origin of the temporal and spatial axes of $\mathscr{R}'$, we can always fix $T = 0$. In this case we can take advantage of the expression of Λ (8.27) in terms of the velocity $\mathbf{v}^{(\mathscr{R}')}_{\mathscr{R}}$ of $\mathscr{R}$ with respect to $\mathscr{R}'$. From (8.27) we have that

$$x'^0 = \dot{\gamma}^0_{\mathscr{R}} x^0 + \sum_{\alpha=1}^{3} \dot{\gamma}^\alpha_{\mathscr{R}} x^\alpha \tag{8.32}$$

and

$$x'^\beta = \dot{\gamma}^\beta_{\mathscr{R}} x^0 + x^\beta + \frac{\dot{\gamma}^\beta_{\mathscr{R}} \sum_{\alpha=1}^{3} \dot{\gamma}^\alpha_{\mathscr{R}} x^\alpha}{1 + \dot{\gamma}^0_{\mathscr{R}}}$$

In the rest space of $\mathscr{R}'$ and at time x'^0, the body B is therefore described by the set

$$B_{x'^0} := \left\{ (x'^1, x'^2, x'^3) \in \mathbb{R}^3 \;\middle|\; x'^\beta = \dot{\gamma}^\beta_{\mathscr{R}} x^0 + x^\beta + \frac{\dot{\gamma}^\beta_{\mathscr{R}} \sum_{\alpha=1}^{3} \dot{\gamma}^\alpha_{\mathscr{R}} x^\alpha}{1 + \dot{\gamma}^0_{\mathscr{R}}} \;\; , \;\; (x^1, x^2, x^3) \in B \right\}$$

where x^0 is obtained by inverting (8.32) as

$$x^0 = \frac{x'^0 - \sum_{\alpha=1}^{3} \dot{\gamma}^\alpha_{\mathscr{R}} x^\alpha}{\dot{\gamma}^0_{\mathscr{R}}}$$

and by inserting the result in the expression above for $B_{x'^0}$. In summary,

$$B_{x'^0} := \left\{ (x'^1, x'^2, x'^3) \in \mathbb{R}^3 \;\middle|\; x'^\beta = \frac{\dot\gamma^\beta_{\mathscr{R}}}{\dot\gamma^0_{\mathscr{R}}} x^0 + x^\beta - \frac{\dot\gamma^\beta_{\mathscr{R}} \sum_{\alpha=1}^3 \dot\gamma^\alpha_{\mathscr{R}} x^\alpha}{\dot\gamma^0_{\mathscr{R}}(1+\dot\gamma^0_{\mathscr{R}})} \;\;,\;\; (x^1, x^2, x^3) \in B \right\} .$$

We conclude that, for $\alpha, \beta = 1, 2, 3$,

$$J_\alpha^\beta = \frac{\partial x'^\beta}{\partial x^\alpha} = \delta^\beta_\alpha - \frac{\dot\gamma^\beta_{\mathscr{R}} \dot\gamma^\alpha_{\mathscr{R}}}{\dot\gamma^0_{\mathscr{R}}(1+\dot\gamma^0_{\mathscr{R}})} .$$

From standard results of the measure theory, using also the fact that $\det J$ is a constant,

$$Vol_{\mathscr{R}'}(B'_{x'^0}) = \int_{B'_t} 1 dx'^2 dx'^2 dx'^3 = \int_B |\det J| dx^1 dx^2 dx^3 = |\det J| \int_B dx^1 dx^2 dx^3 = |\det J| Vol_{\mathscr{R}}(B).$$

The determinant of the matrix $J := [J^\beta_\alpha]_{\alpha,\beta=1,2,3}$ can be immediately computed by observing that the simmetric matrix of components

$$-\frac{\dot\gamma^\beta_{\mathscr{R}} \dot\gamma^\alpha_{\mathscr{R}}}{\dot\gamma^0_{\mathscr{R}}(1+\dot\gamma^0_{\mathscr{R}})}$$

has an evident eigenvector $(\dot\gamma^1_{\mathscr{R}}, \dot\gamma^2_{\mathscr{R}}, \dot\gamma^3_{\mathscr{R}})^t$ with eigenvalue

$$\lambda = -\frac{\sum_\alpha (\dot\gamma^\alpha_{\mathscr{R}})^2}{\dot\gamma^0_{\mathscr{R}}(1+\dot\gamma^0_{\mathscr{R}})} = \frac{1-(\gamma^0_{\mathscr{R}})^2}{\dot\gamma^0_{\mathscr{R}}(1+\dot\gamma^0_{\mathscr{R}})} = \frac{1-\dot\gamma^0_{\mathscr{R}}}{\dot\gamma^0_{\mathscr{R}}}$$

and two vanishing eigenvalues corresponding to two mutually orthogonal vectors normal to $(\dot\gamma^1_{\mathscr{R}}, \dot\gamma^2_{\mathscr{R}}, \dot\gamma^3_{\mathscr{R}})^t$ in $\mathbb{R}^3$. Hence, the $3 \times x$ symmetric matrix J has the eigenvalues $1+\lambda$ and 1 with multiplicity 2. Restoring the value of the light speed, we conclude that

$$\det J = (1+\lambda) \cdot 1 = \frac{1}{\dot\gamma^0_{\mathscr{R}}} = \sqrt{1 - \frac{(\mathbf{v}^{(\mathscr{R}')}_{\mathscr{R}})^2}{c^2}} \,, \tag{8.33}$$

so that we have established the classical formula of the volume contraction,

$$Vol_{\mathscr{R}'}(B'_{x'^0}) = Vol_{\mathscr{R}}(B) \sqrt{1 - \frac{(\mathbf{v}^{(\mathscr{R}')}_{\mathscr{R}})^2}{c^2}} \,. \tag{8.34}$$

This formula is independent of the Minkowskian coordinate systems used in the two Minkowskian reference systems, since it only exploits and refers to intrinsic spetial metrical quantities as discussed above and uses the relative velocity $\mathbf{v}^{(\mathscr{R}')}_{\mathscr{R}}$, which does not depend on any choices of coordinates but only on $\mathscr{R}$ and $\mathscr{R}'$.

8.4 Dynamics I: material points

We briefly discuss here some basic notions useful in relativistic dynamics, always adopting a geometric point of view.

Remarks 8.30. *In this section we explicitly use the value of the light speed $c \neq 1$.* ■

8.4.1 Remarks on Relativistic Local Causality

Before entering the discussion of the formulation of relativistic dynamics a few remarks are necessary about the structure of the arena where that dynamics is formulated. In relativistic theories (macroscopic) interactions are described in terms of future-directed causal curves connecting the event p which is the cause with the event q which is the effect as we shall see shortly.

Definition 8.31. A couple of events p, q of (Minkowski) spacetime is said **causally connected** if $p = q$ or there is a future-directed causal curve joining them in any order. Events which are not causally connected are said to be **spatially-separated** or also **causally-separated**.

Events which are not causally connected *cannot* enjoy causal relations. This requirement is usually named **locality** or **local causality** or **Einstein causality** or **Einstein locality**. Minkowski spacetime possesses the light cone structure and this fact permits to state the definition above into an equivalent but more easy to use version.

Proposition 8.32. *In Minkowski spacetime, the events p and q are causally connected* if and only if *q belongs to the future half $\overline{V_p^+}$ of the closed light-cone emanated from p (so, including its origin and its boundary) or* vice versa.

Proof. If $q - p \in \overline{V^+}$ then $p = q$ or the segment from p to q is a future-directed causal curve. The same is valid swapping q and p. If $q - p \notin \overline{V^+}$ and $p - q \notin \overline{V^+}$, then we can arrange a reference frame $\mathscr{R}$ such that $q - p$ stays in a 3-rest space of $\mathscr{R}$ at time 0. A causal curve (if any) $x^a = x^a(s)$, $a = 0, 1, 2, 3$ with $s \in [a, b]$ joining p and q must satisfy $x^0(a) = x^0(b) = 0$ and thus there is a point $s_0 \in (a, b)$ where $dx^0/ds|_{s_0} = 0$. This is not possible because the curve is causal so that for some choice of the parameter $dx^0/ds \neq 0$ everywhere. □

Some important remarks on these notions are in order.

(i) It is worth stressing that quantum phenomena may give rise to non-causal *correlations* between two spatially-separated events as a consequence of the *Einstein-Podolski-Rosen paradox* (see, e.g., Chapter 5 of [Moretti-b]). These correlations have been experimentally observed and nowadays this type of phenomenology based on the notion of quantum *entanglement* is used in quantum information technology. The 2023 Nobel Prize in Physics has been awarded to Aspect, Shimony, and Zeilinger for the definite experimental proof of the entaglement phenomenon.

(ii) It is not difficult to prove that, if p and q are causally-separated, there are three inertial reference systems $\mathscr{R}$, $\mathscr{R}'$, and $\mathscr{R}''$ equipped with global time coordinates, respectively t, t', and t'', such that $t(p) < t(q)$, $t'(p) = t'(q)$, and $t'(p) > t'(q)$. These reference systems can be constructed starting from a third reference system $\mathscr{R}_0$ such that the segment joining p and q belongs to a rest space of it. As a consequence $t_0(p) = t_0(q)$ and thus $\mathscr{R}' = \mathscr{R}_0$. Next, placing $\mathscr{R}_0$ at the middle point between p and q, we can tilt it towards p or q, defining in this way $\mathscr{R}$ and $\mathscr{R}''$. Notice that we have in particular found the the fact that two spatially-separated events happen simultaneously or not depends on the reference frame used to describe them.

(iii) Conversely, if p, q are causally related, their temporal order *cannot* be reversed with a suitable choice of the Mninkowskian (=inertial) reference system. In fact, if $q \in \overline{V_p^+}$ (the other case is analogous), we have that the segment $\overrightarrow{pq}$ is either the zero vector or it is causal and future-directed and thus $\mathbf{g}(\mathscr{R}, \overrightarrow{pq}) \leq 0$, since also $\mathscr{R}$ is timelike and future-directed just by definition of inertial reference system. Therefore,

$$x^0(q) - x^0(p) = -\mathbf{g}(\mathscr{R}, \overrightarrow{pq}) \geq 0$$

for every global temporal coordinate x^0 of every inertial reference system $\mathscr{R}$.

(iv) The relativity of the chronological order outside the light come is in agreement with the previous remark that no causal relations between spatially separated events can be defined: here the time of the effect could preceed the time of the cause with a suitable choice of the reference system. However, this possibility cannot be used to justify the impossibility to define causal relation in view of the (potential) wrong time ordering. This is because a closer scrutiny to the physical construction of the Special Relativity shows that some conventional elements enter the game when the Einstein syncronization procedure is adopted making discussion on the causality much more subtle[2].

(v) It should be clear that the notion of *rigid body* is untenable in relativity: with a very long rigid rod I could give rise to causal relations between causally-separated events, simply pushing one edge.

8.4.2 The geometric nature of the Relativity Principle in the formulation of relativistic dynamics

As is well known, one of the pillars Einstein exploited to found his Special Relativity theory is the so-called *Relativity Principle*, which can be stated by postulating that *all physical laws can be formulated with the same mathematical form in all inertial reference systems.*
If the physical laws concern the geometry of spacetime, then they can be always formulated with the *intrinsic* and *synthetic* language of the geometry. In turn, we also know that the

[2] A wide discussion on these issues and a construction of the theory from physical requirments appears (in Italian) in [Moretti-d] and also (in English) in Chapter 15 of [Moretti-c].

structure of the (pseudo-)Euclidean spaces selects peculiar coordinate frames – the (pseudo-)orthonormal Cartesian systems of coordinates called Minkowskian coordinate systems – which are completely equivalent for the formulation of all geometric propositions written in coordinates and components: these statements have the same form in every Minkowskian coordinate system. This is a consequence of the fact that they are connected by means of isometries and the physical laws are invariant under the action of isometries.

This invariance-in-form property of the physical laws when represented in Minkowkian reference frames is often called *covariance* by physicists. However, covariance is more precisely related to the fact that the physical laws are written in terms of tensors and operations between tensors. Invariance of the physical laws (the Relativity principle) is a more general property which is not necessarily made explicit through the use of tensors.

In summary, if the physical laws reflect the geometry of the spacetime - i.e., if they have a geometric nature - they should take the same mathematical form in every Minkowskian system of coordinates when all intrinsic geometric objects are written in components[3]. As we have supposed that each Minkowskian system of coordinates defines (up to internal transformations) an inertial reference system, on assuming that the laws of physics are geometric propositions about the geometry of spacetime, we automatically have that the Relativity principle is fulfilled. For this reason, in the rest of the section, we shall formulate the laws of relativistic dynamics using the intrinsic language of the geometry of Minkowski spacetime.

8.4.3 Mass and four momentum

Special Relativity assumes that material points describing timelike worldlines are equipped with a (for the moment) strictly positive constant m called **mass** (or also **rest mass**). The idea is that as Special Relativity should reduce to standard mechanichs when the involved velocities are very smaller than c, then the Newtonian notion of mass should still have some corresponding tool in Special relativity[4]. More precisely the mass m of a material point coincides with its Newtonian mass since, by definition, it is measured in a Minkowskian reference system where the point has a very small velocity if compared with c.
If $\dot{\gamma}(\tau)$ is the four-velocity of the considered material point, the future-directed timelike vector

$$P(\tau) := m\dot{\gamma}(\tau) \tag{8.35}$$

is called **four momentum**.

Remarks 8.33. Exactly as the four-velocity, the four-momentum is *intrinsically defined*, namely without referring to a reference system. That is completely different from the classical notions of velocity and momentum which are defined only when a reference system is given. ■

[3]When the said laws are written in the language of tensor calculus, the fact that the form of the laws in components is identical in all Minkowskian system of coordinates is known as *covariance of the physical laws*.

[4]This sort of arguments are often called "correspondence principle" and play a crucial role in heuristically developing the theory.

From (8.11), we immediately have that, in the reference system $\mathscr{R}$ endowed with co-moving Minkowskian coordinates x^0, x^1, x^2, x^3, the components of P with respect to those coordinates read (restoring the constant c and observing that P is future-directed)

$$0 < P^0_{\mathscr{R}} = \frac{mc}{\sqrt{1 - \left(\frac{\mathbf{v}_{\mathscr{R}}}{c}\right)^2}}\,, \qquad P^\alpha_{\mathscr{R}} = \frac{m\mathbf{v}^\alpha_{\mathscr{R}}}{\sqrt{1 - \left(\frac{\mathbf{v}_{\mathscr{R}}}{c}\right)^2}}\,, \quad \alpha = 1, 2, 3\,, \tag{8.36}$$

where $\mathbf{v}_{\mathscr{R}}$ is the velocity of the material point with respect to $\mathscr{R}$ at the considered event. According to (8.36), if $\|\mathbf{v}_{\mathscr{R}}\| \ll c$, the spatial components of the four momentum in a given Minkowskian reference system $\mathscr{R}$ become an approximation of the components of the classical momentum, since

$$\frac{m\mathbf{v}^\alpha_{\mathscr{R}}}{\sqrt{1 - \left(\frac{\mathbf{v}_{\mathscr{R}}}{c}\right)^2}} = m\mathbf{v}^\alpha_{\mathscr{R}} + O\left(\left(\frac{\mathbf{v}}{c}\right)^2\right)\,.$$

This fact provides a physical intepretation of the spatial components of the four momentum: they are a generalization of the classical momentum. Instead, for the moment the meaning of P^0 is not clear in comparison with classical physics. From (8.36), we however see that the mass coincides with the component P^0 of the four momentum measured in a inertial rest frame at rest (possibly just *istantaneously*) with the particle.

8.4.4 Massless particles

A useful pair of relations immediately arise from (8.36)

$$\mathbf{g}(P, P) = -m^2c^2 \quad \text{i.e., within the abstract index notation} \quad P_aP^a = -m^2c^2\,. \tag{8.37}$$

and

$$P^0 = c\sqrt{\sum_{\alpha=1}^{3} P^\alpha P^\alpha + m^2c^2}\,. \tag{8.38}$$

Equation (8.37) provides a definition of the mass m in case we assume P as a primitive notion. From this viewpoint, namely assuming P as primitive, it seems possible to extend the notion of four momentum to particles satisfying (8.37) with $m = 0$. These particles have ligktlike momentum and are a complete novelty, since they are not permitted in classical physics. It is worth stressing that experimental physics proved that they really exist.

Examples 8.34. *Photons* are an example of massless particles. In that case, in a semiclassical (non-quantum) view

$$P^0 = \hbar\frac{\omega}{c}\,, \quad P^\alpha = \hbar k^\alpha\,, \quad \alpha = 1, 2, 3,$$

where ω is the angular frequency of the light wave associated with the photon and k^α the components of the wave vector in the considered reference system, $\hbar = \frac{h}{2\pi}$ with h the Planck constant ($6.626 \times 10^{-27} erg\, sec$). Since

$$\sum_{\alpha=1}^{3} (k^\alpha)^2 = \left(\frac{\omega}{c}\right)^2$$

from the electromagnetic theory, we have in this case $P_a P^a = 0$, so that *photons must be massless.* ■

Remarks 8.35. Since P is tangent the the worldlines of massless particles, proposition 8.19 implies that massless particles always move with the speed of light. In this case there is no rest frame with the particle since P is lightlike. However (8.38) is still valid with obviously $m = 0$. ■

8.4.5 The principle of inertia

We remind the reader that, if $\mathbb{A}^n$ is an affine space with space of translation V^n, then an **affine segment** is a map of the form

$$I \ni t \mapsto P + t\mathbf{v} \in \mathbb{A}^n =: P(t) , \tag{8.39}$$

where $I \subset \mathbb{R}$ is an interval, $P \in \mathbb{A}^n$ and $\mathbf{v} \in V^n$. The curve above can be reparametrized using another parameter s – defined on another interval I' and related with t through a C^1 map with C^1 inverse – obtaining a new map $P'(s) := P(t(s))$ with the same form as (8.39)

$$I' \ni s \mapsto P' + s\mathbf{v}' \in \mathbb{A}^n = P'(s) ,$$

for suitable P', and $\mathbf{v}'$. This happens if and only if $s = at + b$ for $a \neq 0$ and $b \in \mathbb{R}$ arbitrary constants. All the parameters preserving the form (8.39) are called **affine parameters** of the affine segment.

In classical physics, an *isolated* material point is assumed to evolve with constant velocity in every inertial reference system, i.e., in every Minkowskian reference system. When assuming $m > 0$, from (8.11), we immediately have that histories with constant velocity in a given Minkowskian reference system are actually curves with constant four-velocity $\dot{\gamma}$ so that, in particular, they have constant velocity with respect to every other Minkowskian reference system. Hence, when parametrizing those curves using the proper time τ, we find

$$\gamma(\tau) = \gamma(\tau_0) + (\tau - \tau_0)\dot{\gamma} ,$$

where $\dot{\gamma}$ is a constant timelike future-oriented vector. Those curves are nothing but *affine segments* in $\mathbb{M}^4$ and the proper time is an *affine parameter* for them. We can extend these assumptions for *isolated* massless particles evolving along lightlike worldlines, assuming that also in this case those lines are lightlike segments,

$$\gamma(s) = \gamma(s_0) + (s - s_0)\gamma' .$$

The difference is that now there is no a preferred affine parameter describing them as the proper time of massive particles, since the integrand in (8.7) vanishes. The above affine parameter s can be changed with affine transformations $s' := as + b$ where $a > 0$ obtaining another physically equivalent (future-directed) affine parametrization. In summary, the principle of inertia in Minkowski spacetime states that

Principle of inertia. *Isolated material points have worldlines which are causal future oriented affine segments of* $\mathbb{M}^4$. *They are timelike, with the proper time as an affine parameter, or light-like according to the value of their masses, strictly postive or vanishing respectively.*

Remarks 8.36. A substantial difference between this formulation of the inertia principle and the one of classical physics is that here the notion of reference frame plays no role, since the formulation is now geometrically intrinsic.

8.4.6 Four momentum conservation law

In Special Relativity, as in calssical mechanics, it seems to be palusible to assume valid the conservation of the total four momentum of a system of material points, when these points form an isolated system but they interact each other. This principle is harder to be formulated than in classical mechanics in view of the difficult notion of total four momentum of a system of interacting material points: *a priori*, the total four momentum is computed summing the momenta of the various particles forming the system on a given time slice of a Minkowskian reference system. However, the notion of "at a given time" depends on the reference system and changing reference system we would sum the momenta of the same particles computed at different events. Conversely, the total four momentum is expected to be an intrinsic, reference-independent, notion. Dealing with material points, in a particular case the principle can be however formulated. We are referring to the case of mutual intractions *happening at isolated events*. In this case, the worldlines of the particles of the system define a graph whose lines are causal future-oriented affine segments of $\mathbb{M}^4$, according to the principle of inertia since the material points are isolated between interactions they undergo, and the vertex are the very interaction events. The conservation of total four momentum is stated as follows.

Four momentum conservation. *For a system of material points whose mutual interactions are concentrated in isolated events (and no further interactions exist), the wordlines of the material points of the system form a graph whose lines are causal future-oriented segments and the vertices are the interaction events. In this graph,*

(i) the four momentum is constant along each line,

(ii) the sum of the four momenta entering a vertex from the past is equal to the sum of the four momenta exiting that vertex towards the future.

With this postulate an invariant notion of *total four momentum* can be defined at least when the number of interaction events and lines is finite. Indeed, every rest space $\Sigma^{(\mathscr{R})}_{x^0}$ of every

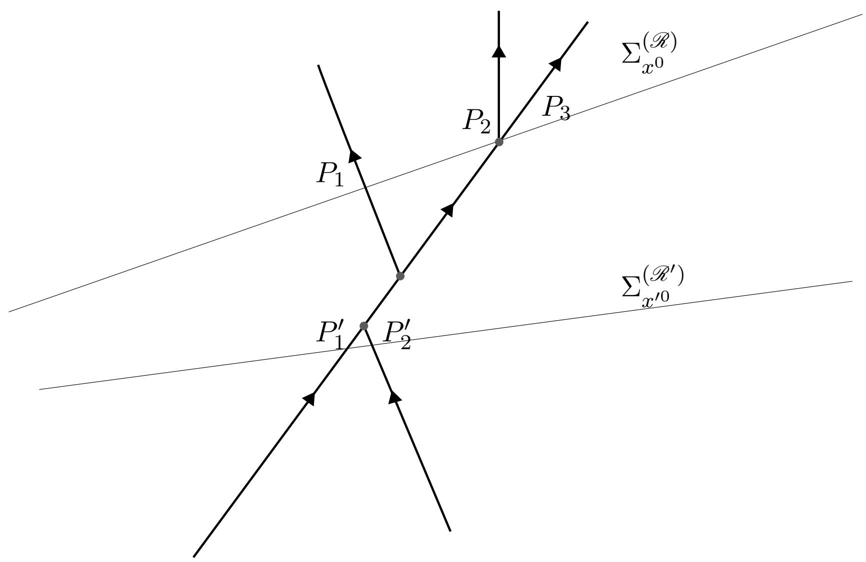

Figure 8.3: Four momentum conservation: $P'_1 + P'_2 = P_1 + P_2 + P_3$

Minkowskian reference system $\mathscr{R}$ is such that the sum of the four momenta $P_1, P_2, \ldots, P_N$ of the worldlines crossing $\Sigma^{(\mathscr{R})}_{x^0}$ (notice that the number N generally depends on $\Sigma^{(\mathscr{R})}_{x^0}$) has a constant value $P = \sum_{i=1}^{N} P_i$ which is independent of *both* x^0 and $\mathscr{R}$. The proof of this fact is not completely straightforward and it is left to the reader as an exercise (use an inductive argument). A crucial observation is that we cannot choose $\Sigma^{(\mathscr{R})}_{x^0}$ that completely includes a segment of the graph, it only may cross each this line in a single event at most. That is because these segments are causal, whereas every segment which is embedded in $\Sigma^{(\mathscr{R})}_{x^0}$ is necessarily spacelike.

Remarks 8.37.
(1) $P = \sum_{i=1}^{N} P_i$ makes sense, even if the momenta P_i are located at different events, just because we are dealing with an *affine space*. However this sort of approach will become problematic once one relaxes the affine structure to that of Lorentzian manifold in General Relativity.
(2) The requirement of conservation of the four momenta as written in the postulate above can be assumed more generally valid when the worldlines entering a vertex and those exiting that vertex become affine segments in an arbitrarily small neighborhood of the vertex where the possible external interaction switch off. In this case, obviously, the four momenta to be considered are computed exactly at the vertex along each worldline. In this more general situation however the notion of total four momentum is more difficult to define. ■

Exercises 8.38.

1. Prove that, even if mass is not conserved, no particles can be constructed out of the vacuum. (*Hint.* If the total final four momentum P must be conserved and its temporal cmponent must be strictly positive because future-oriented.)

2. Prove that a photon cannot decay into a pair of massive particles. (*Hint.* Write down the law of the conservation of the four momentum in components in the reference system $\mathscr{R}_0$ that is parallel to the total conserved 4-momentum P. Notice that P must be timelike because the 4-momenta of the final particles are timelike.)

8.4.7 The interpretation of P^0 and the so-called "Mass-Energy equivalence"

The above formulation of the conservation law of the four momentum permits processes where the number of particles is not conserved, whereas the total momentum is conserved. The masses of the involved particles generally change crossing a vertex. The masses are however constant (possibly vanishing) along the lines due to (8.37). A natural issue in this context is whether or not the total mass of the material points is conserved during the interactions as it happens in classical physics. In particular one may wonder if in processes where particles are created or destroyed – a particle breaks into many particles or many particle join into a single particle – the total mass remains constant. The answer to this question on the one hand suggests an intepretation of the component P^0 of the four momentum, on the other hand it leads to one of the fantastic achievents of modern physics due to Einstein.

Considering the case of a particle decaying into two particles, if P is the four momentum of the initial particle with mass M and P_1, P_2 are the momenta of the two final particles with masses m_1 and m_2 respectively, in the reference system of the initial particle, the conservation law of the four momentum gives (we omit the index $\mathscr{R}$ for the sake of semplicity)

$$Mc = P_1^0 + P_2^0 \,, \quad 0 = P_1^\alpha + P_2^\alpha \,, \quad \alpha = 1,2,3 \,,$$

that is, regarding the first identity,

$$M = \frac{1}{c}P_1^0 + \frac{1}{c}P_2^0$$

which can be rearranged into

$$M = m_1 + m_2 + \left(\frac{1}{c}P_1^0 - m_1\right) + \left(\frac{1}{c}P_2^0 - m_2\right) . \tag{8.40}$$

We see that the mass is *not* conserved (more precisely it is not additive) and the failure to be conserved is represented by the terms in parenthesis which are however very small as the velocities of the two particle are small in comparison with c. In fact, from the former in (8.36)

$$\left(\frac{1}{c}P_1^0 - m_1\right) + \left(\frac{1}{c}P_2^0 - m_2\right) = \frac{1}{c^2}\left[\frac{1}{2}m_1\mathbf{v}_1^2 + \frac{1}{2}m_2\mathbf{v}_2^2 + O\left(\left(\frac{\mathbf{v}_1}{c}\right)^2\right) + O\left(\left(\frac{\mathbf{v}_2}{c}\right)^2\right)\right] . \tag{8.41}$$

We see that the initial mass M is equal to the sum of the final masses with a further contribution due to the classical kinetic energies of the two particles in the reference system of the intial decaying particle and further terms of order v^2/c^2. Expansion (8.41) together with other theoretical results (see Sect.8.4.9) lead to define the **relativistic kinetic energy** in the reference system $\mathscr{R}$ of a particle of mass m and four momentum P as

$$K_{\mathscr{R}} := cP^0_{\mathscr{R}} - mc^2 \,, \tag{8.42}$$

where P^0 is the temporal component of the four momentum P in the reference system $\mathscr{I}$. On the other, it also leads to re-write (8.40) as

$$M = m_1 + m_2 + \frac{K_1}{c^2} + \frac{K_2}{c^2} \,. \tag{8.43}$$

The analogous result pops out when describing the time reversed phenomenon where two particles melt into a unique particle: part of the mass M of the final particle, in addition to the masses m_1 and m_2 is due to the kinetical energies (in the reference system of the particle) of the two colliding particles.

Remarks 8.39. We stress that the discussed phenomena are really observed quite easily when dealing with elementary particles: the measured mass really ceases to be an additive quantity and it receives contributions from the kinetic energy of decay products according to (8.43) or similar equations, where all terms can be experimetally measured separately. ■

This phenomen where the mass of a particle seems to receive a contribution from the kinetic energies of constituents or of the products of decay is corroborated from anoter completely theoretical point of view, if we assume that the mass of a system is c^{-1} times the value of the temporal component of its four momentum P in its co-moving reference system $\mathscr{R}_0$, i.e., in a reference system defined by a normalized vector $\mathscr{R}_0$ parallel to P. (This is equivalent to saying that the three spatial components $P^\alpha_{\mathscr{R}_0}$ vanish in that reference system.) In fact, referring to the system of material points intercting only in isolated events we discussed in the previous section, if $P = \sum_{i=1}^N P_i$ is the total four momentum computed in the co-moving reference system $\mathscr{R}_0$ whose associated rest space intersects N world lines at the considered time), we have in that reference system

$$M := \frac{1}{c} P^0_{\mathscr{R}_0} = \sum_{i=1}^{N} \frac{1}{c} P^0_{\mathscr{R}_0\, i} = \sum_{i=1}^{N} m_i + \frac{K_{\mathscr{R}_0\, i}}{c^2} \,, \qquad P^\alpha = 0 \,, \alpha = 1, 2, 3 \,.$$

Above, m_i and $K_{\mathscr{R}_0\, i}$ are respectively the mass and the kinetic energy of the i-th material point in the co-moving reference system of the whole system at the cosidered time. The system is a "big" material point, at rest in the considered reference system, with a mass which is the sum of the masses of its constituents and a further contribution due to the kinetic energy of them in the said reference system. If such a system of molecules were confined in a box, we would expect that its mass incrases if warming up the box as a consequence of the fact that the temperature is nothing but a macroscopic expression of the (here only kinetic) energy of the molecules.

All that discussion, in particular the fact that $cP^0_{\mathscr{R}}$ is conserved in time and reduces to the kinetic energy in the classical limit, up to an additive constant, lead to the following final postulate which implies the celebrated *mass-energy equivalence principle*.

Energy. *If $P^0_{\mathscr{R}}$ is the time component of the total momentum of a material point in a Minkowskian reference system $\mathscr{R}$, then*

$$E_{\mathscr{R}} := cP^0_{\mathscr{R}}$$

is the mechanical energy of the point in the said reference system.

That physical postulate has several implications we list below.

(i) The mechanical energy is conserved in every Minkowskian reference system when the hypotheses of the principle of conservation of the four momentum are satisfied.

(ii) If the mass of the system does not vanish, then it is

$$m = \frac{1}{c^2} E_{\mathscr{R}_0} ,$$

where $\mathscr{R}_0$ is the reference system co-moving with the system, defined by the condition that $\mathscr{R}_0$ is parallel to P;

(iii) (**Mass-energy equivalence principle**) The mass of a system is not an additive (i.e. conserved) quantity and it contributes to the total mechanical energy $E_{\mathscr{R}}$ of the system, in a given reference system $\mathscr{R}$, with the amount

$$\Delta E_{\mathscr{R}} = mc^2 .$$

(iv) The kinetic energy,

$$K_{\mathscr{R}} = cP^0_{\mathscr{R}} - mc^2 ,$$

in a reference system $\mathscr{R}$ just amounts the the difference of the total mechanical energy $P^0_{\mathscr{R}}$ – which depends on the reference system – and the contribution due to the mass mc^2 – which is independent. In other words, the mechanical energy is the sum of the kinetic energy and a contribution due to the mass according to (iv):

$$P^0_{\mathscr{R}} = mc^2 + K_{\mathscr{R}} .$$

The postulate and the remarks above are extended to all physical systems not necessarily material points. The reason is that, macroscopically speaking, every extend but spatially confined physical system can be described as a material point if dealing with spatial scales much larger than its tipical size. From a general perspective, it is expected that the mass m of a body is nothing but the total amount of energy which can be localized inside the body, up to the conversion factor c^2, when it is measured in its rest frame, provided this reference frame can be defined. This energy, which appears to be a part of the total machanical energy, actually accounts for all types of energies (e.g. chemical, thermodynamical, etc.) which can be localized in the region of the body. Physical transformations may transform the energy ascribed to the mass into different forms of energy and *vice versa*, in particular to kinetic energy. This possibility implies that the mass is not conserved.

8.4.8 The notion of four force

A theoretically interesting attempt to describe interactions which are not localized in events has been proposed in terms of a generalized version of *Newton's second law*. However this approach is quite academic and the only concrete case where it is useful is when considering charged particles in interaction with the electromagnetic field. The idea, restricting ourselves to treating the case of a massive particle only, is to replace Newton's law with

$$F(\gamma(\tau), \dot{\gamma}(\tau)) = \frac{dP}{d\tau}\,, \quad P(\tau) = m(\tau)\dot{\gamma}(\tau)\,, \tag{8.44}$$

where $P = P(\tau)$ is the four momentum at proper time τ of a particle with worldline $\gamma = \gamma(t)$ and the given **four force** $F(p, S_p)$ defines a vector at $T_p\mathbb{M}^4$ for every $p \in \mathbb{M}^4$ and $S_p \in T_p\mathbb{M}^4$ with $\mathbf{g}(S_p, S_p) = -c^2$. It is worth stressing that the above system of differential equations includes the possibility of changing the mass m of the particle due to the interaction described by the four force F. In fact, since

$$m^2 = -\frac{1}{c^2}P_a P^a$$

we also have

$$\frac{dm^2}{d\tau} = -\frac{2}{c^2}P_a\frac{dP^a}{d\tau} = -\frac{2}{c^2}P_a F^a = -\frac{2}{c^2}\mathbf{g}(P, F)\,.$$

Hence, the mass of the particle is constant in (proper) time if and only if F is always orthogonal the the four momentum. A four force satisfying

$$\mathbf{g}(S_p, F(p, S_p)) = 0 \quad \text{for all } p \in \mathbb{M}^4 \text{ and } S_p \in T_p\mathbb{M}^4 \text{ with } \mathbf{g}(S_p, S_p) = -c^2$$

is said to be **mechanical**. A discussion on the existence an uniqueness Cauchy problem for four-forces appears in [Moretti-c].

A fundamental case of mechanical four force is the Lorentz four force acting on a charged particle with mass m, charge q (which is invariant as the mass) and worldline γ, which has the relativistic expression

$$F^a(\gamma(\tau), \dot{\gamma}(\tau)) = q\,\dot{\gamma}_b(\tau)\,F^{ab}\left(\gamma(\tau)\right)\,, \tag{8.45}$$

Above, the antisymmetric tensor $\mathbb{M}^4 \ni p \mapsto F^{ab}(p)$ is the **electromagnetic tensor field** which accounts for the electric and magnetic fields in every Minkowskian reference system

$$[F^{ab}_{\mathscr{R}}]_{a,b=0,1,2,3} = \left[\begin{array}{c|ccc} 0 & E_x/c & E_y/c & E_z/c \\ \hline -E_x/c & 0 & B_z/c & -B_y/c \\ -E_y/c & -B_z/c & 0 & B_x/c \\ -E_z/c & B_y/c & -B_x/c & 0 \end{array}\right] \tag{8.46}$$

where E^α and B^β are the component of the electric and magnetic field in the reference frame $\mathscr{R}$. A direct computation proves that

$$F^0 = \frac{q\mathbf{E}\cdot\mathbf{v}_{\mathscr{R}}}{\sqrt{1-\frac{\mathbf{v}^2_{\mathscr{R}}}{c^2}}}\,, \qquad F^\alpha = \frac{qE^\alpha}{\sqrt{1-\frac{\mathbf{v}^2_{\mathscr{R}}}{c^2}}} + \frac{q}{c}\left(\frac{\mathbf{v}_{\mathscr{R}}}{\sqrt{1-\frac{\mathbf{v}^2_{\mathscr{R}}}{c^2}}}\times\mathbf{B}\right)^\alpha \qquad \alpha = 1,2,3\,,$$

where the spatial components reduce to the standard classical formula when $\|\mathbf{v}_{\mathscr{R}}\| \ll c$, whereas the temporal components tends to the classical power of Lorentz' force.
Antisimmetry of F^{ab} is responsible for the conservation of m during the evolution of the particle:

$$q^{-1}F^a P_a = F^{ba}\dot\gamma_b P_a = mF^{ba}\dot\gamma_a\dot\gamma_b = mF^{ba}\dot\gamma_b\dot\gamma_a = -mF^{ba}\dot\gamma_b\dot\gamma_a = -mF^{ba}\dot\gamma_a\dot\gamma_b = 0\,.$$

8.4.9 The relativistic *vis viva* equation

A mechanical four force satiafies the relativistic version of the relation connecting kinetical energy and power, known as the *vis viva theorem* in classical mechanics. Indeed, if a four force acts on a material particle of (constant) mass $m > 0$, the mechanical-force contraint reads in components of a Minkowskian reference system $\mathscr{R}$ where $\dot\gamma(\tau) = \dot\gamma^0_{\mathscr{R}}\partial_{x^0} + \sum_{\alpha=1}^3 \dot\gamma^\alpha_{\mathscr{R}}\partial_{x^\alpha}$ is the four velocity of the particle,

$$\sum_{\alpha=1}^3 \dot\gamma^\alpha_{\mathscr{R}} F^\alpha_{\mathscr{R}} = \dot\gamma^0_{\mathscr{R}} F^0_{\mathscr{R}}\,,$$

namely

$$\sum_{\alpha=1}^3 \frac{v^\alpha_{\mathscr{R}}}{\sqrt{1-\left(\frac{\mathbf{v}_{\mathscr{R}}}{c^2}\right)^2}} F^\alpha_{\mathscr{R}} = \frac{c}{\sqrt{1-\left(\frac{\mathbf{v}_{\mathscr{R}}}{c^2}\right)^2}} F^0_{\mathscr{R}}$$

which implies

$$\sum_{\alpha=1}^3 v^\alpha_{\mathscr{R}} F^\alpha_{\mathscr{R}} = cF^0_{\mathscr{R}} = \frac{dcP^0_{\mathscr{R}}}{d\tau} = \frac{d(cP^0_{\mathscr{R}} - mc^2)}{d\tau} = \frac{dK_{\mathscr{R}}}{d\tau} \tag{8.47}$$

where we have used (8.42) noticing that m is constant. In summary, we have found that

$$\sum_{\alpha=1}^3 v^\alpha_{\mathscr{R}} F^\alpha_{\mathscr{R}} = \frac{dK_{\mathscr{R}}}{d\tau}\,. \tag{8.48}$$

Since, for $x^0 = ct$,

$$\frac{d}{d\tau} = \frac{dt}{d\tau}\frac{d}{dt} = \frac{1}{\sqrt{1-\left(\frac{\mathbf{v}_{\mathscr{R}}}{c^2}\right)^2}}\frac{d}{dt} = \left(1 + O\left(\frac{\mathbf{v}_{\mathscr{R}}}{c^2}\right)^2\right)\frac{d}{dt}$$

(8.48), for $\|\mathbf{v}_{\mathscr{R}}\| \ll c$, s reduces to the standard result connecting the power of the total force acting on a material point and the derivative of the Kinetical energy known as *vis viva equaion*. On the other hand this identity corroborares the idea that $K_{\mathscr{R}}$ defined in (8.42) can be intepreted as the kinetical energy in the relativistic realm.

Remarks 8.40. As a by product of (8.47) above, we see that the temporal component $F^0_{\mathscr{R}}$ of a mechanical four force, up to the factor $1/c$, always coincides with the power of the spatial component of the four force according to the classical definition,

$$F^0_{\mathscr{R}} = \frac{1}{c}\sum_{\alpha=1}^3 F^\alpha_{\mathscr{R}} v^\alpha_{\mathscr{R}}.$$

We stress that this result is valid for mechanical four forces only. Also observe that in the classical regime $v \ll c$, $F^0_{\mathscr{R}}$ is therefore supposed to be negligible. This fact corresponds to the common classical evidence of the validity of the mass conservation law. ■

8.5 Dynamics II: the stress-energy tensor for macroscopic systems

This section is devoted to the introduction of a fundamental tool in relativistic theories, the so called *stress-energy tensor*. In field theory and particle physics, that mathematical object arises within the variational approach of the dynamics as a byproduct of *Noether's theorem*, when assuming that the Lagrangian density of the system is translationaly invariant in space and time of Minkowski spacetime. Not all physical systems – especially the macroscopic ones – admit a variational formulation and for that reason we adopt here a more general viewpont to justify the introduction of that matematical tool in the formalism.

8.5.1 The non-interacting gas of particles

We pass to focus attention on continuous systems and the most elementary one is a gas of massive non-interacting particle. This system is described by two ingredients,

(i) a C^1 assignment of timelike wordlines whose four velocities define a C^1 timelike future-directed vector field $\mathbb{M}^4 \ni p \mapsto V(p) \in T_p\mathbb{M}^4$ in Minkowski spacetime, everywhere satisfying $V^aV_a = -c^2$. V is the **four-velocity field** of the continuous body we are studying

(ii) a C^1 scalar field $\mu_0 : \mathbb{M}^4 \to [0, +\infty)$ representing the mass in the local rest frames of the particle of the gas. This rest frame is defined at each event p by $\mathscr{R}_0 := \frac{1}{c^2}V(p)$. μ_0 is the **intrinsic density of mass** of the continuous body we are studying.

Heuristically, we can think of $\mu_0(p)$ as the ratio

$$\mu_0(p) = \frac{\delta m_p}{\delta v_{0p}} ,$$

where δv_{0p} is the volume of a small portion of the gas in the rest space $\Sigma^{(\mathscr{R}_0)}_{x^0(p)}$ and δm_p is the total mass of the particles included in that volume: the particles in the said small volume can be considered at rest in $\mathscr{R}_0$. This heuristic (but strongly physically motivated!) view leads to the following requirement on the scalar field μ_0. If we fix a Minkowskian reference frame $\mathscr{R}$ ($\neq \mathscr{R}_0$ in general) and indicate by $\delta v^{(\mathscr{R})}_{0p}$ the volume occupied by the considered small portion of particles measured in $\mathscr{R}$. According to (8.34), the density of mass $\mu_{\mathscr{R}}$ referred to the gas and measured at rest with $\mathscr{R}$ must be,

$$\mu_{\mathscr{R}} = \mu_0 \frac{\delta v_{0p}}{\delta v^{(\mathscr{R})}_{0p}} = \mu_0 \frac{1}{\sqrt{1 - \frac{\|\mathbf{v}_p\|^2}{c^2}}} , \tag{8.49}$$

where $\mathbf{v}_p$ is the velocity associated to the four velocity $V(p)$ in the rest space of $\mathscr{R}$. Observing that

$$\frac{1}{\sqrt{1-\frac{\|\mathbf{v}_p\|^2}{c^2}}} = \frac{1}{c}V^0_{\mathscr{R}} = -\frac{1}{c}\mathbf{g}(V,\mathscr{R}),$$

we find

$$\mu_{\mathscr{R}}(p) := -\frac{\mu_0(p)}{c}\mathbf{g}(V,\mathscr{R})\,. \tag{8.50}$$

We henceforth assume (8.50) as a the *definition* of the **relative density of mass**. For future convenience, we introduce the covariant unit vector,

$$n^{(\mathscr{R})} := -\mathbf{g}(\mathscr{R},\ \ ) \qquad \text{namely} \quad n^{(\mathscr{R})}_a := -g_{ab}\mathscr{R}^b\,.$$

Hence the definition above can be re-phrased to

$$\mu_{\mathscr{R}}(p) := \frac{\mu_0(p)}{c}\langle V(p), n^{(\mathscr{R})}\rangle\,. \tag{8.51}$$

Remarks 8.41. Notice that introducing a Minkowskian system of coordinates x^0, x^2, x^2, x^3 co-moving with $\mathscr{R}$, we immediately have that

$$\mathscr{R} = \partial_{x^0}\,, \qquad n^{(\mathscr{R})} = dx^0\,. \tag{8.52}$$

We shall take advantage of those identities shortly. ■

8.5.2 Mass conservation law in local form for the non-interacting gas

Consider a spatial portion of continuum $B_0 \in \Sigma^{(\mathscr{R})}_0$ at time $t=0$ in a Minkowskian system of coordinates $x^0 = ct, x^1, x^2, x^3$ comoving with an inertial reference system $\mathscr{R}$. B_0 is assumed to be here a generic sufficiently regular measurable set of $\mathbb{R}^3$, for instance a coordinate 3-ball. If V is C^1, its evolution through the *local flow*[5] ϕ_t of V, gives rise to a measurable regular set $B_t := \phi_t(B_0) \in \Sigma^{(\mathscr{R})}_t$ as well since this set is the image of B_0 through a C^1 diffeomorphism. On the physical side, *since we are assuming that the particles do not interact and evolve freely*, we must admit that the mass is conserved. Hence,

$$\int_{\Sigma^{(\mathscr{R})}_t \cap B_t} \mu_{\mathscr{R}}(ct, x^1, x^2, x^3)dx^1dx^2dx^3 = \int_{\Sigma^{(\mathscr{R})}_0 \cap B_0} \mu_{\mathscr{R}}(0, x^1, x^2, x^3)dx^1dx^2dx^3\,, \tag{8.53}$$

for every choice of B_0. We intend to write that requirement into a local form. To this end observe that (8.51) permits to rephrase the identity above as, for $p \equiv (ct, x^1, x^2, x^3)$ and $p_0 \equiv (0, x^1, x^2, x^3)$

$$\int_{\Sigma^{(\mathscr{R})}_t \cap B_t} \mu_0(p)\langle V(p), n^{(\mathscr{R})}\rangle dx^1dx^2dx^3 + \int_{\Sigma^{(\mathscr{R})}_0 \cap B_0} \mu_0(p_0)\langle V(p_0), -n^{(\mathscr{R})}\rangle dx^1dx^2dx^3 = 0\,.$$

[5] We assume that flow of V is complete (and this happens in particular if the C^1 vector field V has bounded components in Cartesian coordinates of $\mathbb{M}^4$) or that the interval $[0,t]$ and B_0 are sufficiently small to guarantee that $B_t \in \Sigma^{(\mathscr{R})}_t$ exists.

In turn, this identity has a geometric intepretation with respect to the Euclidean geometry of the space $\mathbb{R}^4$ where the coordinates of $\mathscr{R}$ range. In fact

(1) the components of $n_i^{(\mathscr{R})} = \delta_i^0$ form the $\mathbb{R}^4$ vector normal to the surfaces $x^0 = 0$ and $x^0 = ct$ and this vector is normalized *with respect to the standard Euclidean metric of* $\mathbb{R}^4$;

(2) the pairing $\langle V(p), n^{(\mathscr{R})} \rangle$ is nothing but the standard $\mathbb{R}^4$ scalar product of the vector $(V^0_{\mathscr{R}}(p), V^1_{\mathscr{R}}(p), V^2_{\mathscr{R}}(p), V^3_{\mathscr{R}}(p))^t$ and the vector $(n_0^{(\mathscr{R})}, n_1^{(\mathscr{R})}, n_2^{(\mathscr{R})} n_3^{(\mathscr{R})})^t$ and the same for the other integrand.

(3) we can add to the two integrals above a vanishing integral computed on the lateral boundary L of a tube $S \subset \mathbb{R}^4$ made of the integral lines of V and joining the two faces $\Sigma_0^{(\mathscr{R})} \cap B_0$ and $\Sigma_t^{(\mathscr{R})} \cap B_t$.

$$\int_L \mu(q) \langle V(q), n_L^{(\mathscr{R})} \rangle d\nu_{\mathscr{R}}(q)$$

$\nu_{\mathscr{R}}$ denoting the natural measure on the 3-surfaces in $\mathbb{R}^4$. This boundary integral simply vanishes because the unit vector $n_L^{(\mathscr{R})}$ (viewed as a vector in $\mathbb{R}^4$) normal to L is consequently normal to $(V^0_{\mathscr{R}}, V^1_{\mathscr{R}}, V^2_{\mathscr{R}}, V^3_{\mathscr{R}})^t$ and produces the result 0 when multiplied with $(V^0_{\mathscr{R}}, V^1_{\mathscr{R}}, V^2_{\mathscr{R}}, V^3_{\mathscr{R}})^t$ in the integrand;

(4) $n^{\mathscr{R}}$ at $\Sigma_t^{(\mathscr{R})} \cap B_t$ and $-n^{\mathscr{R}}$ at $\Sigma_0^{(\mathscr{R})} \cap B_0$ in components define outward vectors the boundary of S contained in the tube with basis B_0 and B_t.

Summing up, we have that, in coordinates (8.53) is equivalent to

$$\int_{+\partial S} \mu_0(q) \langle V(q), n^{(\mathscr{R})}(q) \rangle d\nu_{\mathscr{R}}(q) = 0 \,. \tag{8.54}$$

The divergence theorem in $\mathbb{R}^4$ implies that

$$\int_S \nabla_a [\mu_0(q) V(q)^a] dx^0 dx^1 dx^2 dx^3 = 0 \,.$$

Evidently, since the Lebesgue measure of $\mathbb{R}^4$ is translationally invariant, the found result is also valid if the temporal interval is $[t_1, t_1 + t]$ instead of $[0, t]$. Exploiting arbitrariness of S and the fact that the integrand is continuous, we have that the validity of (8.53) implies the **mass conservation law in local form**

$$\nabla_a(\mu_0 V^a) = 0 \quad \text{everywhere in } \mathbb{M}^4. \tag{8.55}$$

In fact, if this identity were false in an event, in a neighborhood of that event $\nabla_a \mu_0 V^a$ whould be strictly positive or strictly negative. Choosing S as a sufficiently small tube S – with bases B_0 and B_t – constructed around the said event, we would have that (8.53) is false simply going backward along our resoning. Instead we have supposed that the identity holds for every such

tube.
The local form of the mass conservation law implies back the integral form (8.53) of the same physical law again by going backward along our reasoning.

Remarks 8.42.
(1) A crucial observation is that *the identity (8.55) does not depend on the Minkowskian system of coordinates because these coordinates are Cartesian with respect to the affine structure of* $\mathbb{M}^4$ *and we can apply Definition 3.18.*
(2) It is worth stressing that (8.55) is generally false for a gas of interacting particles where particles are destroyed or created. ■

8.5.3 Four-momentum conservation in local form for the non-interacting gas

Let us pass to discuss the conservation of the four-momentum associated to our gas of non-interacting particles. If $\mathscr{R}$ is a given intertial reference frame, the natural definition of **density of four-momentum** in $\mathscr{R}$ is

$$\mathcal{P}_{\mathscr{R}} := \mu_{\mathscr{R}} V \, .$$

Notice that the definition depends on the reference frame just because we want eventually integrate $\mathcal{P}_{\mathscr{R}}$ over the rest space of that reference frame. Yet, for that reason the relevant density of mass is here $\mu_{\mathscr{R}}$ and not μ_0. Taking advantage of (8.51), we can rearrange this expression to

$$\mathcal{P}_{\mathscr{R}} = \frac{\mu_0(p)}{c} V \langle V(p), n^{(\mathscr{R})} \rangle \, . \tag{8.56}$$

This expression is more interesting since it can be intepreted in the following way which will prove to be quite universal. First, we introduce the so called **stress-energy tensor** field for the gas of non-interacting particles,

$$T := \mu_0 V \otimes V \qquad \text{namely} \quad T^{ab} := \mu_0 V^a V^b \, , \tag{8.57}$$

and form now on we use the following notation, if $\omega : \mathbb{M}^4 \ni p \mapsto \omega(p) \in T^*_p\mathbb{M}$ is a C^1 covariant vector field, then $T(\omega)$ is the contravariant vector field obtained by the contraction of ω and the left index of T (we omit to specify the event and the following identity is supposed to hold point by point in $\mathbb{M}^4$)

$$T(\omega)^j := T^{jk}\omega_k \, . \tag{8.58}$$

With this definition we immediately have that, referring to Minkowskian coordinates $x^0 = ct, x^1, x^2, x^3$ co-moving with the inertial reference system $\mathscr{R}$,

$$\mathcal{P}^a_{\mathscr{R}} = \langle T(dx^a), n_{\mathscr{R}} \rangle \, , \tag{8.59}$$

so that, we can define the **components of the total momentum** at time t with respect to Minkowsian coordinates $x^0 = ct, x^1, x^2, x^3$ co-moving with $\mathscr{R}$ as

$$P^a_{\mathscr{R}}(t) := \int_{\Sigma^{(\mathscr{R})}_t} \langle T(dx^a), n_{\mathscr{R}} \rangle dx^1 dx^2 dx^3 \, , \quad a = 0, 1, 2, 3 \, . \tag{8.60}$$

Remarks 8.43. We stress that up to now we do *not* know if these components, which are defined for every fixed Minkowskian coordinate system really define a contravariant vector when varying the coordinate systems. ■

To conclude we prove that the above quantities are in fact conserved, i.e., constant, in time. This result arises *only* from a local equation satisfied by T which, again does not depend on the reference frame. Later we also prove that the said local equation also implies that the components $P^a_{\mathscr{R}}$ actually define a vector P that does *not* depend on $\mathscr{R}$.

Theorem 8.44. *Consider a gas of non-interacting particles with C^1 proper density of mass μ_0 and C^1 field of four-valocities V. If the mass conservation law in local form (8.55) is valid*

$$\nabla_a(\mu_0 V^a) = 0$$

and the particles have inertial motion (see below) so that

$$V^b\nabla_b V^a = 0\,, \tag{8.61}$$

then the stress energy tensor (8.57) satisfies

$$\nabla_a T^{ab} = 0 \quad \textit{everywhere in } \mathbb{M}^4. \tag{8.62}$$

Proof. It is just matter of computations:

$$\nabla_a T^{ab} = \nabla_a\left(\mu_0 V^a V^b\right) = (\nabla_a \mu_0 V^a)V^b + \mu_0 V^a \nabla_a V^b = 0 + 0\,.$$

□

Remarks 8.45.
(1) Equation (8.61) is the so called **geodesical equation** which describes the motion of free falling particles in a general spacetime. The fact that (8.62) is true in view of that equation (and (8.55)) is quite relevant, since it allow us to generalize the statement of the theorem to General Relativity. In Minkowski spacetime, the meaning of (8.61) is the following. In Minkowskian coordinates, if $x^b = x^b(\tau)$ is the worldline integral curve of V[6], we have

$$0 = V^b(p)\nabla_b V^a(p) = \frac{dx^b}{d\tau}\frac{\partial V^a}{\partial x^b} = \frac{d}{d\tau}V^a(x(\tau)) = \frac{d}{d\tau}\frac{d}{d\tau}x^a(\tau) = \frac{d^2x^a}{d\tau^2}\,,$$

so that

$$x^a(\tau) = x^a(0) + V^a(x^0(0), x^1(0), x^2(0), x^3(0))\tau$$

and, as expected, the motion of the particle of continuous body are affine segments according with the inertia principle. *Vice versa*, (8.61) is satisfied if the worldlines of the particles are

[6]We can interpret the parameter τ of the integral lines as the proper time because of the constraint $V^aV_a = \eta_{ab}\frac{dx^a}{d\tau}\frac{dx^a}{d\tau} = -c^2$.

affine segments.
(2) As before, according to Definition 3.18, the equation of local conservatipn (8.62) is intrinsic and it does not depend on the used Minkowskian reference coordinate system. ■

We conclude with this pair of very nice consequences of the only equation (8.62) and definition (8.60) which are completly *independent* of the specific form (8.57) of the stress-energy tensor. From now on, a **tube** $U \subset \mathbb{M}^4$ is a connected open set with C^1 boundary whose intersection with every rest space $\Sigma^{(\mathscr{R})}_{x^0}$ defines an open connected set $U^{(\mathscr{R})}_{x^0}$ with C^1 boundary which is bounded in the $\mathbb{R}^3$ Euclidean topology of the coordinates. As a trivial example of U think of an infinite cylinder defined in Minkowskian coordinates of an inertial reference frame with constant basis $\{(x^1, x^2, x^3) \in \mathbb{R}^3 \mid (x^1)^2 + (x^2)^2 + (x^3)^2 < R^2\}$ for some real $R > 0$.

Theorem 8.46. *Let $\mathbb{M}^4 \ni p \mapsto T(p) \in T_p\mathbb{M}^4 \otimes T_p\mathbb{M}^4$ be a C^1 tensor field and assume that the components $P^a_{\mathscr{R}}(t)$ are defined as in (8.60) for every Minkowskian system of coordinates $x^0 = ct, x^1, x^2, x^3$ co-moving with every inertial reference frame $\mathscr{R}$. If T smootly vanishes before reacing the boundary ∂U of a tube U then,*

- **(a)** $P^a_{\mathscr{R}}(t) = P^a_{\mathscr{R}}(0)$ *($a = 0, 1, 2, 3$).*
- **(b)** *The four functions $P^a_{\mathscr{R}}$ define point-by-point a vector P which does not depend on the choice of inertial reference system $\mathscr{R}$ and the co-moving system of Minkowskian coordinates.*

Proof. (a) Referring to the $\mathbb{R}^4$ space constructed out of the Minkowskian coordinates $x^0 = ct, x^1, x^2, x^3$ of $\mathscr{R}$, define

$$S := U \cap \{(x^0 = c\tau, x^1, x^2, x^3) \in \mathbb{R}^4 \mid 0 \leq \tau < t\} .$$

Working within that $\mathbb{R}^4$ space, with the same argument we used to prove (8.54) but proceeding backward, the indentity $\nabla_a T^{ab} = 0$ implies in particular

$$0 = \int_U \nabla_a T^{ab} dx^0 dx^1 dx^2 dx^3 = \int_{+\partial S} n^{(S\cap U)}_a T^{ab} d\nu$$

$$= \int_{\Sigma^{(\mathscr{R})}_t \cap U^{(\mathscr{R})}_t} n^{(\mathscr{R})}_a T^{ab} dx^1 dx^2 dx^3 + \int_{\Sigma^{(\mathscr{R})}_0 \cap U^{(\mathscr{R})}_0} (-n^{(\mathscr{R})})_a T^{ab} dx^1 dx^2 dx^3 ,$$

where we have used the fact that T^{ab} smootly vanishes before reacing the lateral boundary of S, so we have omitted the surface integral arising from that part of the boundary of S. The found identity can be rarranged to

$$P^a_{\mathscr{R}}(t) = \int_{\Sigma^{(\mathscr{R})}_t} n^{(\mathscr{R})}_a T^{ab} dx^1 dx^2 dx^3 = \int_{\Sigma^{(\mathscr{R})}_0} n^{(\mathscr{R})}_a T^{ab} dx^1 dx^2 dx^3 = P^a_{\mathscr{R}}(0)$$

that is the thesis in (1).
(b) Consider another reference frame $\mathscr{R}'$ together with $\mathscr{R}$ and a rest space of the former $\Sigma^{\mathscr{R}'}_{t'}$

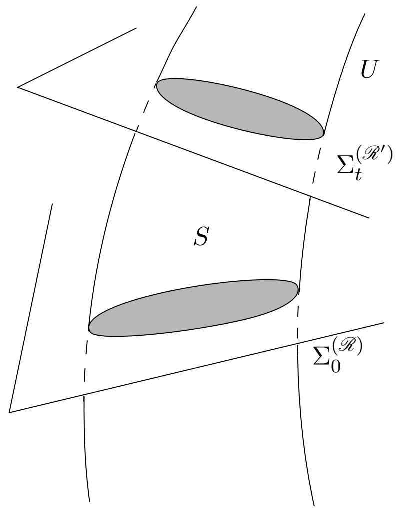

Figure 8.4: Proof of (2) Theorem 8.44

such that the portion S of U between the surfaces $\Sigma_0^{(\mathscr{R})}$ and $\Sigma_{t'}^{(\mathscr{R}')}$ is far form $\Sigma_0^{(\mathscr{R})} \cap \Sigma_{t'}^{(\mathscr{R})'}$. Notice that now S is a bounded solid with non-parallel bases (see the figure). Endow both reference frames with respective systems of co-moving Minkowskian coordinates and suppose that, for every constant covariant vector $\omega \in (T^4)^*$,

$$\int_{\Sigma_0^{(\mathscr{R})}} \langle T(\omega), n^{(\mathscr{R})} \rangle dx^1 dx^2 dx^3 = \int_{\Sigma_{t'}^{(\mathscr{R}')}} \langle T(\omega), n^{(\mathscr{R}')} \rangle dx'^1 dx'^2 dx'^3 \,. \tag{8.63}$$

Under this hypothesis, the linear maps

$$(T^4)^* \ni \omega \mapsto \int_{\Sigma_0^{(\mathscr{R})}} \langle T(\omega), n^{(\mathscr{R})} \rangle dx^1 dx^2 dx^3$$

and

$$(T^4)^* \ni \omega \mapsto \int_{\Sigma_{t'}^{(\mathscr{R}')}} \langle T(\omega), n^{(\mathscr{R}')} \rangle dx'^1 dx'^2 dx'^3$$

coincide, thus they define the same covariant vector field P such that $\langle P, \omega \rangle$ coincides with the common value of the right-hand sides above. By construction, the components of P in every $\mathscr{R}$ and co-moving Minkowskian coordinates are just the constant right-hand side of (8.60) at every time t for the part (a). To conclude the proof it is therefore sufficient to establish (8.63). To this end, we work in the $\mathbb{R}^4$ of the coordinates x^0, x^1, x^2, x^3. They are related to the coordinates x'^0, x'^1, x'^2, x'^3 through a Poincaré transformation

$$x'^a = \Lambda^a{}_b x^b + C^a \,.$$

The three-dimensional surface $\Sigma_{t'}^{(\mathscr{R}')}$ is therefore described as the plane $x'^0 = ct'$ in $\mathbb{R}^4$:

$$\Sigma_{t'}^{(\mathscr{R}')} \equiv \{(x^0, x^1, x^2, x^3) \in \mathbb{R}^4 \mid \Lambda^0{}_b x^b - ct' - C^0 = 0\} \,.$$

We shall suppose that the intersection of $\Sigma_0^{(\mathscr{R})}$ and $\Sigma_{t'}^{(\mathscr{R}')}$ does not pass through the solid S. That is possible, since S is bounded, just by taking $t' > 0$ sufficiently large. The generic case is encompassed when taking statement (a) into account (replacing $\mathscr{R}$ for $\mathscr{R}'$).
Since ω is constant and thus $\nabla_a(T(\omega)^a) = \nabla_a(T^{ab}\omega_b) = (\nabla_a T^{ab})\omega_b = 0$, the $\mathbb{R}^4$ divergence theorem implies, with the same argument as above

$$\int_{\Sigma_0^{(\mathscr{R})}} \langle T(\omega), n^{(\mathscr{R})}\rangle dx^1 dx^2 dx^3 = \int_{\Sigma_{t'}^{(\mathscr{R}')}} T(\omega)^a N_a d\nu \,, \tag{8.64}$$

where ν is the natural measure on $\Sigma_{t'}^{(\mathscr{R}')}$ induced by the one of $\mathbb{R}^4$ and (N_0, N_1, N_2, N_3) is the normal vector to $\Sigma_{t'}^{(\mathscr{R}')}$ in $\mathbb{R}^4$ with unit *Euclidean norm* of $\mathbb{R}^4$. Form the definition of $\Sigma_{t'}^{(\mathscr{R}')}$ it holds

$$N^a = \frac{\Lambda^0{}_a}{\sqrt{\sum_{b=0}^3 (\Lambda^0{}_b)^2}} \,, \quad a = 0,1,2,3, \tag{8.65}$$

As a consequence, if we parametrize $\Sigma_{t'}^{(\mathscr{R}')}$ with the coordinates x^1, x^2, x^3 trough the canonical projection $\pi : \mathbb{R}^4 \ni (x^0, x^1, x^2, x^3) \mapsto (x^1, x^2, x^3)$, the volume $vol(G)$ of a subset $G \subset \Sigma_{t'}^{(\mathscr{R}')}$ with respect to the $\mathbb{R}^4$ measure induced on $\Sigma_{t'}^{(\mathscr{R}')}$ and that of its projection $\pi(G)$ are related by the relation, where the dot denotes the standard scalar product in $\mathbb{R}^4$,

$$\frac{Vol(\pi(G))}{Vol(G)} = N \cdot n_{\mathscr{R}} = \sum_{a=0}^{3} \frac{\Lambda^0{}_a}{\sqrt{\sum_{b=0}^3 (\Lambda^0{}_b)^2}} \delta^0_a = \frac{\Lambda^0{}_0}{\sqrt{\sum_{b=0}^3 (\Lambda^0{}_b)^2}} \,.$$

In other words, the induced measure on $\Sigma_{t'}^{(\mathscr{R}')}$ can be written

$$d\nu = \frac{\sqrt{\sum_{b=0}^3 (\Lambda^0{}_b)^2}}{\Lambda^0{}_0} dx^1 dx^2 dx^3 \,. \tag{8.66}$$

Collecting (8.65) and (8.66), the right hand side of (8.64) has consequently the form

$$\int_{\Sigma_{t'}^{(\mathscr{R}')}} T(\omega)^a N_a d\nu = \int_{\Sigma_{t'}^{(\mathscr{R}')}} \frac{\Lambda^0{}_a T(\omega)^a}{\sqrt{\sum_{b=0}^3 (\Lambda^0{}_b)^2}} \frac{\sqrt{\sum_{b=0}^3 (\Lambda^0{}_b)^2}}{\Lambda^0{}_0} dx^1 dx^2 dx^3 = \int_{\Sigma_{t'}^{(\mathscr{R}')}} \frac{T'(\omega)^0}{\Lambda^0_0} dx^1 dx^2 dx^3 \,.$$

Now suppose that the two Minkowskian coordinate systems are connected by a pure transformation $x'^a = (\Lambda_{\dot\gamma})^a{}_b x^b$. Here $\dot\gamma$ is the four velocitiy of every point at rest with $\mathscr{R}$. Its components in $\mathscr{R}'$ define the pure transformation $\Lambda_{\dot\gamma}$. As before, we parametrize the spatial volumes in $\Sigma_{t'}^{(\mathscr{R}')}$ in terms of the coordinates x^1, x^2, x^3 co-moving with $\mathscr{R}$ taking advantage of the canonical projection $\pi : \Sigma_{t'}^{(\mathscr{R}')} \to \Sigma_0^{(\mathscr{R})}$. The mathematical interplay beteween the two spatial volumes is the same as for the contraction of volumes: there is a spatial volume at rest with $\mathscr{R}$ evolving with the integral lines of $\dot\gamma$ and we want to evaluate the corresponding volume in $\mathscr{R}'$ by crossing the worldlines emanated from the former volume with the rest spaces $\Sigma_{t'}^{(\mathscr{R}')}$ of $\mathscr{R}'$[7]. From the contraction of volumes formula (8.34) we know that

$$\det\left[\frac{\partial x'^{\alpha}}{\partial x^{\beta}}\right]_{\alpha,\beta=1,2,3} = \sqrt{1-\frac{v^2}{c^2}} = \frac{1}{\dot{\gamma}^0/c} = \frac{1}{(\Lambda_{\dot{\gamma}})^0{}_0} \, .$$

If the Poincaré transformation connecting $\mathscr{R}$ and $\mathscr{R}'$

$$x'^a = \Lambda^a{}_b x^b + C^a$$

is not pure, leaving fixed the reference frames but changing the co-moving coordinates, we can always reduce to a pure *Lorentz* transformation by a translation of the origin of four axes in $\mathscr{R}'$ and spatial rotation (8.30) in the same space. Let us discuss this point with some details. First, we stress that these changes of coordinates are isometries, so that they do not change the measure on $\Sigma_{t'}^{(\mathscr{R}')}$ which is the Lebesgue one and thus invariant under isometries. Similarly, a translation along t' does not affect the mesure on the rest spaces of $\mathscr{R}'$. The relevant measure on $\Sigma_{t'}^{(\mathscr{R}')}$ is always written as $dx'^1dx'^2dx'^3$ independently of the choice of the Minkowskian coordinates co-moving with $\mathscr{R}'$. With a four tanslation in $\mathscr{R}'$, without changing the measure on $\Sigma_{t'}^{(\mathscr{R}')}$, we pass from $x'^a = \Lambda^a{}_b x^b + C^a$ to $x'^a = \Lambda^a{}_b x^b$. Next, due to the decomposition theorem of the Lorentz group $\Omega_{R^{-1}}\Lambda$ is pure if $R \in SO(3)$ is suitably chosen. As said above, the measure on $\Sigma_t'^{(\mathscr{R}')}$ is not affected by this change of coodinates. Furthermore, the product with a spatial rotation does not affect the term $\Lambda^0{}_0$ as one immediately proves per direct inspection:

$$(\Omega_{R^{-1}}\Lambda)^0{}_0 = \Lambda^0{}_0 \quad .$$

We conclude that

$$\det\left[\frac{\partial x'^{\alpha}}{\partial x^{\beta}}\right]_{\alpha,\beta=1,2,3} = \frac{1}{\Lambda^0{}_0}$$

is valid in general when $x'^a = \Lambda^a{}_b x^b + C^a$ holds (with $\Lambda \in O(1,3)\uparrow$) and this result gives the transformation laws between the natural measures on the considered rest spaces:

$$dx'^1dx'^2dx'^3 = \frac{1}{\Lambda^0_0}dx^1dx^2dx^3 \, .$$

So that

$$\int_{\Sigma_{t'}^{(\mathscr{R}')}} T(\omega)^a N_a d\nu = \int_{\Sigma_{t'}^{(\mathscr{R}')}} \frac{T'(\omega)^0}{\Lambda^0_0} dx^1dx^2dx^3 = \int_{\Sigma_{t'}^{(\mathscr{R}')}} T'(\omega)^0 dx'^1dx'^2dx'^3$$

$$= \int_{\Sigma_{t'}^{(\mathscr{R}')}} \langle T'(\omega), n^{(\mathscr{R}')}\rangle dx'^1dx'^2dx'^3 \, .$$

[7] We observe that the interplay of the two spatial volumes has nothing to to with the particles of the continuum we are studying, whose worldlines are described by V, and the reasoning above is of pure (lorentzian) geometric nature.

Inserting this result in (8.64) we have (8.63) concluding the proof. $\square$

Remarks 8.47.
(1) It should be evident from the proof that the hypothesis concerning the existence of the tube U can be relaxed by assuming a sufficient rapid decay of T at spatial infinity of an inertial reference frame. We leave to the reader this sort of generalizations.
(2) Introducing the *divergence theorem in covariant form* with respect to the *Levi-Civita connection*, the proof of (8.63) would be trivial.
(3) What is really fundamental in the proof of Theorem 8.46 is the divergence theorem in $\mathbb{R}^4$ which is exploited just in view of is the validity of the equation $\nabla_a T(\omega)^a = 0$, where $T(\omega)^a = \omega_b T^{ab}$ and $\omega = dx^k$ (referred to some Minkowskian coordinate system x^0, x^1, x^2, x^3). Actually, $\nabla_a T(\omega)^a = 0$ is more generally valid if T satisfies (8.62) and ω is a constant covariant field or, more weakly:

(i) T is *symmetric* and

(ii) ω satisfies the identity, known as **Kiling condition**,

$$\nabla_a \omega_b + \nabla_b \omega_b = 0 \,. \tag{8.67}$$

Indeed, under this couple of assumptions,

$$\nabla_a \left(\omega_b T^{ba}\right) = (\nabla_a \omega_b) T^{ba} + \omega_b \nabla_a T^{ba} = \frac{1}{2}(\nabla_a \omega_b)(T^{ba} + T^{ab}) + \omega_b \nabla_a T^{ab}$$
$$= \frac{1}{2}(\nabla_a \omega_b + \nabla_b \omega_a) T^{ba} + \omega_b \nabla_a T^{ab} = 0 \,.$$

When $\nabla_a T(\omega)^a = 0$, the quantities

$$Q_{\mathscr{R}}[\omega] = \int_{\Sigma_t^{(\mathscr{R})}} \langle T(\omega), n^{(\mathscr{R})} \rangle dx^1 dx^2 dx^3$$

are conserved in time (they do not depend on t) and $Q_{\mathscr{R}}[\omega] = Q_{\mathscr{R}'}[\omega]$ for every choice of the reference frames $\mathscr{R}, \mathscr{R}'$. The proof is the same as that of Theorem 8.46.
Condition (8.67) can be stated more generally on manifolds and it determines the continuous isometries of a manifod obtained by constructing the flow of the contravariant version of its solutions ω. In that context, conserved quantities $Q[\omega]$ similarly arise. In this sense, the know relation between symmetries and conserved quantities can be recast in General Relativity.
(4) Within the hypotheses of (3), if $\mathbb{M}^4 \ni e \mapsto T_e(\omega) \in V$ is a C^1 timelike future directed vector field, we can consider a tube between $\Sigma_t^{(\mathscr{R})}$ and $\Sigma_{t'}^{(\mathscr{R}')}$ with a basis $B \subset \Sigma_t^{(\mathscr{R})}$ and the other

$$B' \subset \Sigma_{t'}^{(\mathscr{R}')} \,,$$

where $B' = \{\phi_s(e) \mid e \in B, s \in I\} \cap \Sigma_{t'}^{(\mathscr{R}')}$ and ϕ is the flow of the contravariant vector field $T(\omega)$ where $I \subset \mathbb{R}$ is a sufficiently large interval. If everything goes as expected when choosing

B sufficiently regular and small, so that $\phi_s(B)$ is well defined on some interval $I \ni s$ and B' sufficiently close to B is well defined, a local version of the conservation of Q can be established:

$$\int_{B'} \langle T(\omega), n^{(\mathscr{R}')} \rangle dx'^1 dx'^2 dx'^3 = \int_{B'} \langle T(\omega), n^{(\mathscr{R})} \rangle dx^1 dx^2 dx^3$$

with the same argument used in proving the local conservation of the mass, starting from $\nabla_a T^{ab} = 0$ and also taking advantage of some constructions and arguments exploited in the proof of (b) in Therem 8.46. As a matter of fact, the validity of the integral identity above for a suitable large family of sets B is equivalent to the law $\nabla_a T^{ab} = 0$. The issue with the generality of the above integral identity is that $T(\omega)$ is not timelike for all physical systems. Quite often that vector field is causal and it vanishes somewhere or it is spacelike. In general, the validity of the integral equation above should be checked case-by-case.

(5) Sometime the stress-energy tensor of a *single particle* is defined in terms of a *Dirac measure* (or *Dirac distibution*). Here is the sketch of this formalism. Fix a reference frame $\mathscr{R}$ and co-moving Minkowskian coordinates $ct = x^0, x^1, x^2, x^3$. Let us suppose that the world line γ is described, in coordinates, by $x^a = x^a(\tau)$. As $\frac{dct}{d\tau} > 0$ because the worldline is timelike and future directed, we can re-parametrize the worldline using the coordinate t as new parameter. If the particle has mass $m > 0$, the mass density on $\Sigma_t^{\mathscr{R}}$ can be defined as the Dirac measure $\delta(\vec{x} - \vec{x}(t))$, where $\vec{x} = (x^1, x^2, x^3)^t$ is a generic point on $\Sigma_t^{(\mathscr{R})} \equiv \mathbb{R}^3$. Notice that $\vec{x}(t)$ is assigned at each time t. The measure is therefore concentrated on $\vec{x}(t)$. The stress energy-tensor can be defined, generalizing (8.57), as

$$T^{ab}(t, \vec{x}) = V^a(t) V^b(t) m J(t, \vec{x}) \delta(\vec{x} - \vec{x}(t))$$

where the four velocity V is constructed out of the universe line γ and the Jacobian $J(t, \vec{x})$ is necessary to interpret the Dirac-delta density as a measure on the *instantaneous* rest space of the particle as in (8.49):

$$J(t, \vec{x}) = \sqrt{1 - \frac{1}{c^2} \left(\frac{d\vec{x}}{dt} \right)^2} \,.$$

In summary,

$$T(t, \vec{x}) = m \sqrt{1 - \frac{1}{c^2} \left(\frac{d\vec{x}}{dt} \right)^2} \delta(\vec{x} - \vec{x}(t)) \, V(t) \otimes V(t) \,.$$

The object above, though obtained quite heuristically, possesses good mathematical properties in the appropriate theoretical context. For instance, viewed as a tensor valued distribution on $\mathbb{M}^4$, it takes the same form independently of the choice of the reference frame (and co-moving coordinates). Furthermore, if viewed as a compact support tensor-valued distribution on $\Sigma_t^{(\mathscr{R})}$, the four momentum of the particle is obtained (exercise!) as the integral

$$m V^a(t) = \int_{\Sigma_t^{(\mathscr{R})}} T^{ab}(t, \vec{x}) n_b^{(\mathscr{R})} dx^1 dx^2 dx^3 \,.$$

$\blacksquare$

Exercises 8.48.

1. Prove that, assuming $T^{ab} = \mu_0 V^a V^b$, with V everywhere timelike future-directed and $\mu_0 \geq 0$, then the conservation equation $\nabla_a T^{ab} = 0$ implies both the inertial motion $V^a \nabla_a V^b = 0$ (where $\mu_0 \neq 0$) and the conservation of the mass $\nabla_a(\mu_0 V^a) = 0$.

Solution. $V_b \nabla_a T^{ab} = 0$ can be expanded to $0 = V^b(\nabla_a V^b)\mu_0 V^a + V^b V_b \nabla_a(\mu_0 V^a)$, where $V^b(\nabla_a V^b) = \nabla_a(V^b V_a)/2 = -\nabla_a c^2 = 0$ and thus, it remains $-c^2\nabla_a(\mu_0 V^a) = 0$. Taking this identity into account, by expanding $\nabla_a(\mu_0 V^a V^b) = 0$, one finds $\mu_0 V^a \nabla_a V^b = 0$. Hence $V^a \nabla_a V^b = 0$ where $\mu_0 \neq 0$.

2. Prove that, assuming $T^{ab} = \mu_0 V^a V^b$, with V everywhere timelike future-directed and $\mu_0 \geq 0$, then $\nabla_a T^{ab} = f^b$, where $V_b f^b = 0$, can be intepreted as the motion equation for the particles of the non-interacting gas under the action the density of *mechanical* four-force f, and the density is measured at rest with the particles:

$$\mu_0 \frac{d}{d\tau}\left(\frac{dx^b}{d\tau}\right) = f^b \,, \tag{8.68}$$

where $x^b = x^b(\tau)$ is a worldline of a particle of the continuous body in Minkowskian coordinates. Prove also that the mass conservation law $\nabla_a(\mu_0 V^a) = 0$ is still valid and that μ_0 is constant along the said worldlines.

Solution. $V_b \nabla_a T^{ab} = V_b f^b = 0$ can be expanded to $0 = V_b(\nabla_a V^b)\mu_0 V^a + V^b V_b \nabla_a(\mu_0 V^a)$, where $V_b(\nabla_a V^b) = \nabla_a(V^b V_b)/2 = -\nabla_a c^2 = 0$ and thus, it remains $-c^2\nabla_a(\mu_0 V^a) = 0$, so that the mass is conserved. Using this result in $\nabla_a T^{ab} = f^b$, we obtain, $\mu_0 V^a \nabla_a V^b = f^b$. To conclude, using Minkowskian coordinates as in (1) Comments 8.45 to describe those worldlines, the equation $\mu_0 V^a \nabla_a V^b = f^b$ obtained above can be rephrased to $\mu_0 \frac{d^2 x^b}{d\tau^2} = f^b$.

3. Prove that Eq. (8.68) has the heuristic meaning of the equation of motion of an "infinitesimal" portion of continuum in agreement with Eq. (8.44).

Solution. Eq. (8.68) has the heuristic meaning of the equation of motion of a portion of continuous body of mass δm referred to a co-moving volume δv_0 such that $\mu_0 = \frac{\delta m}{\delta v_0}$. Inserting this ration in (8.68) we have

$$\delta\mu \frac{d}{d\tau}\left(\frac{dx^b}{d\tau}\right) = f^b \delta v_0 \,,$$

that is, since the mass of the volume does not change along its worldline,

$$\frac{d}{d\tau}\delta P^b = f^b \delta v_0 \,, \quad \text{where} \quad \delta P^b = \delta\mu \frac{dx^b}{d\tau} \,,$$

this equation is nothing but (8.44) for the considered portion of continuous body. We can pass from the proper time to the global time t of the used Minkowski frame $\mathscr{R}$

$$\frac{d}{dt}\delta P^b = f^b \frac{\delta v_0}{V^0} .$$

Since the volume of the portion of continuum we are considering is $\delta v = \frac{\delta v_0}{V^0}$ when referred to the rest space of our reference frame $\mathscr{R}$, the previous equation becomes

$$\frac{d}{dt}\delta P^b = f^b \delta v .$$

8.5.4 The macroscopic stress energy tensor in the general case

The pair of theorems proved above suggest the idea that the energy and momentum content of every isolated continuous system should be described by means of a $(2,0)$-type tensor field T satisfying the local equation (8.62)

$$\nabla_a T^{ab} = 0 \quad \text{everywhere in } \mathbb{M}^4 ,$$

and such that the total momentum of the system P should be defined by (8.60)

$$P^b_{\mathscr{R}}(t) := \int_{\Sigma_t^{(\mathscr{R})}} \langle T(dx^b), n^{(\mathscr{R})}\rangle dx^1 dx^2 dx^3 , \quad b = 0, 1, 2, 3 ,$$

where we have fixed a reference system $\mathscr{R}$ and co-moving Minkowskian coordinates. As established before, these choices do not affect the definition of the vector P, which does not depend on $\mathscr{R}$, on the used co-moving coordinates, and it is also constant in time, due to the validity of (8.62) as proved in Theorem 8.46. These properties of P are true provided the stress energy tensor vanishes outside compact sets in the rest spaces of reference frames (or it vanishes sufficiently rapidly at infinity on those 3-spaces). Conversely $\nabla_a T^{ab} = 0$ is a local equation and can be considered valid even if P cannot be defined.

Remarks 8.49. With this intepretation T^{00} and $\frac{1}{c}T^{0\alpha}$ respectively have the meaning of spatial energy density and the spatial density of the component $\alpha = 1, 2, 3$ of the momentum in a reference frame $\mathscr{R}$ with co-moving coordinates x^0, x^1, x^2, x^3. ■

The most direct way to construct the stress-energy tensor of a physical system in Special Relativity passes through the variational approaches and the various formulations of the Noether theorem when this formalism is suitable for the considered system. Tipically this happens for a field to be quantized and often describing elementary particles. This approach concerns (quantum) *microscopic* systems. However, the existence of the stress-energy tensor for *macroscopic* systems can be assumed as a direct generalization of the *stress tensor* (see Examples 3.19) of classical mechanics together with the general classical laws of the continuum mechanics as we are going to illustrate.

The fundamental equations of *classical* continuum mechanics in classical mechanics are the following three ones, valid in a *classical* intertial reference frame and referring to orthonormal Cartesian coordinates x^1, x^2, x^3 in the rest space of the reference system and to the absolute time t,

$$\frac{\partial \mu}{\partial t} + \sum_{\beta=1}^{3} \frac{\partial}{\partial x^\beta} \mu v^\beta = 0 \,, \tag{8.69}$$

$$\mu \left(\frac{\partial v^\alpha}{\partial t} + \sum_{\beta=1}^{3} v^\beta \frac{\partial}{\partial x^\beta} v^\alpha \right) = \sum_{\beta=1}^{3} \frac{\partial}{\partial x^\beta} \sigma^{\alpha\beta} + f^\alpha \,, \quad \alpha = 1, 2, 3 \,. \tag{8.70}$$

Above, v^α is a component of the velocity field vector of the continuum, μ is the mass density, and σ is Cauchy's stress tensor describing the *internal* stresses of the system. Furthermore f^α is α-th components of the spatial density of *external* forces. Eq. (8.69) is the (classical) *conservation law of the mass*. Eq. (8.70) is *Newton's second law* for every particle of the continuous body taking into account both internal stresses and external forces. The (classical) conservation law of the angular momentum is equivalent to the fact that σ is symmetric, when the previous two equations are taken into account:

$$\sigma^{\alpha\beta} = \sigma^{\beta\alpha} \,. \tag{8.71}$$

When one specifies the type of continuous body, further information must be supplied regarding the relation between the stress tensor, the field of velocities and other geometrical structures of the continuous body as the *deformation tensor*. However, the above three general equations are completely sufficient for an elementary heuristic formulation.

In order to compare (8.69)-(8.70) with the corresponding relativistic equations, we are going re-write the second equation into another form. The final system is however equivalent to (8.69)-(8.70).

Taking advantage of the trivial identity

$$\mu \frac{\partial v^\alpha}{\partial t} = \frac{\partial}{\partial t} \mu v^\alpha - v^\alpha \frac{\partial \mu}{\partial t} \,,$$

expanding the above derivative $\frac{\partial \mu}{\partial t}$ in terms of the spatial derivativs using (8.69), we have

$$\mu \frac{\partial v^\alpha}{\partial t} = \frac{\partial}{\partial t} \mu v^\alpha + v^\alpha \sum_{\beta=1}^{3} \frac{\partial}{\partial x^\beta} \mu v^\beta \,.$$

Inserting this result in the left-hand side of (8.70), we find

$$\frac{\partial}{\partial t} \mu v^\alpha + v^\alpha \sum_{\beta=1}^{3} \frac{\partial}{\partial x^\beta} \mu v^\beta + \sum_{\beta=1}^{3} \mu v^\beta \frac{\partial}{\partial x^\beta} v^\alpha = \sum_{\beta=1}^{3} \frac{\partial}{\partial x^\beta} \sigma^{\alpha\beta} + f^\alpha \,.$$

That is

$$\frac{\partial}{\partial t}\mu v^\beta + \sum_{\alpha=1}^{3} \frac{\partial}{\partial x^\alpha}\left(\mu v^\alpha v^\beta - \sigma^{\alpha\beta}\right) = f^\beta \,.$$

In summary, since the passages above are reversible (when keeping both equations), the set of equations (8.69)-(8.70) can be equivalently re-arranged as follows

$$\frac{\partial \mu}{\partial t} + \sum_{\beta=1}^{3} \frac{\partial}{\partial x^\alpha}\mu v^\alpha = 0 \,, \tag{8.72}$$

$$\frac{\partial}{\partial t}\mu v^\beta + \sum_{\alpha=1}^{3} \frac{\partial}{\partial x^\alpha}\left(\mu v^\alpha v^\beta - \sigma^{\alpha\beta}\right) = f^\beta \,, \quad \beta = 1,2,3 \,. \tag{8.73}$$

Let us finally assume that the continuous body is nothing but the *classical limit* of a relativistic system admitting a stress-energy tensor T satisfying a natural generalization of (8.62),

$$\nabla_a T^{ab} = f^b \,, \tag{8.74}$$

where the vector field f on $\mathbb{M}^4$ is intepreted as the *density of external four force* and T^{00} and $\frac{1}{c}T^{0\beta}$ respectively represent the energy density and the momentum densities. From this perspective, (8.74) can be interpreted as a direct relativistic generalization of the set of equations (8.69)-(8.70). To show it, fix a Minkowskian reference frame $\mathscr{R}$ equipped which co-moving Minkowskian coordinates $x^0 = ct$, x^1, x^2, x^3 identified with the classical ones and expand (8.74) in components:

$$\frac{\partial}{\partial t}\frac{1}{c^2}T^{00} + \sum_{\beta=1}^{3} \frac{\partial}{\partial x^\alpha}\frac{1}{c}T^{\alpha 0} = \frac{1}{c}f^0 \,, \tag{8.75}$$

$$\frac{\partial}{\partial t}\frac{1}{c}T^{0\beta} + \sum_{\alpha=1}^{3} \frac{\partial}{\partial x^\alpha}T^{\alpha\beta} = f^\beta \,, \quad \beta = 1,2,3 \,. \tag{8.76}$$

When we say that the relativistic system admits a classical limit we just mean that, within the regime of small velocity v^a (with respect to c), T^{00}/c^2 and $T^{\beta 0}/c$ become the corresponding classical quantities

$$T^{00}/c^2 \simeq \mu \,, \quad T^{\beta 0}/c \simeq \mu v^\beta$$

and $\frac{1}{c}f^0$ is negligible (see Comment 8.40) that is equivalent to postulate the validity of the (classical!) mass conservation law. In that case, equation (8.72) becomes (8.75). Similarly, (8.73) tends to become (8.70) provided one assumes, in that regime,

$$T^{\alpha\beta} \simeq \mu v^\alpha v^\beta - \sigma^{\alpha\beta} \,, \quad T^{0\beta}/c \simeq T^{\beta 0}/c \simeq \mu v^\beta \,. \tag{8.77}$$

According to this intepretation, let us study the special choice of a reference frame exactly *at rest* with a particle of the continuum at a given event $e \in \mathbb{M}^4$. Here (8.77) boils down to some *exact* relations valid exactly at e (we will omit to write it)

$$T^{\alpha\beta} = -\sigma^{\alpha\beta} \,, \quad T^{0\beta} = T^{\beta 0} = 0 \,, \quad \alpha, \beta = 1,2,3 \tag{8.78}$$

Since $\sigma^{\alpha\beta} = \sigma^{\beta\alpha}$, in particular $T^{ab} = T^{ba}$ is valid in the said reference frame, i.e. the considered basis of $T_e\mathbb{M}^4 \otimes T_e\mathbb{M}^4$. However, as we known, the symmetry property of a tensor does not depend on the chosen basis, so that we are commited to assume that *the stress-energy tensor is a symmetric tensor*, at least when dealing with macroscopic continuous media. Symmetry also permits to present the law of conservation of the relativistic angular momentum into an easy formulation (see Exercises (8.53) below).

Remarks 8.50. In the very general case, the symmetry of the stress energy tensor is a delicate issue especially when dealing with variational approaches for systems which do not admit a classical limit in terms of continuous bodies. The stress-energy tensor obtaioned in that way is not symmetric in general and it must be made symmetric, preserving the total conserved quantities, obtaining the so-called *Belinfante-Rosenberg stress-energy tensor*. ■

Examples 8.51. Referring to the gas of non-interacting particles we have $\sigma^{\alpha\beta} = 0$ and, indeed, $T^{\alpha\beta} = \mu_0 V^\alpha V^\beta = 0$ if $\alpha, \beta = 1, 2, 3$ in the rest frame with a particle of continuum, since V has only the temporal component. ■

This discussion suggests that the notion of stress energy-tensor is a very promising tool to describe the content of momentum, energy, but also internal stresses for a relativistic extended continuous system. We expect that T^{ab} can be also defined for systems which are not continuous systems in classical sense: (quantum) fields, in particular. In that case the classical limit (8.72)-(8.73) does not make sense. We can give the following general interpretation of the terms which decompose T^{ab} in a given reference frame $\mathscr{R}$.

(i) $T^{00}_{\mathscr{R}}$ is the *density of energy* of the system referred to rest space $\Sigma^{(\mathscr{R})}_t$,

(ii) $cT^{\alpha 0}_{\mathscr{R}}$ is the α-component of the *internal flux of energy* of the system referred to rest space $\Sigma^{(\mathscr{R})}_t$,

(iii) $\frac{1}{c}T^{0\beta}_{\mathscr{R}}$ is the *density of the spatial momentum* of the system along the β axis referred to rest space $\Sigma^{(\mathscr{R})}_t$,

(iv) $T^{\alpha\beta}_{\mathscr{R}}$ is the α-component of the *internal flux of spatial momentum* along the β axis of the system referred to rest space $\Sigma^{(\mathscr{R})}_t$ (it).

Remarks 8.52. It is important to notice that the internal flux of spatial momentum is not a purely kinematic object as it includes the internal stresses in case of a macroscopic continuous body. Indeed, in the classical limit it becomes $T^{\alpha\beta} \simeq \mu v^\alpha v^\beta - \sigma^{\alpha\beta}$. ■

The intepretation above is a consequence of the following integral representation of $\nabla_a T^{ab} = 0$. Fix a spatial bounded geometrical volume B – an open set whose closure is compact with orientable regular boundary – at rest with an inertial reference frame $\mathscr{R}$. Working in Minkowskian coordinates $x^0 = ct, x^1, x^2, x^3$, let us integrate the equation above in B. We have

$$\int_B \nabla_0 T^{0a} dx^1 dx^2 dx^3 = -\sum_{\alpha=1}^{3} \int_B \nabla_\alpha T^{\alpha a} dx^1 dx^2 dx^3 \, .$$

Since $\overline{B}$ is compact, if assuming the integrands above are continuous in all coordinates, we can extract the temporal derivative (for instance through a suitable use of Lebesgue's dominated convergence theorem and Lagrange's theorem) and use the divergence theorem on the right-hand side obtaining

$$\frac{d}{dt}\int_B \frac{1}{c} T^{0a} dx^1 dx^2 dx^3 = -\oint_{+\partial B} \sum_{\alpha=1}^{3} T^{\alpha a} n_\alpha d\nu \, ,$$

where n_α denotes the α-th component of the outward normal vector to ∂B. It is now evident that the variation per unit of time of the total quantity in the volume B (energy or spatial momentum up to constant factors) amounts to the flow of the corresponding quantity passing through the boundary ∂B of the voulume. All that corroborates the interpretation proposed above for the integrands in the right-hand side.

8.5.5 Interacting macroscopic continuous systems

When we add a density of external four force density, the physical interpretation of T^{ab} and its conservation law still hold, but the components of the four force act as *external sources* of energy and momentum. To explain this extension of the formalism consider the case of two continuous systems, where each subsystem has its own stress energy tensor T, T'. Here, the total stress-energy tensor T_{tot} is the sum of those terms with possibly a third part describing the interaction between them,

$$T_{tot} = T + T' + T_{int} \, .$$

If the overall system is isolated, then T_{tot} satisfies (8.62), but T, T', T_{int} do not separately. Equivalently the *density of external force* acting on the first and the second system are respectively

$$f^b = -\nabla_a (T'^{ab} + T^{ab}_{int}) \quad \text{and} \quad f'^b = -\nabla_a (T^{ab} + T^{ab}_{int}) \, .$$

Notice that $f^b \neq -f'^b$ unless $\nabla_a T^{ab}_{int} = 0$.

Exercises 8.53.

1. Consider a Minkowskian coordinate system co-moving with an inertial reference frame $\mathscr{R}$. Assume that both $\nabla_a T^{ab} = 0$ and $T^{ab} = T^{ba}$ and define

$$M^{ab}_{\mathscr{R}}(t) := \frac{1}{c}\int_{\Sigma^{(\mathscr{R})}_t} \left(x^a T^{0b} - x^b T^{0a} \right) \, dx^1 dx^2 dx^3 \, .$$

Above, the integrand, and more strongly the functions $\mathbb{R}^4 \ni x \mapsto T^{ab}(x) \in \mathbb{R}$, are C^1 and has spatial compact support in a neighborhood of the considered instant t. Prove that

$$M^{ab}_{\mathscr{R}}(t) = M^{ab}_{\mathscr{R}}(0) \, .$$

2. Provide a physical interpretation of the quantities

$$L_\gamma^{(\mathscr{R})} := \frac{1}{2}\sum_{\alpha,\beta=1}^{3} \epsilon_{\gamma\alpha\beta} M_{\mathscr{R}}^{\alpha\beta} \quad \text{where } \gamma = 1,2,3.$$

(*Hint.* $L_\gamma^{(\mathscr{R})}$ is the γ-th component of the total angular momentum in $\mathscr{R}$ computed with respect to the origin of the coordinates.)

3. Prove that if $\mathscr{R}$ and $\mathscr{R}'$ are connected by the Poincaré transformation $x'^a = \Lambda^a{}_b x^b + C^a$ and $M_{\mathscr{R}}^{ab}$ and $M_{\mathscr{R}'}^{ab}$ are defined as above, then

$$M_{\mathscr{R}'}^{ab} = \Lambda^a{}_c \Lambda^b{}_d M_{\mathscr{R}}^{cd} + C^a P_{\mathscr{R}'}^b - C^b P_{\mathscr{R}'}^a \,.$$

4. Specialising $M_{\mathscr{R}}^{ab}(t) = M_{\mathscr{R}}^{ab}(0)$ to the case $a = 0, b = \beta$ prove that

$$\frac{1}{c}\int_{\Sigma_t^{(\mathscr{R})}} T^{0\beta}(t) dx^1 dx^2 dx^3 = \frac{d}{dt}\int_{\Sigma_t^{(\mathscr{R})}} x^\beta \frac{T^{00}}{c^2} dx^1 dx^2 dx^3$$

and provide a physical intepretation of this equation.
(*Hint.* The argument of the temporal derivative in right hand side is Mx_G^β where x_G^β is the β-th coordinate of the *center of mass* of the physical system in the reference frame $\mathscr{R}$, where the *total mass* is $M = \int_{\Sigma_t^{(\mathscr{R})}} \frac{T^{00}}{c^2} dx^1 dx^2 dx^3$, the left hand side is the β-th component of the total momentum.)

8.5.6 The stress-energy tensor of the ideal fluid and of the Electromagnetic field

In classical mechanics, the simplest example of continuum is an *ideal fluid* which is defined by the requirment that its stress tensor is completly isotropic: in orthonormal Cartesian coordinates of an inertial reference frame

$$\sigma^{\alpha\beta} = -p\delta^{\alpha\beta} \,, \quad \alpha,\beta = 1,2,3 \,, \tag{8.79}$$

where $p \geq 0$ is the **pressure** of the fluid at the considerd time and place in space (see Section 4.2.3 where we adopted a different notation). The pressure p is related to the other properties of the fluid as the mass density μ_0 and the velocity field v through some *constitutive relations* which depend on the type of fluid. To relativistically generalize this physical system, we observe that in the instataneous rest frame $\mathscr{R}_e$ of a particle of fluid at the event $e \in \mathbb{M}^4$, according to the discussion in the previous section,

$$(T_0)^{\alpha\beta} = p\delta^{\alpha\beta} \,, \quad (T_0)^{00} = c^2\mu_0 \,, \quad (T_0)^{0\alpha} = (T_0)^{\alpha 0} = 0 \,, \quad \alpha,\beta = 1,2,3 \,.$$

Hence, in a generic inertial reference frame $\mathscr{R}$ and co-moving Minkowskian coordinates x^0, x^1, x^2, x^3 we therefore have at the event e

$$T^{ab} = \Lambda^a{}_c \Lambda^a{}_d (T_0)^{cd} \,,$$

where Λ is the Lorentz part of the orthochronous Poincaré transformation connecting a Minkowkian coordinate frame at rest with $\mathscr{R}_e$ with the Minkowskian coordinates co-moving with $\mathscr{R}$. If the worldlines of the particles of the fluid have components V^a in $\mathscr{R}$, we can always arrange a co-moving Minkowkian coordinate system such that this reference frame and the one at instantaneous rest with a particle is a pure Lorentz matrix $\Lambda = \Lambda_V$. From (8.27),

$$\Lambda_V = \left[\begin{array}{c|c} V^0/c & \vec{V}^t/c \\ \hline \vec{V}/c & I + \frac{\vec{V}\vec{V}^t}{c^2(1+V^0/c)} \end{array}\right] .$$

In fact, this matrix transforms the vector $(c, 0, 0, 0)^t \in \mathbb{R}^4$, *representing the components of the four velocity of a particle in its rest frame*, to the vector of $\mathbb{R}^4$ whose components are the ones of V. By direct computation, the components of T in $\mathscr{R}$ are

$$T^{ab} = \mu_0 V^a V^b + p\left(g^{ab} + \frac{V^a V^b}{c^2}\right) . \tag{8.80}$$

Notice that choosing a different co-moving Minkowskian coordinate system in $\mathscr{R}$ is completely equivalent (due to the decomposition theorem of the Lorentz group) to (twice) compose the right-hand side with a spatial rotation Ω_R and to include a spacetime translation which however has no effct on the components of tensors. These operations do not change the structure of the found tensor: in all reference frames and co-moving Minkowskian coordinates T always admits the form (8.80) in components. This form is therefore intrinsic, if $\tilde{g}$ is the completely covariant representation of the metric tensor,

$$T = \mu_0 V \otimes V + p\left(\tilde{g} + \frac{1}{c^2} V \otimes V\right) ,$$

where $\mathbb{M}^4 \ni e \mapsto V(e) \in T_e\mathbb{M}^4$ is the vector field of four-velocities of the continuum and p and μ_0 are scalar functions on $\mathbb{M}^4$ with the said physical meaning.
This type of stress-energy tensor is said the **stress-energy tensor of the ideal fluid**. In (8.80), μ_0 and p are respectively the mass density and the pressure measured at rest with the particles of the fluid and V is its four velocity field.

Remarks 8.54. The form (8.80) of the stress-energy tensor is also exploited in General Relativity and cosmological applications: there the fluid is made of galaxies. In that context, the classical constraint $p \geq 0$ is relaxed, and generally speaking, negative pressures are admitted. ■

There are continuous systems which do not admit a four velocity vector field. These systems *cannot* be intepreted as continuous systems of particles and they are in particular *fields* as the *electromagnetic field*. However, also these physical systems admit their own stress-energy tensor usually produced by the Noether theorem, when the system admits a variational description.

Usually some further manipulations are necessary to assure that the final tensor is symmetric. The *symmetric stress-energy tensor of the electromagnetic field* reads

$$T^{ab}_{EM} = F^{ac}F^{b}{}_{c} - \frac{1}{4}g^{ab}F_{cd}F^{cd} \,. \tag{8.81}$$

Above F^{ab} is the electromagnetic tensor introduced in (8.46). The interpretation of the densities T^{00}_{EM} and $T^{0\alpha}_{EM}$ as the density of energy and the α-th component of the density of momentum (up to multiplicative constants) is extended to this stress-energy tensor and it is in agreement with experimental facts. $T^{\alpha 0} = T^{0\alpha}$ is in fact the corresponding component of the *Poynting vector*,

$$T^{0\alpha} = \frac{1}{c^2}\vec{E} \times \vec{B} \,,$$

which describes the density of momentum of the electromagnetic field. $T^{00}_{EM} \geq 0$ is satisfied as requested from the intpretation of T^{00}_{EM} as a density of energy up positive factors because T^{00}_{EM} has just the classical expression

$$T^{00}_{EM} = \frac{1}{c^2}\left(\vec{E}^2 + \vec{B}^2\right) \,.$$

Finally $T^{a}_{EM\,a} = 0$. When quantizing the electromagnetic field, this condition is related to the fact that photons have zero mass.

Exercises 8.55.

1. Maxwell equations are written in terms of the electromagnetic tensor F^{ab}

$$\nabla_a F^{ab} = -J^b \,, \qquad \epsilon^{abcd}\nabla_b F_{cd} = 0 \,,$$

where J is four-current density $J^a = \rho_0 V^a$, where ρ_0 is the charge density computed at rest with the particles of a continuous charged body. Prove that

$$\nabla_a T^{ab}_{EM} = -F^{bc}J_c \,.$$

So that, in particular, the conservation equation $\nabla_a T^{ab}_{EM} = 0$ is valid in the absence of charges. Observe that, according to (8.45), $F^{bc}J_c$ is the Lorentz force density acting on a particle of continuum.

2. Consider an electrically charged gas of non-interacting particles with $\rho_0 = \kappa\mu_0$ (for a constant κ with physical dimensions [charge]/[mass]) subjected to the density of Lorentz force

$$f^a = \rho_0 F^{ab}V_b$$

due to the electromagnetic field, prove that

$$\nabla_a\left(T^{ab}_{EM} + T^{ab}\right) = 0$$

where $T^{ab} = \mu_0 V^a V^b$.

Solution. $\nabla_a T^{ab}_{EM} = J_c F^{cb} = \kappa\mu_0 V_c F^{cb} = -f^b$ and $\nabla_a T^{ab} = f^b$ according to (2) Exercises 8.48.

Chapter 9

Lorentz group structure

The final goal of this chapter is to discuss the interplay of standard physical decomposition of Lorentz group in boost and spatial rotation and the polar decomposition theorem proved in chapter 7. We start by focussing again on the features of Lorentz and Poincaré groups, referring to the Lie group structure in particular.

Remarks 9.1. We assume henceforth that the reader is familiar with the basic notions of matrix Lie groups [KNS]. ■

9.1 Lie group structure, distinguished subgroups, and connected components of the Lorentz group

We have a first elementary but important result concerning Lorentz group and its Lie group structure.

Proposition 9.2. *The Lorentz group $O(1,3)$ is a Lie group which is a Lie subgroup of $GL(4,\mathbb{R})$. Similarly, the Poincaré group $IO(1,3)$ is a Lie group which can be viewed as a Lie subgroup of $GL(5,\mathbb{R})$.*

Proof. From the general theory of Lie groups [KNS] we know that to show that $O(1,3)$ is a Lie subgroup of the Lie group of $GL(2,\mathbb{R})$ it is sufficient to prove that the former is a topologically-closed algebraic subgroup of the latter. The fact that $O(1,3)$ is a closed subset of $GL(2,\mathbb{R})$, where the latter is equipped with the topology (and the differentiable structure) induced by $\mathbb{R}^{16}$, is obvious from $\Lambda^t\eta\Lambda = \eta$ since the product of matrices and the transposition of a matrix are continuous operations. Concerning the second statement, it is possible to provide $IO(1,3)$ with the structure of Lie group which is also subgroup of $GL(5,\mathbb{R})$ and that includes $O(1,3)$ and $\mathbb{R}^4$ as Lie subgroups, as follows. One start with the injective map

$$IO(1,3) \ni (\Lambda, T) \mapsto \left[\begin{array}{c|c} 1 & 0 \\ \hline T & \Lambda \end{array}\right] \in GL(5,\mathbb{R})\,, \tag{9.1}$$

and he verifies that the map is in fact an injective group homomorphism. The matrix group in the right hand side of (9.1) define, in fact a closed subset of $\mathbb{R}^{25}$ and thus of $GL(5,\mathbb{R})$. As a consequence this matrix group is a Lie group which is a Lie subgroup of $GL(5,\mathbb{R})$. $\Box$

Consider two Minkowskian frames $\mathscr{R}$ and $\mathscr{R}'$ and let x^0, x^1, x^2, x^3, x'^0, x'^1, x'^2, x'^3 be two respectively co-moving Minkowski coordinate frames. We know that $\partial_{x^0} = \mathscr{R}$ and $\partial_{x'^0} = \mathscr{R}'$, finally we know that $\mathscr{R}$ and $\mathscr{R}'$ must have the same time-orientation they being future directed. We conclude that it must be $g(\partial_{x^0}, \partial_{x'^0}) < 0$. Only transformations (8.17) satisfying that constraint may make sense physically speaking. From (8.17) one has $\partial_{x^0} = \Lambda^\mu{}_0 \partial_{x'^\mu}$, so that the requirement $g(\partial_{x^0}, \partial_{x'^0}) < 0$ is equivalent to $\Lambda_0{}^0 > 0$ which is, in turn, equivalent to $\Lambda_0{}^0 > 1$ because of the first statement in (e) in proposition 8.25. One expect that the Poincaré or Lorentz transformations fulfilling this constraint form a subgroup. Indeed this is the case and the group is called the *orthochronous subgroup*: it embodies all physically sensible transformations of coordinates between inertial frames in special relativity. The following proposition states that results introducing also two other relevant subgroups. The proof of the following proposition is immediate from (e) and (d) of proposition 8.25.

Proposition 9.3. *The subsets of $IO(1,3)$ and $O(1,3)$ defined by*

$$IO(1,3)\!\uparrow := \{(\Lambda, T) \in IO(1,3) \mid \Lambda^0{}_0 \geq 1\}\,, \quad O(1,3)\!\uparrow := \{\Lambda \in O(1,3) \mid \Lambda^0{}_0 \geq 1\}\,, \tag{9.2}$$

and called respectively **orthochronous Poincaré group** *and* **orthochronous Lorentz group**, *are (Lie) subgroups of $IO(1,3)$ and $O(1,3)$ respectively.*
The subsets of $IO(1,3)$ and $O(1,3)$ defined by

$$ISO(1,3) := \{(\Lambda, T) \in IO(1,3) \mid \det \Lambda = 1\}\,, \quad SO(1,3) := \{\Lambda \in O(1,3) \mid \det \Lambda = 1\}\,, \tag{9.3}$$

and called respectively **proper Poincaré group** *and* **proper Lorentz group**, *are (Lie) subgroups of $IO(1,3)$ and $O(1,3)$ respectively.*

We remark that the condition $\Lambda^0{}_0 \geq 1$ can be replaced with the equivalent constraint $\Lambda^0{}_0 > 0$, whereas the condition $\det \Lambda = 1$ can be replaced with the equivalent constraint $\det \Lambda > 0$. Since the intersection of a pair of (Lie) groups is a (Lie) group, we can give the following final definition.

Definition 9.4. The (Lie) subgroups of $IO(1,3)$ and $O(1,3)$ defined by

$$ISO(1,3)\!\uparrow := IO(1,3)\!\uparrow \cap ISO(1,3) \quad SO(1,3)\!\uparrow := O(1,3)\!\uparrow \cap SO(1,3) \tag{9.4}$$

are called respectively **orthochronous proper Poincaré group** and **orthochronous proper Lorentz group**. $\blacksquare$

To conclude the short landscape of properties of Lorentz group initiated in the previous chapter, we state and partially prove (the complete proof needs a result we shall achieve later) the following proposition about connected components of Lorentz group. These are obtained by starting from $SO(1,3)\uparrow$ and transforming it under the left action of the elements of discrete subgroup of $O(1,3)$: $\{I, \mathscr{P}, \mathscr{T}, \mathscr{P}\mathscr{T}\}$ where $\mathscr{T} := \eta$ and $\mathscr{P} := -\eta$ (so that $\mathscr{P}\mathscr{T} = \mathscr{T}\mathscr{P} = -I$, $\mathscr{P}\mathscr{P} = \mathscr{T}\mathscr{T} = I$). In this context $\mathscr{T}$ is called *time reversal* operator – since it changes the time orientation of causal vectors – and $\mathscr{P}$ is also said to be the *(parity) inversion* operator – since it corresponds to the spatial inversion in the rest space.

Proposition 9.5. *$SO(1,3)$ admits four connected components which are respectively, with obvious notation, $SO(1,3)\uparrow$, $\mathscr{P}SO(1,3)\uparrow$, $\mathscr{T}SO(1,3)\uparrow$, $\mathscr{P}\mathscr{T}SO(1,3)\uparrow$. Only the first is a subgroup.*

Proof. By construction: (1) if $\Lambda \in \mathscr{P}SO(1,3)\uparrow$, both $\det\Lambda = -1$ and ${\Lambda^0}_0 \geq 1$, (2) if $\Lambda \in \mathscr{T}SO(1,3)\uparrow$, both ${\Lambda^0}_0 \leq -1$ and $\det\Lambda = -1$, (3) if $\Lambda \in \mathscr{P}\mathscr{T}SO(1,3)\uparrow$, both ${\Lambda^0}_0 \leq -1$ and $\det\Lambda = 1$.
Thus the last statement is an immediate consequence of the fact that, as I satisfies $\det I = 1$ and $(I)^0_0 = 1$, it cannot belong to the three sets by construction $\mathscr{P}SO(1,3)\uparrow$, $\mathscr{T}SO(1,3)\uparrow$, $\mathscr{P}\mathscr{T}SO(1,3)\uparrow$. Assume that $SO(1,3)\uparrow$ is connected. We shall prove it later. Since the maps $O(1,3) \ni \Lambda \mapsto \mathscr{T}\Lambda$ and $O(1,3) \ni \Lambda \mapsto \mathscr{P}\Lambda$ are continuous, they transform connected sets to connected sets. As a consequence $\mathscr{P}SO(1,3)\uparrow$, $\mathscr{T}SO(1,3)\uparrow$, $\mathscr{P}\mathscr{T}SO(1,3)\uparrow$ are connected sets. To conclude, it is sufficient to prove that the considered sets are pairwise disconnected. To this end it is sufficient to exhibit continuous real-valued function defined on $O(1,3)$ which cannot vanish but they change their sign passing from a set to the other[1]. By construction two functions are sufficient $O(1,3) \ni \Lambda \mapsto \det\Lambda$ and $O(1,3) \ni \Lambda \mapsto {\Lambda^0}_0$. $\square$

9.2 Spatial rotations and boosts

The final aim of this chapter is to state and prove the decomposition theorem for the group $SO(1,3)\uparrow$. To this end, two ingredients have to be introduced: spatial rotations and boosts, which are distinguished types of Lorentz transforms in $SO(1,3)\uparrow$.

9.2.1 Spatial rotations

Consider a reference system $\mathscr{R}$, what is the relation between two Minkowskian coordinate frames x^0, x^1, x^2, x^3 and x'^0, x'^1, x'^2, x'^3 both co-moving with $\mathscr{R}$? These transformations are called **internal** to $\mathscr{R}$. The answer is quite simple. The class of all internal Poincaré transformations is completely obtained by imposing the further constraint $\partial_{x^0} = \partial_{x'^0} (= \mathscr{R})$ on the equations (8.17) and assuming that $\Lambda \in O(1,3)\uparrow$.

[1]Indeed, assuming that X is a topological space and $f : X \to \mathbb{R}$ a continuous function, if X is connected $f(X)$ is so. Therefore, if $a, b \in f(X)$ with $a < 0$ and $b > 0$, then $f(X)$ is a connected subset of $\mathbb{R}$ including a, b. The connected subsets of $\mathbb{R}$ are the intervals, so that $f(X)$ has to contain all reals between $a < 0$ and $b > 0$, in particular $0 \in f(X)$. If f cannot vanish, X cannot be connected.

Since $\partial_{x^0} = \Lambda^\mu{}_0 \partial_{x'^\mu}$, the constraint is equivalent to impose $\Lambda^0{}_0 = 1$ and $\Lambda^i{}_0 = 0$ for $i = 1, 2, 3$. Lorentz condition $\Lambda^t \eta \Lambda = \eta$ implies in particular that:

$$(\Lambda^0{}_0)^2 = 1 + \sum_{i=1}^{3} \Lambda^0{}_i \Lambda^0{}_i ,$$

and thus, since $\Lambda^0{}_0 = 1$, we find that $\Lambda^0{}_i = 0$. Summing up, internal Lorentz transformations must have the already seen form

$$\Omega_R = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & R \end{array}\right] , \tag{9.5}$$

By direct inspection one finds that, in this case, $\Omega^t \eta \Omega = \eta$ reduces to

$$R^t R = I , \tag{9.6}$$

This is nothing but the equation determining the orthogonal group $O(3)$. Conversely, starting from any matrix $R \in O(3)$ and thus satisfying (9.6), and defining Ω_R as in (9.5), it is immediate to verify that $\Omega_R \in O(1,3)\uparrow$ and

$$x'^\mu = (\Omega_R)^\mu{}_\nu (x^\nu + T^\nu) ,$$

with $T \in \mathbb{R}^4$ fixed arbitrarily, is an internal Poincaré transformation. It is immediate to show also that Ω_R with form (9.5) belongs to $SO(1,3)\uparrow$ if and only if $R \in SO(3)$, the group of special rotations made of rotations of $O(3)$ with unitary determinant.

Remarks 9.6. It is worthwhile noticing, from a kinematic point of view that, the velocity of $\mathscr{R}$ with respect to $\mathscr{R}'$ seen in Section 8.3.3 is invariant under changes of co-moving Minkowskian coordinates when the transformations of coordinates are internal to $\mathscr{R}$ and $\mathscr{R}'$. We leave the simple proof of this fact to the reader. ■

Definition 9.7. The Lorentz transformations Ω_R defined in (9.5) with $R \in O(3)$ are called **spatial rotations**. If $R \in SO(3)$, Ω_R is called **spatial proper rotations**. ■

Since the translational part is trivial, from now on we will focus on the Lorentz part of Poincaré group only.

9.2.2 Lie algebra analysis

Focusing on the Lorentz group, we wish to extract the *non internal* part of a Lorentz transformation, i.e. what remains after one has taken spatial rotations into account. This goal will be achieved after a preliminary analysis of the Lie algebra of $SO(1,3)\uparrow$ and the corresponding

exponentiated operators. As $SO(1,3)\!\uparrow$ is a matrix Lie group which is Lie subgroup of $GL(4,\mathbb{R})$, its Lie algebra can be obtained as a Lie algebra of matrices in $M(4,\mathbb{R})$ with the commutator $[\cdot,\cdot]$ given by the usual matrix commutator. The topology and the differential structure on $SO(1,3)\!\uparrow$ are those induced by $\mathbb{R}^{16}$.
As the maps $O(1,3)\ni\Lambda\mapsto\det\Lambda$ $O(1,3)\ni\Lambda\mapsto{\Lambda^0}_0$ are continuous and $\det I=1$ and ${(I)^0}_0>0$, every $\Lambda\in O(1,3)$ sufficiently close to I must belong to $SO(1,3)\!\uparrow$ and *viceversa*. Hence the Lie algebra of $O(1,3)$, $o(1,3)$, coincides with that of its Lie subgroup $SO(1,3)\!\uparrow$ because it is completely determined by the behavior of the group in an arbitrarily small neighborhood of the identity.

Proposition 9.8. *The Lie algebra of $SO(1,3)\!\uparrow$ admits a vector basis made of the following 6 matrices called* **boost generators** K_1,K_2,K_3 *and* **spatial rotation generators** S_1,S_2,S_3*:*

$$K_1=\left[\begin{array}{c|ccc}0&1&0&0\\\hline 1&0&0&0\\0&0&0&0\\0&0&0&0\end{array}\right]\;,\;K_2=\left[\begin{array}{c|ccc}0&0&1&0\\\hline 0&0&0&0\\1&0&0&0\\0&0&0&0\end{array}\right]\;,\;K_3=\left[\begin{array}{c|ccc}0&0&0&1\\\hline 0&0&0&0\\0&0&0&0\\1&0&0&0\end{array}\right]\;. \tag{9.7}$$

$$S_i=\left[\begin{array}{c|ccc}0&0&0&0\\\hline 0&&&\\0&&T_i&\\0&&&\end{array}\right]\quad\textit{with}\;\;T_1=\begin{bmatrix}0&0&0\\0&0&-1\\0&1&0\end{bmatrix}\;,\;T_2=\begin{bmatrix}0&0&1\\0&0&0\\-1&0&0\end{bmatrix}\;,\;T_3=\begin{bmatrix}0&-1&0\\1&0&0\\0&0&0\end{bmatrix} \tag{9.8}$$

These generators enjoy the following commutation relations, which, as a matter of facts, determines the structure tensor of $o(1,3)$:

$$[S_i,K_j]=\sum_{k=1}^{3}\epsilon_{ijk}K_k\;,\quad [S_i,S_j]=\sum_{k=1}^{3}\epsilon_{ijk}S_k\quad [K_i,K_j]=-\sum_{k=1}^{3}\epsilon_{ijk}S_k\;. \tag{9.9}$$

Above ϵ_{ijk} is the usual completely antisymmetric Ricci indicator with $\epsilon_{123}=1$.

Proof. If a matrix $N\in M(4,\mathbb{R})$ is in $o(1,3)$, the generated one-parameter subgroup $\{e^{uN}\}_{u\in\mathbb{R}}$ in $GL(4,\mathbb{R})$ satisfies $(e^{uN})^t\eta e^{uN}=\eta$ for u in a neighborhood of 0, that is $e^{uN^t}\eta e^{uN}=\eta$ for the same values of u. Taking the derivative at $t=0$ one gets the necessary condition

$$N^t\eta+\eta N=0\,. \tag{9.10}$$

These equations are also sufficient. Indeed, from standard properties of the exponential map of matrices, one has

$$\frac{d}{du}\left(e^{uN^t}\eta e^{uN}-\eta\right)=e^{uN^t}\left(N^t\eta+\eta N\right)e^{uN}\,.$$

Thus, the validity of (9.10) implies that $e^{uN^t}\eta e^{uN}-\eta=$ constant. For $u=0$ one recognizes that the constant is 0 and so $(e^{uN})^t\eta e^{uN}=\eta$ is valid (for every $u\in\mathbb{R}$). Eq (9.10) supplies 10 linearly independent conditions so that it determines a subspace of $M(4,\mathbb{R})$ with dimension 6. The 6 matrices $S_i, K_j \in M(4.\mathbb{R})$ are linearly independent and satisfy (9.10), so they are a basis for $o(1,3)$. The relations (9.9) can be checked by direct inspection. $\Box$

From now on $\mathbf{K}, \mathbf{S}, \mathbf{T}$ respectively denote the formal vector with components K_1, K_2, K_3, the formal vector with components S_1, S_2, S_3 and the formal vector with components T_1, T_2, T_3. $\mathbb{S}^2$ will indicate the sphere of three-dimensional unit vectors.
Generators S_1, S_2, S_3 produce proper spatial rotations as stated in the following proposition.

Proposition 9.9. *The following facts about proper rotations hold.*

(a) Every proper spatial rotations has the form $\Omega_R = e^{\theta \mathbf{n}\cdot\mathbf{S}}$ – or equivalently $R = e^{\theta\mathbf{n}\cdot\mathbf{T}}$ for all $R \in SO(3)$ – with suitable $\theta \in \mathbb{R}$ and $\mathbf{n} \in \mathbb{S}^2$ depending on R.

(b) Every matrix $e^{\theta\mathbf{n}\cdot\mathbf{S}}$ with $\theta \in \mathbb{R}$ and $\mathbf{n} \in \mathbb{S}^2$ is a proper rotation Ω_R, and is associated with $R = e^{\theta\mathbf{n}\cdot\mathbf{T}} \in SO(3)$.

(c) The following equivalent identities hold true, for every $U \in SO(3)$, $\mathbb{N} \in \mathbb{S}^2$, $\theta \in \mathbb{R}$:

$$\Omega_U e^{\theta\mathbf{n}\cdot S}\Omega_U{}^t = e^{\theta(U\mathbf{n})\cdot S}\,, \quad U e^{\theta\mathbf{n}\cdot T}U^t = e^{\theta(U\mathbf{n})\cdot T}\,. \tag{9.11}$$

The latter holds, more generally, also if $U \in SL(3,\mathbb{C})$.

(d) The Lie group of the spatial proper rotations $SO(3)$ is connected, but not simply connected, its fundamental group being $\pi_1(SO(1)) = \mathbb{Z}_2$.

Proof. (a) and (c). Since, from the given definitions,

$$e^{i\theta\mathbf{n}\cdot S} = \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & e^{\mathbf{n}\cdot\mathbf{T}} & \\ 0 & & & \end{array}\right], \tag{9.12}$$

it is obvious that $\Omega_R = e^{\theta\mathbf{n}\cdot\mathbf{S}}$ are completely equivalent $R = e^{\theta\mathbf{n}\cdot\mathbf{T}}$, so we deal with the latter.
If $R \in SO(3)$, the induced operator in $\mathbb{R}+i\mathbb{R}$ is unitary and thus it admits a base of eigenvectors with eigenvalues λ_i with $|\lambda_i| = 1$, $i = 1,2,3$. As the characteristic polynomial of R is real, an eigenvalue must be real, the remaining pair of eigenvalues being either real or complex and conjugates. Since $\det R = \lambda_1\lambda_2\lambda_3 = 1$, 1 is one of the eigenvalues. If another eigenvalue coincides with 1 all three eigenvalues must do it and $R = I$. In this case every non-vanishing real vector is an eigenvector of R. Otherwise, the eigenspace of $\lambda = 1$ must be one-dimensional and thus, as R is real, it must contain a real eigenvector. We conclude that, in every case, R has a real normalized eigenvector $\mathbf{n}$ with eigenvalue 1. Consider an orthonormal basis $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3 := \mathbf{n}$, related with the initial one by means of $R' \in SO(3)$, and represent R in the new basis. Imposing the requirement that $\mathbf{n}_3$ is an eigenvector with eigenvalue 1 as well as that the represented transformation belong to $SO(3)$, one can easily prove that, in such a base, R

is represented by the matrix $e^{\theta T_3}$ for some $\theta \in [0, 2\pi]$. In other words, coming back to the initial basis, $R = R' e^{\theta T_3} R'^t$ for some $R' \in SO(3)$. Now notice that $(T_i)_{jk} = -\epsilon_{ijk}$. This fact entails that $\sum_{i,j,k} U_{pi} U_{qj} U_{rk} \epsilon_{ijk} = \epsilon_{pqr}$ for all $U \in SL(3, \mathbb{C})$. That identity can be re-written as $\mathbf{n} \cdot U\mathbf{T}U^t = (U\mathbf{n}) \cdot \mathbf{T}$ for every $U \in SL(3, \mathbb{C})$. By consequence, if $U \in SO(3)$ in particular, it also holds $Ue^{\theta \mathbf{n} \cdot \mathbf{T}} U^t = e^{\theta (U\mathbf{n}) \cdot \mathbf{T}}$. (This proves the latter in (9.11), the former is a trivial consequence of the given definitions). Therefore, the identity found above for any $R \in SO(3)$, $R = R' e^{\theta T_3} R'^t$ with $R' \in SO(3)$, can equivalently be written as $R = e^{\theta \mathbf{n} \cdot \mathbf{T}}$ for some versor $\mathbf{n} = R' \mathbf{e}_3$.
(b) Finally, every matrix $e^{\theta \mathbf{n} \cdot \mathbf{T}}$ belongs to $SO(3)$ because $(e^{\theta \mathbf{n} \cdot \mathbf{T}})^t = e^{\theta \mathbf{n} \cdot \mathbf{T}^t} = e^{-\theta \mathbf{n} \cdot \mathbf{T}} = (e^{\theta \mathbf{n} \cdot \mathbf{T}})^{-1}$ and $\det e^{\theta \mathbf{n} \cdot \mathbf{T}} = e^{\theta \mathbf{n} \cdot \mathrm{tr}\, \mathbf{T}} = e^0 = 1$.
(e) We sketch here the idea of the proof only. First consider the subgroup made of marices $e^{\theta \mathbf{e}_3 \cdot \mathbf{T}}$ with $\theta \in \mathbb{R}$. By the explicit form of these matrices, i.e.

$$T_3 = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{9.13}$$

one see that, all these matrices are biunivocally determined by couples $(\theta, \mathbf{n}) \in [0, \pi] \times \{-\mathbf{e}_3, \mathbf{e}_3\}$ with the only exception given by the pairs $(\pi, -\mathbf{e}_3)$ and $(\pi, \mathbf{e}_3)$ which produces the same rotation. This result generalizes to the subgroup $\{e^{\theta \mathbf{n} \cdot \mathbf{T}}\}_{\theta \in \mathbb{R}}$ for every fixed $\mathbf{n} \in \mathbb{S}^2$, in view of (c). From now on we restrict θ to range in $[0, \pi]$. By direct inspection, reducing to the case $\mathbf{n} = \mathbf{e}_3$ using (c) again, one sees that, $e^{\theta \mathbf{n} \cdot \mathbf{T}} = e^{\theta' \mathbf{n}' \cdot \mathbf{T}}$ is possible only if $\mathbf{n} = \mathbf{n}'$ and $\theta = \theta'$, with the only exception $\theta = \theta' = \pi$, where also $\mathbf{n} = -\mathbf{n}'$ is allowed as a consequence of the analysis above. The proof of this fact is an immediate consequence of the fact that $\mathbf{n}'$ is an eigenvector (with eigenvalue 1) of $e^{\theta' \mathbf{n}' \cdot \mathbf{T}}$, so that, if $e^{\theta \mathbf{e}_3 \cdot \mathbf{T}} = e^{\theta' \mathbf{n}' \cdot \mathbf{T}}$ $\mathbf{n}'$ must be an eigenvector of the matrix (9.13) (with eingenvalue 1). As a conclusion we obtain that $SO(3)$ in biunivocally defined by the points of a set B constructed as follows. B is the ball in $\mathbb{R}^3$ with radius π, where each two points on the surface of the ball which belong to the same diameter are identified (in other words, the pairs $(\pi, \mathbf{n})$ and $(\pi, -\mathbf{n})$ determins the same element of $SO(3)$ as we said above). A closer scrutiny proves that this one-to-one correspondence is actually a homeomorphism when B is endowed with the natural topology indiced by $\mathbb{R}^3$ and the said identifications. The group of continuous closed paths in B is $\mathbb{Z}_2$ as one may symply prove. □

9.2.3 Boosts

From the analysis performed above, we conclude that what remains of a Lorentz transformation in $SO(1,3)\uparrow$, one one has taken spatial rotations into account, are transformations obtained by exponentiating the generators $\mathbf{K}$.

Definition 9.10. The elements of $SO(1,3)\uparrow$ with the form $\Lambda = e^{\chi \mathbf{m} \cdot \mathbf{K}}$, with $\chi \in \mathbb{R}$ and $\mathbf{m} \in \mathbb{S}^2$, are called **boosts** or **pure transformations**. ■

Let us investigate the basic properties of boosts. These are given by the following proposition.

Proposition 9.11. *The boost enjoy the following properties.*

(a) All matrices $\Lambda = e^{\chi \mathbf{m}\cdot\mathbf{K}}$ with arbitrary $\chi \in \mathbb{R}$ and $\mathbf{m} \in \mathbb{S}^2$ belong to $SO(1,3)\!\uparrow$ and thus are boosts.

(b) For every pair $\mathbf{m} \in \mathbb{S}^1$, $\chi \in \mathbb{R}$ and every $R \in SO(3)$ one has

$$\Omega_R e^{\chi \mathbf{m}\cdot\mathbf{K}} \Omega_R{}^t = e^{\chi (R\mathbf{m})\cdot\mathbf{K}} . \tag{9.14}$$

(c) For $\mathbf{n} \in \mathbb{S}^2$, the explicit form of $e^{\chi \mathbf{n}\cdot\mathbf{K}}$ reads:

$$e^{\chi \mathbf{n}\cdot\mathbf{K}} = \left[\begin{array}{c|c} \cosh\chi & (\sinh\chi)\mathbf{n}^t \\ \hline (\sinh\chi)\mathbf{n} & I - (1-\cosh\chi)\mathbf{n}\mathbf{n}^t \end{array}\right] , \tag{9.15}$$

(d) Every boost is symmetric and (strictly) positive definite.

(e) For every fixed $\mathbf{m} \in \mathbb{S}^2$, $\{e^{\chi \mathbf{m}\cdot\mathbf{K}}\}_{\chi\in\mathbb{R}}$ is a subgroup of $SO(1,3)$. However, the set of all the boost transformations is not a subgroup of $SO(1,3)$ and it is homeomorphic to $\mathbb{R}^3$ when equipped with the topology induced by $SO(1,3)\!\uparrow$.

Proof. (a) It has been proved in the proof of proposition 9.8, taking into account that $N := \chi \mathbf{m}\cdot\mathbf{K} \in o(1,3)$.
(b) Fix $\mathbf{n} \in \mathbb{S}^2$ and $j = 1,2,3$. Now, for $i = 1,2,3$ define the functions

$$f_j(\theta) := e^{\theta \mathbf{n}\cdot S} K_j e^{\theta \mathbf{n}\cdot S} , \quad g_j(\theta) := \sum_{k=1}^{3} \left(e^{\theta \mathbf{n}\cdot T}\right)_{jk} K_k .$$

Taking the first derivative in θ and using both (9.9) and the explicit form of the matrices T_h, one finds that the smooth functions f_k and the smooth functions g_k satisfies the same system of differential equation of order 1 written in normal form: for $j = 1,2,3$,

$$\frac{df_j}{d\theta} = \sum_{k=1}^{3} n_k \epsilon_{kjp} f_p(\theta) , \quad \frac{dg_j}{d\theta} = \sum_{k=1}^{3} n_k \epsilon_{kjp} g_p(\theta) .$$

Since $f_j(0) = g_j(0)$ for $j = 1,2,3$, we conclude that these functions coincide for every $\theta \in \mathbb{R}$:

$$e^{\theta \mathbf{n}\cdot S} K_j e^{\theta \mathbf{n}\cdot S} = \sum_{k=1}^{3} \left(e^{\theta \mathbf{n}\cdot T}\right)_{jk} K_k .$$

Now, by means of exponentiation we get (9.14) exploiting (c) of proposition 9.9.
(c) First consider the case $\mathbf{n} = \mathbf{e}_3$. In this case directly from Taylor's expansion formula

$$e^{\chi \mathbf{e}_3\cdot\mathbf{K}} = \left[\begin{array}{c|ccc} \cosh\chi & 0 & 0 & \sinh\chi \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh\chi & 0 & 0 & \cosh\chi \end{array}\right] = \left[\begin{array}{c|c} \cosh\chi & (\sinh\chi)\mathbf{e}_3^t \\ \hline (\sinh\chi)\mathbf{e}_3 & I - (1-\cosh\chi)\mathbf{e}_3\mathbf{e}_3^t \end{array}\right] . \tag{9.16}$$

If $\mathbf{n} \in \mathbb{S}^2$ there is $R \in SO(3)$ such that $\mathbf{n} = R\mathbf{e}_3$. Using this R in (b) with $\mathbf{m} = \mathbf{e}_3$ one gets (9.15) with $\mathbf{n}$.
(d) Symmetry is evident from (9.15). Using (b), it is sufficient to prove positivity if $\mathbf{m} = \mathbf{e}_3$. In this case strictly positivity can be checked by direct inspection using (9.16).
(e) The first statement is trivial, since $\{e^{\chi \mathbf{m}\cdot\mathbf{K}}\}_{\chi\in\mathbb{R}}$ is a one-parameter subgroup of $SO(1,3)\uparrow$. By direct inspection using (9.15) for $\mathbf{n} = \mathbf{e}_1$ and $\mathbf{n} = \mathbf{e}_3$ one verifies that the product of these boosts cannot be represented as in (9.15). Concerning the last statement, one proves that $e^{\mathbf{u}\cdot\mathbf{K}} = e^{\mathbf{u}'\cdot\mathbf{K}}$ implies $\mathbf{u} = \mathbf{u}'$, reducing to study the simpler case $e^{\chi\mathbf{e}_3\cdot\mathbf{K}} = e^{\mathbf{u}'\cdot\mathbf{K}}$, making use of (b). This case can be examined by taking advantage of (9.16) and (9.15), yielding to $\chi\mathbf{e}_3 = \mathbf{u}'$ straightforwardly. Summarizing, $\mathbb{R}^3 \ni \mathbf{u} \mapsto e^{\mathbf{u}\cdot\mathbf{K}}$ is bijective onto the set of all boost transformations. Directly from (9.16) one see that the map $\mathbb{R}^3 \ni \mathbf{v} \to e^{\mathbf{v}\cdot\mathbf{K}} \in \mathbb{R}^{16}$ is a smooth embedding and defines a submanifold of $\mathbb{R}^{16}$. A way to verify it is to identify the three components of $\mathbf{v} = (\sinh\chi)\mathbf{n}$ with the coordinates x^{ij}, when $i = 1$ and $j = 2,3,4$, in the space $\mathbb{R}^{16}$ of the real 4×4 matrices, and prove that the hypotheses of the theorem of regular values are valid for the equations $f^{ij} := x^{ij} - x^{ij}(x^{12}, x^{13}, x^{14}) = 0$, which determines the other 12 components of the boost $e^{\mathbf{v}\cdot\mathbf{K}} \in \mathbb{R}^{16}$, referring to the its explicit experssion as given in (9.16). In particular the Jacobian sub-matrix with elements $\{\partial f^{rs}/\partial x^{ij}\}_{r,s=1,\ldots,16, i=1, j=2,3,4}$ has range 3 and thus $(x^{12}, x^{13}, x^{14}) = \mathbf{v}$ are (global) admissible coordinates on the space of boosts, and the map $\mathbb{R}^3 \ni \mathbf{v} \to e^{\mathbf{v}\cdot\mathbf{K}} \in \mathbb{R}^{16}$, restricted to its image in the co-domain, define a diffeomorphism from $\mathbb{R}^3$ onto the space of the boost. Thus in particular, the set of the boosts is homeomorphic to $\mathbb{R}^3$ when endowed with the topology induced by $\mathbb{R}^{16}$. Since the topology of $SO(1,3)\uparrow$ is that induced by $\mathbb{R}^{16}$ itself, we conclude that the set of the boosts is homeomorphic to $\mathbb{R}^3$ when endowed with the topology induced by $SO(1,3)\uparrow$. $\square$

Recalling in Section 8.3.3, the kinematic meaning of parameters χ and $\mathbf{m}$ in boosts is clear from the following last proposition.

Proposition 9.12. *Let $\mathscr{R}$ and $\mathscr{R}'$ be Minkowskian reference with associated co-moving Minkowskian coordinate frames x^0, x^1, x^2, x^3 and x'^0, x'^1, x'^2, x'^3 respectively and suppose that (8.17) hold with*

$$\Lambda = e^{\chi\mathbf{m}\cdot\mathbf{K}}\,.$$

Let γ represent the world line of a point matter at rest with respect to $\mathscr{R}$, that is, γ admits parametrization $x^i(\xi) = x^i_0$ constant for $i = 1,2,3$, $x^0 = x^0(\xi)$. The velocity of γ with respect to $\mathscr{R}'$ does not depend on x^i_0 and it is constant in $\mathscr{R}'$-time so that, indicating it by $\mathbf{v}^{(\mathscr{R}')}_{\mathscr{R}}$, it holds

$$\mathbf{v}^{(\mathscr{R}')}_{\mathscr{R}} = (\tanh\chi)\mathbf{m}\,.$$

Proof. The proof follows immediately from (9.15) and definition 8.18. $\square$

Remarks 9.13. Boosts along $\mathbf{n} := \mathbf{e}_3$ are know in the literature as "special Lorentz transformations" along z (an analogous name is given replacing $\mathbf{e}_3$ with $\mathbf{e}_1$ and $\mathbf{e}_2$). ■

9.2.4 Decomposition theorem for $SO(1,3)\uparrow$

In the previous subsection we have determined two classes of transformations in $SO(1,3)\uparrow$: spatial pure rotations and boosts. It is natural to wonder if those transformations encompass, up to products of some of them, the whole group $SO(1,3)\uparrow$. The answer is positive and is stated in a theorem of decomposition of $SO(1,3)\uparrow$ we go to state. The proof of the theorem will be given in a specialized next section using a nonstandard approach based on the so-called polar decomposition theorem.

Theorem 9.14. *Take $\Lambda \in SO(1,3)\uparrow$, the following holds.*

(a) There is exactly one boost P and exactly a spatial pure rotation U (so that $P = e^{\chi \mathbf{m}\cdot\mathbf{K}}$, $U = e^{\theta \mathbf{m}\cdot\mathbf{S}}$ for some $\chi, \theta \in \mathbb{R}$ and $\mathbf{m}, \mathbf{n} \in \mathbb{S}^2$) such that

$$\Lambda = UP\,.$$

(b) There is exactly one boost P' and exactly a spatial pure rotation U' (so that $P' = e^{\chi' \mathbf{m}'\cdot\mathbf{K}}$, $U' = e^{\theta' \mathbf{m}'\cdot\mathbf{S}}$ for some $\chi', \theta' \in \mathbb{R}$ and $\mathbf{m}', \mathbf{n}' \in \mathbb{S}^2$) such that

$$\Lambda = P'U'\,.$$

(c) It holds $U' = U$ and $P' = UPU^t$.

9.2.5 Proof of decomposition theorem by polar decomposition of $SO(1,3)\uparrow$

$\Lambda \in SO(1,3)\uparrow$ can be seen as a linear operator in the finite-dimensional real vector space $\mathbb{R}^4$. Therefore one may consider the polar decomposition $\Lambda = PU = U'P'$ given in theorem 7.10. $U = U'$ are now orthogonal operators of $O(4)$ and P, P' are a symmetric positive operators. *A priori* those decompositions could be physically meaningless because U and P, P' could not to belong to $SO(1,3)\uparrow$: the notions of symmetry, positiveness, orthogonal group $O(4)$ are refereed to the positive scalar product of $\mathbb{R}^4$ instead of the indefinite Lorentz scalar product. Nevertheless we shall show that the polar decompositions of $\Lambda \in SO(1,3)\uparrow$ are in fact physically meaningful. Indeed, they coincides with the known decompositions of Λ in spatial-rotation and boost parts as in theorem 9.14.
Let us focus attention on the real vector space $V = \mathbb{R}^4$ endowed with the usual positive scalar product. In that case $\mathscr{L}(V|V) = M(4,\mathbb{R})$, orthogonal operators are the matrices of $O(4)$ and the adjoint $A^\dagger$ of $A \in \mathscr{L}(V|V)$ (see chapter 7) coincides with the transposed matrix A^t, therefore symmetric operators are symmetric matrices. We have the following theorem which proves theorem 9.14 as an immediate consequence.

Theorem 9.15. *If $UP = P'U = \Lambda$ (with $P' = UPU^t$) are polar decompositions of $\Lambda \in SO(1,3)\uparrow$,* *(a) $P, P', U \in SO(1,3)\uparrow$, more precisely P, P' are boosts and U a spatial proper rotation;*

(b) there are no other decompositions of Λ as a product of a Lorentz boost and a spatial proper rotation different from the two polar decompositions above.

Proof. If $P \in M(4,\mathbb{R})$ we exploit the representation:

$$P = \left[\begin{array}{c|c} g & B^t \\ \hline C & A \end{array}\right] , \tag{9.17}$$

where $g \in \mathbb{R}$, $B, C \in \mathbb{R}^3$ and $A \in M(3,\mathbb{R})$.
(a) We start by showing that $P, U \in O(1,3)$. As $P = P^t$, $\Lambda^t\eta\Lambda = \Lambda$ entails $PU^t\eta UP = \eta$. As $U^t = U^{-1}$ and $\eta^{-1} = \eta$, the obtained identity is equivalent to $P^{-1}U^t\eta UP^{-1} = \eta$ which, together with $PU^t\eta UP = \eta$, implies $P\eta P = P^{-1}\eta P^{-1}$, namely $\eta P^2\eta = P^{-2}$, where we have used $\eta = \eta^{-1}$ once again. Both sides are symmetric (notice that $\eta = \eta^t$) and positive by construction, by theorem 7.9 they admit unique square roots which must coincide. The square root of P^{-2} is P^{-1} while the square root of $\eta P^2\eta$ is $\eta P\eta$ since $\eta P\eta$ is symmetric positive and $\eta P\eta\eta P\eta = \eta PP\eta = \eta P^2\eta$. We conclude that $P^{-1} = \eta P\eta$ and thus $\eta = P\eta P$ because $\eta = \eta^{-1}$. Since $P = P^t$ we have found that $P \in O(1,3)$ and thus $U = \Lambda P^{-1} \in O(1,3)$. Let us prove that $P, U \in SO(1,3)\uparrow$. $\eta = P^t\eta P$ entails det $P = \pm 1$, on the other hand $P = P^t$ is positive and thus det $P \geq 0$ and $P^0{}_0 \geq 0$. As a consequence det $P = 1$ and $P^0{}_0 \geq 0$. We have found that $P \in SO(1,3)\uparrow$. Let us determine the form of P using (9.17). $P = P^t$, $P \geq 0$ and $P\eta P = \eta$ give rise to the following equations: $C = B$, $0 < g = \sqrt{1+B^2}$, $AB = gB$, $A = A^*$, $A \geq 0$ and $A^2 = I + BB^t$. Since $I + BB^t$ is positive, the solution of the last equation $A = \sqrt{A^2} = I + BB^t/(1+g) \geq 0$ is the unique solution by theorem 7.9. We have found that a matrix $P \in O(1,3)$ with $P \geq 0$, $P = P^*$ must have the form

$$P = \left[\begin{array}{c|c} \cosh\chi & (\sinh\chi)\mathbf{n}^t \\ \hline (\sinh\chi)\mathbf{n} & I - (1-\cosh\chi)\mathbf{n}\mathbf{n}^t \end{array}\right] = e^{\chi\mathbf{n}\cdot\mathbf{K}} , \tag{9.18}$$

where we have used the parameterization $B = (\sinh\chi)\mathbf{n}$, $\mathbf{n}$ being any versor in $\mathbb{R}^3$ and $\chi \in \mathbb{R}$. By (c) in proposition we have found that P is a boost. (The same proofs apply to P'.)
Let us pass to consider U. Since $\Lambda, P \in SO(1,3)\uparrow$, from $\Lambda P^{-1} = U$, we conclude that $U \in SO(1,3)\uparrow$. $U\eta = \eta(U^t)^{-1}$ (i.e. $U \in O(1,3)$) and $U^t = U^{-1}$ (i.e. $U \in O(4)$) entail that $U\eta = \eta U$ and thus the eigenspaces of η, E_λ (with eigenvalue λ), are invariant under the action of U. In those spaces U acts as an element of $O(\dim(E_\lambda))$ and the whole matrix U has a block-diagonal form. $E_{\lambda=-1}$ is generated by $\mathbf{e}_0$ and thus U reduces to $\pm I$ therein. The sign must be $+$ because of the requirement $U^0{}_0 > 0$. The eigenspace $E_{\lambda=1}$ is generated by $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and therein U reduces to an element of $R \in O(3)$. Actually the requirement $\det U = 1$ (together with $U^0{}_0 = 1$) implies that $R \in SO(3)$ and thus one has that $U = \Omega_R$ is a spatial pure rotation

as well.
(b) Suppose that $\Omega B = \Lambda \in SO(1,3)\uparrow$ where B is a pure boost and Ω is a spatial proper rotation. $\Omega \in O(4)$ by construction, on the other hand, from (d) of proposition 9.11: $B^t = B > 0$ Thus $\Omega B = \Lambda$ is a polar decomposition of Λ. Uniqueness of polar decomposition (theorem 7.10) implies that $\Omega = U$ and $B = P$. The other case is analogous. $\Box$

We emphasize here a topological consequence of this theorem with the following proposition.

Proposition 9.16. *$SO(1,3)\uparrow$ is homeomorphic to $SO(3) \times \mathbb{R}^3$, as a consequence it is connected, arch-connected and $\pi_1(SO(1,3)\uparrow) = \mathbb{Z}_2$, so that it is not simply connected.*

Proof. Using the fact that $SO(1,3)\uparrow$ is bijectively identified with the product $SO(3) \times \mathbb{R}^3$, where the $\mathbb{R}^3$ is the space of the boosts as in (c) of proposition 9.11, it arises that $SO(1,3)\uparrow$ is homeomorphic to the topological space $SO(3) \times \mathbb{R}^3$. Indeed, the map $\mathbb{R}^{16} \ni \Lambda \mapsto (U,P) \in \mathbb{R}^{16} \times \mathbb{R}^{16}$ is continuous since the functions used to construct the the factors U and P of the polar decomposition of a nonsingular matrix are continuous in the $\mathbb{R}^{16}$ topology; on the other hand the multiplication of matrices $\mathbb{R}^{16} \times \mathbb{R}^{16} \ni (U,P) \mapsto UP \in \mathbb{R}^{16}$ is trivially continuos in the same topology; and finally the topologies used in the restricted spaces of the matrices here employed, i.e. $\Lambda \in SO(1,3)\uparrow$, $U \in SO(3)$ and $P \in \mathbb{R}^3$ (space of the boosts), are actually those induced by $\mathbb{R}^{16}$. As soon as both the factors of $SO(3) \times \mathbb{R}^3$ are connected, $SO(1,3)\uparrow$ is connected too. It is also arch connected, it being a differentiable manifold and thus admitting a topological base made of (smooth-)arch-connected open sets. Moreover, the fundamental group of $SO(1,3)\uparrow$ is, as it happens for product manifolds [Sernesi], the product of the fundamental groups of the factors. As $\pi_1(SO(3)) = \mathbb{Z}_2$ by (d) in proposition 9.9, whereas $\pi_1\left(\mathbb{R}^3\right) = \{1\}$, one finds again $\pi_1(SO(1,3)\uparrow) = \mathbb{Z}_2$. $\Box$

Chapter 10

$SL(2,\mathbb{C})$ and $SO(1,3)\uparrow$

In this chapter we introduce some elementary results about the interplay of $SL(2,\mathbb{C})$ and $SO(1,3)\uparrow$. The final goal is to prepare the background to develop the theory of spinors and spinorial representations.

10.1 Elementary properties of $SL(2,\mathbb{C})$ and $SU(2)$

As is well known, $SL(2,\mathbb{C})$ denotes the Lie subgroup of $GL(2,\mathbb{C})$ made of all the 2×2 complex matrices with unital determinant [KNS, Ruhl, Streater-Wightman, Wightman]. $SL(2,\mathbb{C})$ can be viewd as a real Lie group referring to real coordinates. We remind the reader that $U(2)$ is the group of complex 2×2 unitary matrices. $SU(2) := U(2)\cap SL(2,\mathbb{C})$. Evidently, both are (Lie) subgroup of $SL(2,\mathbb{C})$.

10.1.1 Almost all on Pauli matrices

The (real) Lie algebra of $SL(2,\mathbb{C})$, $sl(2,\mathbb{C})$ and that of $SU(2)$, $su(2)$ are 6-dimensional and 3-dimensional; respectively. Performing an analysis similar to the one about Lorentz group presented in the previous chapter, one sees that

$$sl(2,\mathbb{C}) = \{M\in M(2,\mathbb{C}) \mid trM = 0\}\,, \quad su(2) = \{M\in M(2,\mathbb{C}) \mid M = M^\dagger\,,\ trM = 0\}\,. \tag{10.1}$$

Above † denotes the Hermitian conjugate: $M^\dagger := \overline{M}^t$ and the bar indicates the complex conjugation. By direct inspection one verifies that $sl(2,\mathbb{C})$ admits a basis made of the following six matrices

$$-i\sigma_1/2,\ -i\sigma_2/2,\ -i\sigma_3/2,\ \sigma_1/2,\ \ \sigma_2/2,\ \ \sigma_3/2\,, \tag{10.2}$$

and the first three define a basis of $su(2)$ as well. We have introduced the well-known *Pauli matrices*

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\,,\ \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}\,,\ \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\,. \tag{10.3}$$

By direct inspection one sees that these matrices fulfill the following identities

$$tr\sigma_i = 0\,,\ \ \sigma_i = (\sigma_i)^\dagger\,, \qquad \sigma_i\sigma_j = \delta_{ij}I + i\sum_{k=1}^{3}\epsilon_{ijk}\sigma_k\,, \quad \text{for } i,j = 1,2,3, \tag{10.4}$$

where, as usual, ϵ_{ijk} is the completely antisymmetric Ricci indicator with $\epsilon_{123} = 1$.

Remarks 10.1. As a very important consequence of the definition of Pauli matrices, one immediately obtains the following commutation relations of the generators of $sl(2,\mathbb{C})$.

$$\left[-i\frac{\sigma_i}{2}, \frac{\sigma_j}{2}\right] = \sum_{k=1}^{3}\epsilon_{ijk}\frac{\sigma_k}{2}\,, \quad \left[-i\frac{\sigma_i}{2}, -i\frac{\sigma_j}{2}\right] = \sum_{k=1}^{3}\epsilon_{ijk} - i\frac{\sigma_k}{2}\,, \quad \left[\frac{\sigma_i}{2}, \frac{\sigma_j}{2}\right] = -\sum_{k=1}^{3}\epsilon_{ijk}\frac{\sigma_k}{2}\,. \tag{10.5}$$

It is then evident that the unique real vector space isomorphism from $sl(2,\mathbb{C})$ to the lie algebra of $SO(1,3)\uparrow$, $so(1,3)\uparrow$ that identifies S_i with $-i\sigma_i/2$ and $\sigma_i/2$ with K_i, $i = 1,2,3$, preserves the Lie commutator in view of (9.9). In other words $sl(2,\mathbb{C})$ and $so(1,3)\uparrow$ are isomorphic as real Lie algebras. The analog arises by comparison of the Lie algebra of $su(2)$ and that of $so(3)$. As is known [KNS] this implies that there exists a local Lie-group isomorphism form a neighborhood of the identity of $SL(2,\mathbb{C})$ to a similar neighborhood of $SO(1,3)\uparrow$, and from a neighborhood of the identity of $SU(2)$ to a similar neighborhood of $SO(3)$ respectively, whose differential maps reduce to the found Lie-algebra isomorphisms. In the following we shall study those Lie-group isomorphisms proving that, actually, they extend to surjective global Lie-group homomorphisms, which are the starting point for the theory of (relativistic) spinors. ■

For the applications to relativity, it is convenient to define some other "Pauli matrices". First of all, define

$$\sigma_0 := I\,, \tag{10.6}$$

so that σ_μ makes sense if $\mu = 0,1,2,3$. In the following, as usual, $\eta_{\mu\nu}$ denotes the components of the Minkowskian metric tensor in canonical form (i.e. the elements of the matrix $diag(-1,1,1,1,))$, and $\eta^{\mu\nu}$ are the components of the inverse Minkowskian metric tensor. Extending the procedure of raising indices, it is customary to define the **Pauli matrices with raised indices**:

$$\sigma^\mu := \eta^{\mu\nu}\sigma_\nu\,. \tag{10.7}$$

Above $\nu,\mu := 0,1,2,3$ and it is assumed the convention of summation of the repeated indices. Finally, essentially for technical reasons, it is also customary to define the **primed Pauli matrices**

$$\sigma'_0 := I\,,\ \sigma'_i := -\sigma_i\,, \quad i = 1,2,3\,. \tag{10.8}$$

and those with raised indices

$$\sigma'^\mu := \eta^{\mu\nu}\sigma'_\nu\,. \tag{10.9}$$

It is worth noticing that $\{\sigma_\mu\}_{\mu=0,1,2,3}$ is a vector basis of the real space $\mathcal{H}(2,\mathbb{C})$ of Hermitian complex 2×2 matrices and $\{\sigma_\mu, i\sigma_\mu\}_{\mu=0,1,2,3}$ is a vector basis of the algebra of complex 2×2

matrices $M(2,\mathbb{C})$ viewed as complex vector space.

Exercises 10.2.

1. Prove that $\sigma_0, \sigma_1, \sigma_2, \sigma_3$ is a basis for the *real* vector space of the 2×2 complex Hermitian matrices.

2. Prove that $\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}I$ and $\sigma'_i\sigma'_j + \sigma'_j\sigma'_i = 2\delta_{ij}I$, $i = 1,2,3$.

3. Prove that:

$$tr(\sigma'_\mu\sigma_\nu) = -\eta_{\mu\nu}\ , \ \ tr(\sigma'^\mu\sigma^\nu) = -\eta^{\mu\nu}\ , \ \ tr(\sigma'^\mu\sigma_\nu) = -\delta^\mu_\nu\ . \tag{10.10}$$

and

$$tr(\sigma_\mu\sigma'_\nu) = -\eta_{\mu\nu}\ , \ \ tr(\sigma^\mu\sigma'^\nu) = -\eta^{\mu\nu}\ , \ \ tr(\sigma^\mu\sigma'_\nu) = -\delta^\mu_\nu\ . \tag{10.11}$$

(*Hint*. Use the result in exercise 2 and the cyclic property of the trace.)

4. Defining $\epsilon := i\sigma_2$, prove that:

$$\epsilon\sigma^t_\mu\epsilon = \epsilon\overline{\sigma_\mu}\epsilon = -\sigma'_\mu\ . \tag{10.12}$$

5. Prove that $\{\sigma_\mu\}_{\mu=0,1,2,3}$ is a vector basis of $\mathcal{H}(2,\mathbb{C})$ and that $\{\sigma_\mu, i\sigma_\mu\}_{\mu=0,1,2,3}$ is a vector basis of the algebra $M(2,\mathbb{C})$ viewed as complex vector space.

10.1.2 Properties of exponentials of Pauli matrices and consequences for $SL(2,\mathbb{C})$ and $SU(2)$

We have a technical, but very important, result stated in the following proposition.

Proposition 10.3. *The following results are valid.*

(a) If $\sigma := (\sigma_1, \sigma_2, \sigma_3)$ and the $so(3)$ generators T_j being defined as in proposition 9.8,

$$e^{-i\theta\mathbf{m}\cdot\sigma/2}\mathbf{n}\cdot\sigma e^{i\theta\mathbf{m}\cdot\sigma/2} = \left(e^{\theta\mathbf{m}\cdot\mathbf{T}}\mathbf{n}\right)\cdot\sigma\ , \quad \textit{for all } \mathbf{n},\mathbf{m}\in\mathbb{S}^2, \theta\in\mathbb{R}. \tag{10.13}$$

(b) The following pair of decompositions are valid:

$$e^{\chi\mathbf{n}\cdot\sigma/2} = \cosh\frac{\chi}{2}\, I + \sinh\frac{\chi}{2}\,\mathbf{n}\cdot\sigma\ , \quad \textit{for all } \mathbf{n}\in\mathbb{S}^2, \chi\in\mathbb{R}, \tag{10.14}$$

$$e^{-i\theta\mathbf{m}\cdot\sigma/2} = \cos\frac{\theta}{2}\, I - i\sin\frac{\theta}{2}\,\mathbf{m}\cdot\sigma\ , \quad \textit{for all } \mathbf{m}\in\mathbb{S}^2, \theta\in\mathbb{R}, \tag{10.15}$$

(c) A matrix $A\in M(2,\mathbb{C})$ has the form $e^{-i\theta\mathbf{n}\cdot\sigma/2}$, for some $\theta\in\mathbb{R}, \mathbf{n}\in\mathbb{S}^2$, if and only if $A\in SU(2)$. The Lie group $SU(2)$ is homeomorphic to $\mathbb{S}^3$.

(d) A matrix $A\in M(2,\mathbb{C})$ has the form $e^{\chi\mathbf{n}\cdot\sigma/2}$, for some $\chi\in\mathbb{R}, \mathbf{n}\in\mathbb{S}^2$, if and only if A is a Hermitian positive element of $SL(2,\mathbb{C})$. The set of all Hermitian positive element of $SL(2,\mathbb{C})$ equipped with the topology induced by $SL(2,\mathbb{C})$ is homeomorphic to $\mathbb{R}^3$ and it is not a subgroup of $SL(2,\mathbb{C})$.

Proof. (a) and (b) Fix $\mathbf{m} \in \mathbb{S}^2$ and consider the smooth functions

$$f_i(\theta) = e^{-i\theta\mathbf{m}\cdot\sigma/2}\sigma_i e^{i\theta\mathbf{m}\cdot\sigma/2} .$$

By direct inspection, making use of the first set commutation rules in (10.5), one obtains

$$\frac{df_i}{d\theta} = \sum_{k,p=1}^{3} m^k \epsilon_{kip} f_p(\theta) .$$

Reminding that $(T_i)_{rs} = -\epsilon_{irs}$ and considering the other set of smooth functions

$$g_i(\theta) = \left(e^{\theta\mathbf{m}\cdot\mathbf{T}}\mathbf{e}_i\right) \cdot \sigma ,$$

one veryfies that, again,

$$\frac{dg_i}{d\theta} = \sum_{k,p=1}^{3} m^k \epsilon_{kip} g_p(\theta) .$$

Since both the set of smooth functions satisfy the same system of first-order differential equations, in normal form, with the same initial condition $f_i(0) = g_i(0) = \sigma_i$, we conclude that $f_i(\theta) = g_i(\theta)$ for all $\theta \in \mathbb{R}$ by uniqueness. This results yields (10.13) immediately. The established identity implies, by exponentiation

$$e^{-i\theta\mathbf{m}\cdot\sigma/2} e^{s\mathbf{n}\cdot\sigma} e^{i\theta\mathbf{m}\cdot\sigma/2} = \exp\{s\left(e^{\theta\mathbf{m}\cdot\mathbf{T}}\mathbf{n}\right)\cdot\sigma\}$$

for either $s \in \mathbb{R}$ or $s \in i\mathbb{R}$. Using these improved results, and rotating $\mathbf{n}$ into $\mathbf{e}_3$ with a suitable element $e^{-i\theta\mathbf{m}\cdot\sigma/2}$ of $SO(3)$, one realizes immediately that (10.14) and (10.15) are equivalent to

$$\begin{aligned} e^{\chi\sigma_3/2} &= \cosh\frac{\chi}{2}\, I + \sinh\frac{\chi}{2}\,\sigma_3 , \quad \text{for all } \chi \in \mathbb{R}, \\ e^{-i\theta\sigma_3/2} &= \cos\frac{\theta}{2}\, I - i\sin\frac{\theta}{2}\sigma_3 , \quad \text{for all } \theta \in \mathbb{R}. \end{aligned}$$

In turn, these identities can be proved by direct inspection, using the (diagonal) explicit form of σ_3.

(c) If $A := e^{-i\theta\mathbf{n}\cdot\sigma/2}$, then, making use of the well-known identity, $\det(e^F) = e^{trF}$, we have $\det A = e^{tr(-i\theta\mathbf{n}\cdot\sigma/2)} = e^{-i\theta\mathbf{n}\cdot tr(\sigma)/2} = e^0 = 1$, so that $A \in SL(2,\mathbb{C})$. We also have

$$A^\dagger = (e^{-i\theta\mathbf{n}\cdot\sigma/2})^\dagger = e^{(-i\theta\mathbf{n}\cdot\sigma/2)^\dagger} = e^{i\theta\mathbf{n}\cdot\sigma^\dagger/2} = e^{i\theta\mathbf{n}\cdot\sigma^\dagger/2} = A^{-1} .$$

This proves that every matrix of the form $e^{-i\theta\mathbf{n}\cdot\sigma/2}$ belongs to $SU(2)$. Now we go to establish also the converse fact. Consider a matrix $V \in SU(2)$. We know by the spectral theorem that it is decomposable as $V = e^{-i\lambda}P_- + e^{i\lambda}P_+$, where $P_\pm$ are orthonormal projectors onto one-dimensional subspaces and $\lambda \in \mathbb{R}$. Consider the Hermitian operator $H := -\lambda P_- + \lambda P_+$. By construction $e^{iH} = V$. On the other hand, every Hermiten 2×2 matrix H can be written as $H = t^0 I + \sum_{i=1}^3 t^i\sigma_i$ for some reals t^0, t^1, t^2, t^3, that is $H = t^0 I - \theta\mathbf{n}\cdot\sigma/2$ for some $\theta \in \mathbb{R}$ and

$\mathbf{n} \in \mathbb{S}^2$. We have found that $V = e^{t^0} e^{-i\theta\mathbf{n}\cdot\sigma/2}$. The requirement $\det V = 1$ impose $t^0 = 0$, so that we have found that $V = e^{-i\theta\mathbf{n}\cdot\sigma/2}$ if and only if $V \in SU(2)$ as requested. Consider now the generic $V \in SU(2)$. In view of (b), it can be written as the matrix:

$$V(\mathbf{n}, \theta) = \cos\frac{\theta}{2}\, I - i \sin\frac{\theta}{2}\, \mathbf{n}\cdot\sigma\,, \quad \text{for any } \mathbf{n} \in \mathbb{S}^2, \theta \in \mathbb{R}.$$

Making explicit the form of V with the help of the explicit expression of Pauli matrices, it turns out to be is that the assignment of such a $V(\mathbf{n}, \theta)$ is equivalent to the assignment a point on the surface in $\mathbb{R}^4$:

$$X^0 := \cos\frac{\theta}{2}\,,\ X^1 = n^1 \sin\frac{\theta}{2}\,,\ X^2 = n^2 \sin\frac{\theta}{2}\,, X^3 = n^3 \sin\frac{\theta}{2}\,, \quad \text{for any } \mathbf{n} \in \mathbb{S}^2, \theta \in \mathbb{R}.$$

This surface it is nothing but the 3-sphere $\mathbb{S}^3$, $(X)^2+(X^1)^2+(X^2)^2+(X^3)^2 = 1$. This one-to-one correspondence between matrices $V(\mathbf{n}, \theta)$ and the points of 3-sphere $\mathbb{S}^3$ can be proved to be a homeomorphism from $\mathbb{S}^3$ and the $SU(2)$, the idea of the proof being the same as that given for the space of the boost in the proof of (e) in proposition 9.11.
(d) With the same procedure as in the case (c), one finds that every matrix $A := e^{\chi\mathbf{n}\cdot\sigma/2}$, with $\chi \in \mathbb{R}$ belongs to $SL(2, \mathbb{C})$ and fulfills $A = A^\dagger$. Let us prove that it is positive, too. Positivity means that $(x|Ax) \geq 0$ for every $x \in \mathbb{C}^2$, where $(|)$ is the standard Hermitian scalar product in $\mathbb{C}^2$. Notice that, if V is unitary (and thus preserves the scalar product), $(x|Ax) \geq 0$ for every $x \in \mathbb{C}^2$ if and only if $(x|V^\dagger AVx) \geq 0$ for every $x \in \mathbb{C}^2$. Using (a), one can fix the matrix $V = e^{-i\theta\mathbf{m}\cdot\sigma}$, in order that $V^\dagger\mathbf{n}\cdot\sigma V = \sigma_3$. Therefore $V^\dagger e^{\chi\mathbf{n}\cdot\sigma/2} V = e^{\chi\sigma_3/2}$. Since σ_3 is diagonal, $e^{\chi\sigma_3/2}$ is diagonal as well and positivity of $e^{\chi\sigma_3/2}$ can be checked by direct inspection, and it result to be verified trivially. We have established that every matrix of the form $e^{\chi\mathbf{n}\cdot\sigma/2}$ is a positive Hermitian element of $SL(2, \mathbb{C})$. Now we go to prove the converse fact. Consider a Hermitian positive matrix $P \in SL(2, C)$. We know, by the spectral theorem of Hermitian operators that it is decomposable as $P = \lambda_1 P_1 + \lambda_2 P_2$, where $P_{1,2}$ are orthonormal projectors onto one-dimensional subspaces and $\lambda_1, 2 > 0$ from positivity and the requirement $1 = \det V = \lambda_1\lambda_2$. Consider the Hermitian operator $H := \ln\lambda_1 P_1 + \ln\lambda_2 P_2$. By construction $e^H = P$. As before, every Hermitian 2×2 matrix H can be written as $H = t^0 I + \sum_{i=1}^3 t^i\sigma_i$ for some reals t^0, t^1, t^2, t^3, that is $H = t^0 I + \chi\mathbf{n}\cdot\sigma/2$ for some $\chi \in \mathbb{R}$ and $\mathbf{n} \in \mathbb{S}^2$. We have found that $P = e^{t^0} e^{\chi\mathbf{n}\cdot\sigma/2}$. The requirement $\det P = 1$ imposes $t^0 = 0$, so that we have found that $P = e^{\chi\mathbf{n}\cdot\sigma/2}$ if and only if $P \in SL(2, \mathbb{C})$ is Hermitian and positive as requested. Consider now the generic positive Hermitian element $P \in SL(2, \mathbb{C})$. In view of (b), it can be written as the matrix:

$$P(\mathbf{n}, \chi) = \cosh\frac{\chi}{2}\, I + \sinh\frac{\chi}{2}\, \mathbf{n}\cdot\sigma\,, \quad \text{for any } \mathbf{n} \in \mathbb{S}^2, \chi \in \mathbb{R}.$$

Making explicit the form of P with the help of the explicit expression of Pauli matrices, it turns out that the assignment of such a P is equivalent to the assignment a point on the surface in $\mathbb{R}^4$:

$$X^0 := \cosh\frac{\chi}{2}\,,\ X^1 = n^1 \sinh\frac{\chi}{2}\,,\ X^2 = n^2 \sinh\frac{\chi}{2}\,,\ X^3 = n^3 \sinh\frac{\chi}{2}\,, \quad \text{for any } \mathbf{n} \in \mathbb{S}^2, \chi \in \mathbb{R}.$$

This surface it is nothing but the hyperboloid $-(X)^2 + (X^1)^2 + (X^2)^2 + (X^3)^2 = -1$ which is trivially homeomorphic to $\mathbb{R}^3$. This one-to-one correspondence between matrices $P(\mathbf{n}, \chi)$ and the points of the hyperboloid can be proved to be a homeomorphism from $\mathbb{R}^3$ and the set of all Hermitian positive element of $SL(2, \mathbb{C})$ equipped with the topology induced by $SL(2, \mathbb{C})$, the idea of the proof being the same as that given for the space of the boost in the proof of (e) in proposition 9.11. $\Box$

10.1.3 Polar decomposition theorem and topology of $SL(2, \mathbb{C})$

We are in a position to state and prove a decomposition theorem for $SL(2, \mathbb{C})$ which will play an important role in the following section.

Theorem 10.4. *Take $L \in SL(2, \mathbb{C})$, the following holds.*

(a) There is exactly one positive Hermitian matrix $H \in SL(2, \mathbb{C})$ (so that $H = e^{\chi \mathbf{n} \cdot \sigma/2}$ for some $\chi \in \mathbb{R}$ and $\mathbf{n} \in \mathbb{S}^2$) and exactly a unitary matrix $V \in U(2)$ (so that $V = e^{-i\theta \mathbf{m} \cdot \sigma}$ for some $\theta \in \mathbb{R}$ and $\mathbf{m} \in \mathbb{S}^2$) such that

$$L = VH \, .$$

(b) There is exactly one positive Hermitian matrix $H' \in SL(2, \mathbb{C})$ (so that $H' = e^{\chi' \mathbf{n}' \cdot \sigma/2}$ for some $\chi' \in \mathbb{R}$ and $\mathbf{n}' \in \mathbb{S}^2$) and exactly a unitary matrix $V' \in U(2)$ (so that $V' = e^{-i\theta' \mathbf{m}' \cdot \sigma}$ for some $\theta' \in \mathbb{R}$ and $\mathbf{m}' \in \mathbb{S}^2$) such that

$$L = H'V' \, .$$

(c) It holds $V' = V$ and $H' = VHV^t$.

Proof. This is nothing but the specialization of polar decomposition theorem (theorem 7.10) applied to complex 2×2 matrices. The explicit expressions for H, V, H', V' arise from (b) and (c) in proposition 10.3. $\Box$

We want emphasize here a topological consequence of this theorem. Using the fact that $SL(2, \mathbb{C})$ is bijectively identified with the product $SU(2) \times \mathbb{R}^3$ i.e. $\mathbb{S}^3 \times \mathbb{R}^3$, where the $\mathbb{R}^3$ is the topological space of the Hermitian positive elements of $SL(2, \mathbb{C})$ as in (c) of proposition 10.3, it arises (the proof is the same as that in proposition 9.16) that $SL(2, \mathbb{C})$ is homeomorphic to the topological space $\mathbb{S}^3 \times \mathbb{R}^3$. Since the factors are connected, $SL(2, \mathbb{C})$ is connected too. It is also arch connected, it being a differentiable manifold and thus admitting a topological base made of (smooth-)arch-connected open sets. Moreover, the fundamental group of $SL(2, \mathbb{C})$ is, as happens for product manifolds [Sernesi], the product of the fundamental groups of the factors. As $\pi_1 (SL(2, \mathbb{C})) = \pi_1 \left(\mathbb{R}^3\right) = \{1\}$, one finds again $\pi_1 (SL(2, \mathbb{C})) = \{1\}$. We proved the following proposition.

Proposition 10.5. *$SL(2,\mathbb{C})$ is homeomorphic to $\mathbb{S}^3 \times \mathbb{R}^3$, as a consequence it is connected, arch-connected and $\pi_1(SL(2,\mathbb{C})) = \{1\}$, so that it is simply connected.*

Remarks 10.6. We can draw further conclusions from the proposition above, making use of general theorems on Lie groups [KNS]. Since $SL(2,\mathbb{C})$ is simply connected, it is the only simply-connected Lie group (up to Lie-group isomorphisms) which admits $sl(2,\mathbb{C})$ as Lie algebra. As a consequence, it has to coincide with the universal covering of $SO(1,3)\uparrow$ (since $so(1,3)\uparrow$ is isomorphic to $sl(2,\mathbb{C})$, as noticed at the beginning of this section, and since $SO(1,3)\uparrow$ is not simply connected due to proposition 9.16). Thus there must be a Lie-group homomorphism from $SL(2,\mathbb{C})$ onto $SO(1,3)\uparrow$ which is a local Lie-group isomorphism about the units and whose differential computed in the tangent space on the group unit coincides with the Lie-algebra isomorphism which identifies $sl(2,\mathbb{C})$ and $so(1,3)\uparrow$. ■

10.2 The interplay of $SL(2,\mathbb{C})$ and $SO(1,3)\uparrow$

In this subsection we establish the main results enabling one to introduce the notion of *spinor* in relation with vectors. This result just concerns the covering Lie-group homomorphism from $SL(2,\mathbb{C})$ onto $SO(1,3)\uparrow$ already mentioned in remarks 10.6.

10.2.1 Construction of the covering homomorphism $\Pi : SL(2,\mathbb{C}) \to SO(1,3)\uparrow$

First of all we notice that the real vector space $\mathcal{H}(4,\mathbb{C})$ of 2×2 complex Hermitian matrices admits a basis made of the extended Pauli matrices σ_μ with $\mu = 0,1,2,3$. So that, if t^μ denotes the μ-th component of $t \in \mathbb{R}^4$, referred to the canonical basis, there is a linear bijective map

$$\mathbb{R}^4 \ni t \to H(t) := t^\mu \sigma_\mu \in \mathcal{H}(4,\mathbb{C}) \, . \tag{10.16}$$

Above, we have adopted the convention of summation over repeated indices and that convention will be always assumed henceforth in reference to Greek indices. As a second step we notice the following remarkable identity, which arises by direct inspection from the given definitions,

$$\det H(t) = -\eta_{\mu\nu} t^\mu t^\nu \, , \quad \text{for every } t \in \mathbb{R}^4. \tag{10.17}$$

$$tr(\sigma'_\mu \sigma_\nu) = -\eta_{\mu\nu} \, , \;\; tr(\sigma'^\mu \sigma^\nu) = -\eta^{\mu\nu} \, , \;\; tr(\sigma'^\mu \sigma_\nu) = -\delta^\mu_\nu \, . \tag{10.18}$$

As a further step, we observe that, if $H(t) \in \mathcal{H}(4,\mathbb{C})$, then $LH(t)L^\dagger \in \mathcal{H}(4,\mathbb{C})$ because

$$(LH(t)L^\dagger)^\dagger = (L^\dagger)^\dagger H(t)^\dagger L^\dagger = LH(t)L^\dagger \, ,$$

and thus $LH(t)L^\dagger$ has the form $H(t')$ for some $t' \in \mathbb{R}$. However, if $L \in SL(2,\mathbb{C})$, it also holds $L^\dagger \in SL(2,\mathbb{C})$ (because $\det L^\dagger = \overline{\det L^t} = \overline{\det L} = 1$) and thus the determinant of $H(t)$ coincides with that of $H(t') = LH(t)L^\dagger$ by Binet's rule:

$$\det H(t') = \det(LH(t)L^\dagger) = (\det L)(\det H(t)) \det L^\dagger = 1(\det H(t))1 = \det H(t) \, .$$

Now (10.17) implies that $LH(t)L^\dagger = H(t')$ with $\eta_{\mu\nu}t^\mu t^\nu = \eta_{\mu\nu}t'^\mu t'^\nu$. This result has a quite immediate fundamental consequence.

Proposition 10.7. *There is a group homomorphism* $\Pi : SL(2,\mathbb{C}) \ni L \mapsto \Lambda_L \in O(1,3)$ *which is uniquely determined by*

$$LH(t)L^\dagger = H(\Lambda_L t) \ , \quad \textit{for all } t \in \mathbb{R}^4. \tag{10.19}$$

Proof. As noticed above, $LH(t)L^\dagger = H(t')$. Since the correspondence (10.16) is bijective, for a fixed $L \in SL(2,\mathbb{C})$, a map $f_L : \mathbb{R}^4 \ni t \mapsto t'$ is well defined (there is a unique t' for each fixed t). So, we can write $LH(t)L^\dagger = H(f_L(t))$. On the other hand f_L is linear due to the linearity and bijectivity of $t \mapsto H(t)$, in particular. Indeed, one has $LH(at+bs)L^\dagger = aLH(t)L^\dagger + bLH(s)L^\dagger$, i.e $H(f_L(at+bs)) = aH(f_L(t)) + bH(f_L(s)) = H(af_L(t) + bf_L(s))$, so that $f_L(at+bs) = af_L(t)+bf_L(s)$, because $t \mapsto H(t)$ is invertible. It remains to prove that $f_L \in O(1,3)$. We know that, as discussed above, $\eta_{\mu\nu}t^\mu t^\nu = \eta_{\mu\nu}(f_Lt)^\mu(f_Lt)^\nu$ for every $t \in \mathbb{R}^4$, namely

$$\mathbf{g}(t,t) = \mathbf{g}\left(f_Lt, f_Lt\right) \ , \quad \text{for every } t \in \mathbb{R}^4. \tag{10.20}$$

However, as a general fact, if $s, t \in \mathbb{R}^4$, symmetry of the pseudo scalar product $\mathbf{g}$ entails:

$$\mathbf{g}(s,t) = \frac{1}{4}\left(\mathbf{g}(s+t, s+t) - \mathbf{g}(s-t, s-t)\right)$$

so that, (10.20) implies

$$\mathbf{g}(s,t) = \mathbf{g}\left(f_Ls, f_Lt\right) \ , \quad \text{for every } s,t \in \mathbb{R}^4.$$

Let us indicate f_L by Λ_L and prove that $\Pi : L \mapsto \Lambda_L$ is a group homomorphism. Trivially, $L_I = I$ because $IH(t)I^\dagger = H(It)$ for every $t \in \mathbb{R}^4$. Moreover

$$H(\Lambda_{LL'}t) = (LL')H(t)(LL')^\dagger = L(L'H(t)L'^\dagger)L^\dagger = L(H(\Lambda_{L'}t))L^\dagger = H(\Lambda_L\Lambda_{L'}t), \quad \text{for all } t \in \mathbb{R}^4,$$

and thus $\Lambda_{LL'} = \Lambda_L\Lambda_{L'}$. The proof is concluded. $\Box$

10.2.2 Properties of Π

The obtained result can be made stronger with several steps. First of all, we prove that Π it is a Lie-group homomorphism.

Proposition 10.8. *The group homomorphism* $\Pi : SL(2,\mathbb{C}) \ni L \mapsto \Lambda_L \in O(1,3)$ *determined by (10.19) enjoys the following properties.*

(a) It is a Lie-group homomorphism.

(b) the following explicit formula holds:

$$(\Lambda_L)^\nu{}_\mu = -tr\left(\sigma'^\nu L\sigma_\mu L^\dagger\right) \ , \quad \textit{for every } L \in SL(2,\mathbb{C}), \tag{10.21}$$

(c) The kernel of Π *includes* $\{\pm I\}$ *since*

$$\Pi(L) = \Pi(L') \quad \textit{if } L' = -L. \tag{10.22}$$

Proof. (a) and (b) Since the differentiable structures of $SL(2,\mathbb{C})$ and $O(1,3)$ are those induced by $\mathbb{R}^{16}$ and the the operations of taking the trace, taking the Hermitian conjugate and multiplying matrices are trivially differentiable, (10.21) implies that the group homomorphism Π is differentiable and thus is a Lie-group homomorphism. To prove (10.21). Choosing $t = \delta^\alpha_\mu e_\alpha$, where $\{e_\lambda\}_{\lambda=0,1,2,3}$ is the canonical base of $\mathbb{R}^4$, (10.19) produces:

$$L\sigma_\mu L^\dagger = Lt^\alpha\sigma_\alpha L^\dagger = (\Lambda_L)^\gamma{}_\beta\, t^\beta\sigma_\gamma = (\Lambda_L)^\gamma{}_\beta \delta^\beta_\mu \sigma_\gamma = (\Lambda_L)^\gamma{}_\mu\, \sigma_\gamma \ .$$

As a consequence

$$\sigma'^\nu L\sigma_\mu L^\dagger = (\Lambda_L)^\gamma{}_\mu\, \sigma'^\nu\sigma_\gamma \ .$$

To conclude, it is enough applying the identity $tr(\sigma'^\nu\sigma_\gamma) = -\delta^\nu_\gamma$ (see (10.10) in exercises 10.2).
(c) Using the very definition of Π, it is evident that $\Pi(L) = \Pi(-L)$. $\Box$

Now we study the interplay of Π and exponentials $\exp\{a\mathbf{n}\cdot\sigma\}$ as those appearing in the decomposition theorem 10.4.

Proposition 10.9. *The Lie-group homomorphism* $\Pi : SL(2,\mathbb{C}) \ni L \mapsto \Lambda_L \in O(1,3)$ *determined by (10.19) enjoys the following features, using that notation of subsection 9.2.2.*

(a) For every $\theta \in \mathbb{R}$ *and* $\mathbf{n} \in \mathbb{S}^2$*, one has*

$$\Pi\left(e^{-i\theta\mathbf{n}\cdot\sigma/2}\right) = e^{\theta\mathbf{n}\cdot\mathbf{S}} \ , \tag{10.23}$$

so that $\Pi(SU(2)) = SO(3)$*. Moreover, if* $V, V' \in SU(2)$*,* $\Pi(V) = Pi(V')$ *if and only if* $V = \pm V'$*.*

(b) For every $\chi \in \mathbb{R}$ *and* $\mathbf{n} \in \mathbb{S}^2$*, one has*

$$\Pi\left(e^{\chi\mathbf{n}\cdot\sigma/2}\right) = e^{\chi\mathbf{n}\cdot\mathbf{K}} \ , \tag{10.24}$$

so that Π *maps the set of* 2×2 *complex Hermitian positive elements of* $SL(2,\mathbb{C})$ *onto the set of Lorentz boosts. Moreover this map is injective.*

Proof. (a) The identity (10.13) implies immediately that, with obvious notation

$$e^{-i\theta\mathbf{n}\cdot\sigma/2}t^\mu\sigma_\mu\left(e^{-i\theta\mathbf{n}\cdot\sigma/2}\right)^\dagger = e^{-i\theta\mathbf{n}\cdot\sigma/2}(t^0I+\mathbf{t}\cdot\sigma)\sigma_\mu e^{i\theta\mathbf{n}\cdot\sigma/2} = t^0I+\left(e^{\theta\mathbf{n}\cdot\mathbf{T}}\mathbf{t}\right)\cdot\sigma = \left(e^{\theta\mathbf{n}\cdot\mathbf{S}}t\right)^\mu\sigma_\mu\,.$$

This results, in view of the definition of Π, and since t is arbitrary, implies (10.23). Remembering that every element of $SU(2)$ has the form $e^{\theta\mathbf{n}\cdot\mathbf{S}}$ (see proposition 9.9), we have also obtained that $\Pi(SU(2)) = SO(3)$. To conclude, we have to prove that $\Pi(V) = \Pi(V')$ implies $V' = -V$ when $V, V' \in SU(2)$. To this end, we notice that, from the definition of Π, $\Pi(V) = \Pi(V')$ is equivalent to $V_1\sigma_iV_1^\dagger = \sigma_i$ for $i = 1, 2, 3$, where $V_1 := V'V^\dagger$. Since $V_1^\dagger = V_1^{-1}$ because $V_1 \in SU(2)$, it also holds $V_1\sigma_k = \sigma_kV_1$ for $i = 1, 2, 3$. Thus V_1 commutes with all complex combinations of I and the Pauli matrices. But these combinations amount to all of the elements of $M(2,\mathbb{C})$. As a consequence it must be $V_1 = \lambda I$ for some $\lambda \in \mathbb{C}$. The requirements $\det V_1 = 1$ and $V_1^\dagger = V^{-1}$ imply $\lambda = \pm 1$, so that $V = \pm V'$.
(b) The identity to be proved is now

$$e^{\chi\mathbf{n}\cdot\sigma/2}t^\mu\sigma_\mu\left(e^{\chi\mathbf{n}\cdot\sigma/2}\right)^\dagger = \left(e^{\chi\mathbf{n}\cdot\mathbf{K}}t\right)^\mu\sigma_\mu\,,$$

that is, since $\sigma_\mu^\dagger = \sigma_\mu$,

$$e^{\chi\mathbf{n}\cdot\sigma/2}t^\mu\sigma_\mu e^{\chi\mathbf{n}\cdot\sigma/2} = \left(e^{\chi\mathbf{n}\cdot\mathbf{K}}t\right)^\mu\sigma_\mu\,.$$

It can be equivalently re-written:

$$e^{-i\theta\mathbf{m}\cdot\sigma/2}e^{\chi\mathbf{n}\cdot\sigma/2}e^{i\theta\mathbf{m}\cdot\sigma/2}e^{-i\theta\mathbf{m}\cdot\sigma/2}t^\mu\sigma_\mu e^{i\theta\mathbf{m}\cdot\sigma/2}e^{-i\theta\mathbf{m}\cdot\sigma/2}e^{\chi\mathbf{n}\cdot\sigma/2}e^{-i\theta\mathbf{m}\cdot\sigma/2} = \left(e^{\chi\mathbf{n}\cdot\mathbf{K}}t\right)^\mu\sigma_\mu\,.$$

That is, fixing $\mathbf{m}$ and θ suitably, and then making use of (10.13),

$$e^{\chi\sigma_3/2}e^{-i\theta\mathbf{m}\cdot\sigma/2}t^\mu\sigma_\mu e^{i\theta\mathbf{m}\cdot\sigma/2}e^{\chi\sigma_3/2} = \left(e^{\chi\mathbf{n}\cdot\mathbf{K}}t\right)^\mu\sigma_\mu\,.$$

In turn, employing (a) in this proposition, and (b) of Proposition 9.11 this turns out to be equivalent to:

$$e^{\chi\sigma_3/2}s^\mu\sigma_\mu e^{\chi\sigma_3/2} = \left(e^{\chi K_3}s\right)^\mu\sigma_\mu\,, \tag{10.25}$$

where $s = e^{\theta\mathbf{m}\cdot\mathbf{S}}t$ is generic. (10.25) can be proved to hold by direct inspection, expanding the exponentials via (10.14), the last formula in (10.4), and (9.16). Since all the boost of Lorentz group have the form $e^{\chi\mathbf{n}\cdot\sigma/2}$, it remains to prove that Π restricted to the Hermitian positive matrices of $SL(2,\mathbb{C})$ is injective. To this end suppose that $\Pi(e^{\chi\mathbf{n}\cdot\sigma/2}) = \Pi(e^{\chi'\mathbf{n}'\cdot\sigma/2})$. This is equivalent to, for every $t \in \mathbb{R}^4$:

$$e^{\chi\mathbf{n}\cdot\sigma/2}t^\mu\sigma_\mu e^{\chi\mathbf{n}\cdot\sigma/2} = e^{\chi'\mathbf{n}'\cdot\sigma/2}t^\mu\sigma_\mu e^{\chi'\mathbf{n}'\cdot\sigma/2}\,.$$

This is equivalent to write, for every $\mu = 0, 1, 2, 3$:

$$e^{\chi\mathbf{n}\cdot\sigma/2}\sigma_\mu e^{\chi\mathbf{n}\cdot\sigma/2} = e^{\chi'\mathbf{n}'\cdot\sigma/2}\sigma_\mu e^{\chi'\mathbf{n}'\cdot\sigma/2}\,.$$

Taking $\mu = 0$, and using $\sigma_0 = I$, these identities produce:

$$e^{\chi \mathbf{n}\cdot\sigma} = e^{\chi' \mathbf{n}'\cdot\sigma} .$$

By direct inspection, making use of (10.14), one finds that this is possible only if

$$\cosh\frac{\chi}{2}\,\sigma_0 + \sinh\frac{\chi}{2}\sum_{j=1}^{3} n^j\sigma_j = \cosh\frac{\chi'}{2}\,\sigma_0 + \sinh\frac{\chi'}{2}\sum_{j=1}^{3} n'^j\sigma_j$$

and thus $\chi n^j = \chi' n'^j$, because the σ_μ are a basis of the space of Hermitian matrices and using well-known properties of hyperbolic functions. As a consequence, it finally holds: $\chi\mathbf{n}\cdot\sigma = \chi'\mathbf{n}'\cdot\sigma$, We have found that $\Pi(e^{\chi\mathbf{n}\cdot\sigma/2}) = \Pi(e^{\chi'\mathbf{n}'\cdot\sigma/2})$ implies $e^{\chi\mathbf{n}\cdot\sigma/2} = e^{\chi'\mathbf{n}'\cdot\sigma/2}$ as wanted. $\square$

We are in a position to state and prove the conclusive theorem.

Theorem 10.10. *The Lie-group homomorphism $\Pi : SL(2,\mathbb{C}) \ni L \mapsto \Lambda_L \in O(1,3)$ determined by (10.19) satisfies the following facts hold.*

(a) $\Pi(SL(2,\mathbb{C})) = SO(1,3){\uparrow}$.

(b) $\Pi(SU(2)) = SO(3)$.

(c) The kernel of Π *is* $\{\pm I\}$, *that is* $\Pi(L') = \Pi(L")$ *if and only if* $L' = \pm L"$.

(d) Π *defines a local Lie-group isomorphism (about the unit) between* $SL(2,\mathbb{C})$ *and* $SO(1,3){\uparrow}$, *so that* $d\Pi|_I : sl(2,\mathbb{C}) \to so(1,3){\uparrow}$ *is a Lie-algebra isomorphism.*

(e) $\Pi{\restriction}_{SU(2)}$ *defines a local Lie-group isomorphism (about the unit) between* $SU(2)$ *and* $SO(3)$, *so that* $d\Pi{\restriction}_{SU(2)}|_I : su(2) \to so(3)$ *is a Lie-algebra isomorphism.*

(f) If $L = HV$ *is the polar decomposition of* $L \in SL(2,\mathbb{C})$, *with* V *unitary and* H *positive Hermitian,* $\Pi(H)\Pi(V)$ *is the polar decomposition of* $\Pi(L) \in SO(1,3){\uparrow}$, *where* $\Pi(V)$ *is a spatial rotation and* $\Pi(H)$ *a boost.*

Proof. We start from (f). The proof of this statement is almost evident: if $L \in SL(2,\mathbb{C})$ the matrices of its polar decomposition in $SL(2,\mathbb{C})$ have the form $H = e^{\chi\mathbf{n}\cdot\sigma/2}$ and $V = e^{-i\theta\mathbf{m}\cdot\sigma/2}$ due to theorem 10.4. Thus, the action of Π produces, in view of proposition 10.9, $\Pi(V) = e^{\theta\mathbf{m}\cdot\mathbf{S}}$ and $\Pi(H) = e^{\chi\mathbf{n}\cdot\mathbf{K}}$, so that $\Pi(L) = \Pi(H)\Pi(V) = e^{\chi\mathbf{n}\cdot\mathbf{K}}e^{\theta\mathbf{m}\cdot\mathbf{S}}$. The latter is, trivially, a polar decomposition in $SO(1,3){\uparrow}$ of the product of these factors and thus the *unique* one (theorems 9.14 and 9.15). But the product of these factors is $\Pi(L)$ by construction. So $\Pi(H)\Pi(V)$ is the polar decomposition of $\Pi(L)$ in $SO(1,3){\uparrow}$ as wanted.
(a) Let $\Lambda \in SO(1,3){\uparrow}$, so that $\Lambda = e^{\theta\mathbf{m}\cdot\mathbf{S}}e^{\chi\mathbf{n}\cdot\mathbf{K}}$ via polar decomposition. Due to (f), we have immediately: $\Pi(e^{-i\theta\mathbf{m}\cdot\sigma/2}e^{\chi\mathbf{n}\cdot\sigma/2}) = \Lambda$, and thus $\Pi(SL(2,\mathbb{C})) \supset SO(1,3){\uparrow}$. On the other hand, if $L \in SL(2,\mathbb{C})$, as established in the proof of (f), $\Pi(L) = \Pi(H)\Pi(V) = e^{\chi\mathbf{n}\cdot\mathbf{K}}e^{\theta\mathbf{m}\cdot\mathbf{S}}$

so that $\Pi(L) \in SO(1,3)\uparrow$ because is the product of two elements of that group. Therefore $\Pi(SL(2,\mathbb{C})) \subset SO(1,3)\uparrow$ and thus $\Pi(SL(2,\mathbb{C})) = SO(1,3)\uparrow$. (b) It follows from (a) in proposition 10.9 and (a) and (b) of proposition 9.9.
(c) Suppose that $\Pi(L') = \Pi(L")$. This is equivalent to say that $\Pi(L) = I$, where $L = L'L"^{-1}$. By polar decomposition $L = HV$ and, for (f), $\Pi(H)\Pi(V)$ must be the polar decomposition of $I \in SO(1,3)\uparrow$. A polar decomposition of I is trivially obtained as the product of $I \in SO(3)$ and the trivial boost I. By uniqueness, this is the only polar decomposition of I. (f) entails that $\Pi(H) = I$ so and $\Pi(V) = I$. In view of proposition 10.9, $H = I$ and $V = \pm I$, so that $L = \pm I$ and $L' = \pm L"$.
(d) and (e). As is well-known from the general theory of Lie groups, given a basis $e_1, \ldots, e_n$ in the tangent space at the unit element (i.e the Lie algebra of the group), the set of parameters $(t_1, \ldots, t_n)$ of the one-parameter subgroups $t_k \mapsto \exp\{t_k e_k\}$ generated by the elements of the basis, for a sufficiently small range of the parameters about 0, i.e. $|t_k| < \delta$, defines a coordinate patch, compatible with the differentiable structure, in a neighborhood of the unit element 1 and centered on that point, i.e. 1 corresponds to $(0, \ldots, 0)$. Fix the basis $-i\sigma_1/2$, $-i\sigma_2/2$, $-i\sigma_3/2$, $\sigma_1/2$, $\sigma_2/2$, $\sigma_3/2$ in $sl(2,\mathbb{C})$, associated with the coordinates system $(s^1, s^2, s^3, t^1, t^2, t^3)$ about $I \in SL(2,\mathbb{C})$, and the basis $S_1, S_2, S_3, K_1, K_2, K_3$ in $so(1,3)\uparrow$, associated with the coordinates system $(x^1, x^2, x^3, y^1, y^2, y^3)$ about $I \in SO(1,3)\uparrow$. By proposition 10.9, $\Pi(e^{-is_k\sigma_k/2}) = e^{s_k S_k}$ and $\Pi(e^{t_j\sigma_j/2}) = e^{t_j K_j}$, so that, in the said coordinates, the action of Π, is nothing but:

$$\Pi(s^1, s^2, s^3, t^1, t^2, t^3) = (s^1, s^2, s^3, t^1, t^2, t^3)\,,$$

that is, the identity map. In other words, Π is a diffeomorphism, and thus a Lie-group isomorphism in the constructed coordinate patches about the units of the two groups.
By construction the differential $d\Pi|_I sl(2,\mathbb{C}) \to so(1,3)\uparrow$ maps each element of the first basis $-i\sigma_1/2$, $-i\sigma_2/2$, $-i\sigma_3/2$, $\sigma_1/2$, $\sigma_2/2$, $\sigma_3/2$ into the corresponding element of the other basis $S_1, S_2, S_3, K_1, K_2, K_3$. As noticed early in the comment 10.1, this map is also an isomorphism of Lie algebras since it preserves the commutations rules. The case concerning $SU(2)$ and $SO(3)$ has the same proof, restricting to the bases $-i\sigma_1/2$, $-i\sigma_2/2$, $-i\sigma_3/2$ and S_1, S_2, S_3. $\square$

Chapter 11

Introduction to Spinors

Spinors are important in relativity – because, in a sense, they generalize the notion of four-vector – in quantum mechanics and in quantum field theory. They are the mathematical tool used to describe the spin (or the helicity) of particles and to formulate the quantum relativistic equation of the electron, the so called Dirac equation, but also of other semi-integer particles as neutrinos and, more generally, fermions. Nowadays, their relevance includes several other fields of pure mathematics, especially in relations with the Clifford algebras, but also noncommutative geometry. However, our approach will be very elementary.

11.1 The space of Weyl spinors

In this section we use again the definition of conjugate space $\overline{W}$ and conjugate dual space $\overline{W^*}$ (which is naturally isomorphic to $\overline{W}^*$) as defined in definition 2.1, making use of their properties as stated in theorem 2.12. In general we make use of various definitions, notions and results contained in chapters 2 and 3. However we will confine ourselves to the case of a two-dimensional vector space W on the complex field $\mathbb{C}$. Another ingredient will be a preferred non degenerate anti symmetric tensor $\epsilon \in W^* \otimes W^*$ called the *metric spinor*. Non degenerate means as usual that, viewing ϵ as a linear map $\epsilon : W \to W^*$, it turns out to be bijective. The metric spinor will be used to raise and lower indices, similarly to the metric tensor.

Definition 11.1. A (Weyl) **spinor space** is a two-dimensional vector space W on the complex field $\mathbb{C}$, equipped with a preferred non degenerate anti symmetric tensor $\epsilon \in W^* \otimes W^*$, called the **metric spinor**. The elements of W are called **spinors** (or, equivalently, **Weyl spinors**), those of $\overline{W}$ are called **conjugate spinors**, those of W^* are called **dual spinors**, and those of $\overline{W^*}$ are called **dual conjugate spinors**. The elements belonging to a tensor products of $W, W^*, \overline{W}, \overline{W^*}$ are generically called **spinorial tensors**. ■

Notation 11.2.

(a) Referring to a basis in W and the associated in the spaces W^*, $\overline{W}$ and $\overline{W^*}$, the components of the spinors, i.e. vectors in W, are denoted by ξ^A (where χ may be replaced with another Greek letter). The components of dual spinors, that is vectors of W^*, are denoted by ξ_A. The components of conjugate spinors, that is vectors of $\overline{W}$, are denoted by $\xi^{A'}$. The components of dual conjugate spinors, that is vectors of $\overline{W^*}$, are denoted either by $\xi_{A'}$. Notation for spinorial tensor is similar.

(b) We make sometimes use of the *abstract index notation*, so, for instance $\Xi^A{}_{B'}$ denotes a spinorial tensor in $W \otimes \overline{W^*}$.

(c) Sometimes will be more convenient to use an intrinsic notation, in that case a spinor or a spinorial tensor will be indicated with a Greek letter without indices, e.g. $\xi \in W$ or $\Xi \in W \otimes W^*$.

(d) If Ξ is a spinorial tensor, and $\mathcal{B} := \{e_A\}_{A=1,2} \subset W$ is a basis, ${}_{\mathcal{B}}\Xi$ denotes the *matrix* whose elements are the components of Ξ with respect to $\mathcal{B}$, or with respect the relevant basis, canonically associated with $\mathcal{B}$ in the tensor space containing Ξ. For instance, if $\Xi \in W \otimes \overline{W}$, so that $\{e_A \otimes \overline{e}_{B'}\}_{A,B'=1,2} \subset W \otimes \overline{W}$ is the basis in $W \otimes \overline{W}$ canonically associated with $\mathcal{B}$, ${}_{\mathcal{B}}\Xi$ is the matrix of elements $\Xi^{AB'}$, where $\Xi = \Xi^{AB'} e_A \otimes \overline{e}_{B'}$. ∎

11.1.1 The metric spinor to lower and raise indices

Let us come to the use of the metric spinor ϵ_{AB}. Exactly as in the case of the metric tensor (see section 5.2), since ϵ is non degenerate, the components of

$$\epsilon = \epsilon_{AB} e^{*A} \otimes e^{*B} \,, \quad \epsilon_{AB} = \epsilon(e_A, e_B) = -\epsilon_{BA} \,,$$

gives rise to a nonsingular antisymmetric matrix referring to a fixed basis of W, $\{e_A\}_{A=1,2}$ and the associated canonical bases in $W^* \otimes W^*$. The matrix whose components are ϵ^{AB}, and satisfy

$$\epsilon_{AB}\, \epsilon^{BC} = -\delta_A^C \,, \tag{11.1}$$

defines a second spinorial tensor, the **inverse metric spinor**:

$$\epsilon^{AB} e_A \otimes e_B = -\epsilon^{BA} e_A \otimes e_B \,.$$

The definition turns out to be independent from the used basis, exactly as for the metric tensor. Notice the sign $-$ in (11.1). Following the same procedure as that for the metric tensor, one sees that the metric spinor and the inverse metric spinor define a natural isomorphism ad its inverse respectively, from W to W^*. In components, or referring to the abstract index notation, the isomorphism corresponds to the procedure of lowering indices, and its inverse correspond to the procedure of raising indices. These are defined, respectively, as

$$W \ni \xi^A \mapsto \xi_A := \xi^B \epsilon_{BA} \in W^* \quad \text{and} \quad W^* \ni \eta_A \mapsto \eta^A := \epsilon^{AB} \eta_B \in W \,. \tag{11.2}$$

Notice that these two procedures work by summing over *different* indices. *The order cannot be interchanged here*, because ϵ_{AB} and ϵ^{AB} are antisymmetric, differently from the corresponding metric tensors. With the given choices, it results

$$\xi^A = \epsilon^{AC}(\xi^B \epsilon_{BC}) .$$

The procedures extend to spinorial tensors constructed with several factors V and V^*, exactly as for the metric tensor. In particular one finds, in view of (11.1) again:

$$\epsilon_{AB} = \epsilon^{CD}\epsilon_{CA}\epsilon_{DB} \quad \text{and} \quad \epsilon^{AB} = \epsilon_{CD}\epsilon^{CA}\epsilon^{DB} .$$

11.1.2 The metric spinor in the conjugate spaces of spinors

The space $\overline{W}$ is naturally anti-isomorphic to W, and the same happens for W^* and $\overline{W^*}$, as established in theorem 2.12. As a consequence a preferred metric spinor is defined on $\overline{W}$, induced by ϵ. If $F : V \to \overline{V}$ is the natural anti isomorphism described in theorem 2.12,

$$\overline{\epsilon}(\xi', \eta') := \overline{\epsilon(F^{-1}(\xi'), F^{-1}(\eta'))} , \quad \text{for all } \xi', \eta' \in \overline{W}, \tag{11.3}$$

defines an anti symmetric bilinear map from $\overline{W} \times \overline{W}$ to $\mathbb{C}$. The corresponding tensor of $\overline{W^*} \otimes \overline{W^*}$, is indicated by $\epsilon_{A'B'}$ or, indifferently, $\overline{\epsilon}_{A'B'}$ in the abstract index notation. Fixing a basis $\{e_A\}_{A=1,2}$ and working in the canonically associated basis $\{\overline{e^{*A'}} \otimes \overline{e^{*B'}}\}_{A',B'=1,2}$ in $\overline{W^*} \otimes \overline{W^*}$, in view of the last statement of theorem 2.12, one finds

$$\overline{\epsilon}(\overline{e}_A, \overline{e}_B) = \epsilon_{AB} ,$$

that is, the components of the spinorial tensor $\overline{\epsilon} \in \overline{W^*} \otimes \overline{W^*}$ are the same as those of the metric spinor. As a consequence $\overline{\epsilon}$ is non degenerate and thus it defines a metric spinor on $\overline{W}$. There is, as a consequence, an inverse metric spinor $\overline{\epsilon}^{A'B'}$ also for $\overline{\epsilon}_{A'B'}$, and the procedure of raising and lowering indices can be performed in the conjugated spaces, too.
Finally, one can consider spinorial tensors defined in tensor spaces with four types of factors: V, V^*, $\overline{V}$ and $\overline{V^*}$, and the procedure of raising and lowering indices can be performed in those spaces employing the relevant metric spinors.

Exercises 11.3.
Prove that the inverse metric spinor $\overline{\epsilon}^{A'B'}$ defined as above, may be obtained equivalently from ϵ^{AB} and the natural anti isomorphism $G : V^* \to \overline{V^*}$ (theorem 2.12), following the analog of the procedure to define $\overline{\epsilon}_{A'B'}$ from ϵ_{AB} and the natural anti isomorphism $F : V \to \overline{V}$, as done in (11.3).

11.1.3 Orthonormal bases and the role of $SL(2, \mathbb{C})$

Consider the spinor space W with metric tensor ϵ. The components of the non degenerate anti symmetric metric tensor ϵ, for a fixed basis $\{e_A\}_{A=1,2} \subset W$, are

$$\epsilon_{11} = \epsilon_{22} = 0 , \quad \text{and} \quad \epsilon_{12} = -\epsilon_{21} = \epsilon(e_1, e_2) .$$

Notice that $\epsilon(e_1, e_2) \neq 0$, otherwise ϵ would be degenerate. Therefore, we can rescale e_1 and/or e_2, in order to achieve

$$\epsilon_{11} = \epsilon_{22} = 0 \,, \quad \text{and} \quad \epsilon_{12} = -\epsilon_{21} = 1 \,.$$

As we shall see later, these bases are important, especially in relation with standard tensors in Minkowski spacetime, so we state a formal definition.

Definition 11.4. Referring to a spinor space W with metric spinor ϵ, a basis $\{e_A\}_{A=1,2} \subset W$ is said (ϵ-)**orthonormal** if the components of ϵ, referred to the associated canonical basis $\{e^{*A} \otimes e^{*B}\}_{A,B=1,2} \subset W$, read

$$\epsilon_{11} = \epsilon_{22} = 0 \,, \quad \text{and} \quad \epsilon_{12} = -\epsilon_{21} = 1 \,. \tag{11.4}$$

■

If $\{e_A\}_{A=1,2} \subset W$ is ϵ-orthonormal, the components of ϵ^{AB} with respect to the canonically associated basis in $V^* \otimes V^*$ are again:

$$\epsilon^{11} = \epsilon^{22} = 0 \,, \quad \text{and} \quad \epsilon^{12} = -\epsilon^{21} = 1 \,, \tag{11.5}$$

This happens thanks to the sign $-$ in the right-hand side of (11.1). Conversely, if, referring to some basis $\{e_A\}_{A=1,2} \subset W$ and to the canonically associated ones, one finds that (11.5) is satisfied, he/she is sure that $\{e_A\}_{A=1,2} \subset W$ is ϵ-orthonormal, because it arises from (11.1) that (11.5) implies (11.4). We conclude that (11.4) and (11.5) are equivalent.
This result has an important implication concerning the role of $SL(2,\mathbb{C})$ in spinor theory. It plays the same role as the Lorentz group plays in vector theory, in relation to pseudo orthonormal frames. If $\{e_A\}_{A=1,2} \subset W$ is ϵ-orthonormal, so that (11.5) is true, consider another basis, $\{\tilde{e}_A\}_{A=1,2} \subset W$, with $e_A = L^B{}_A \tilde{e}_B$ for some matrix $L \in GL(2,\mathbb{C})$, whose components are the coefficients $L^B{}_A$. As soon as (11.4) and (11.5) are equivalent, $\{f_A\}_{A=1,2}$ is orthonormal if and only if the coefficients

$$\widetilde{\epsilon}^{AB} = L^A{}_C L^B{}_D \epsilon^{CD} \tag{11.6}$$

satisfy the four requirements

$$\widetilde{\epsilon}^{11} = 0 \,, \quad \widetilde{\epsilon}^{22} = 0 \,, \quad \widetilde{\epsilon}^{12} = -\widetilde{\epsilon}^{21} \,, \quad \widetilde{\epsilon}^{12} = 1 \,.$$

Taking (11.5) into account in (11.6), the only nontrivial condition among the requirements written above is the last one, which is equivalent to say that

$$1 = L^1{}_1 L^2{}_2 - L^1{}_2 L^2{}_1 \,.$$

We recognize the determinant of L in the right-hand side. In other words, if $\{e_A\}_{A=1,2} \subset W$ is ϵ-orthonormal, another basis $\{\tilde{e}_A\}_{A=1,2} \subset W$, with $e_A = L^B{}_A \tilde{e}_B$ for some matrix $L \in GL(2,\mathbb{C})$ (whose components are the coefficients $L^B{}_A$), is ϵ-orthonormal if and only if

$$\det L = 1 \,.$$

Let us summarize the obtained results within a proposition that includes some further immediate results.

Proposition 11.5. *Consider the spinor space W with metric tensor ϵ. Let $\{e_A\}_{A=1,2} \subset W$ be a basis. The following facts hold.*

(a) Referring to the bases canonically associated with $\{e_A\}_{A=1,2} \subset W$, the following conditions are equivalent:

(i) $\{e_A\}_{A=1,2}$ is ϵ-orthonormal;

(ii) $\epsilon^{11} = \epsilon^{22} = 0$, $\epsilon^{12} = -\epsilon^{21} = 1$;

(iii) $\overline{\epsilon}_{1'1'} = \overline{\epsilon}_{2'2'} = 0$, $\overline{\epsilon}_{1'2'} = -\overline{\epsilon}_{2'1'} = 1$;

(iv) $\overline{\epsilon}^{1'1'} = \overline{\epsilon}^{2'2'} = 0$, $\overline{\epsilon}^{1'2'} = -\overline{\epsilon}^{2'1'} = 1$.

(b) Assume that $\{e_A\}_{A=1,2}$ is ϵ-orthonormal and let $\{\tilde{e}_A\}_{A=1,2} \subset W$ be another basis, with

$$e_A = L^B{}_A \tilde{e}_B\,, \quad \text{for some } L \in GL(2,\mathbb{C}) \text{ with components } L^B{}_A.$$

$\{\tilde{e}_A\}_{A=1,2}$ is ϵ-orthonormal if and only if $L \in SL(2,\mathbb{C})$.

11.2 Four-Vectors constructed with spinors in Special Relativity

In this subsection we show how it is possible to build up Minkowskian four-vectors starting from spinors. This is the starting point to construct, in theoretical physics, the differential equations describing the motion of relativistic particles with spin.
First of all we focus on the Hermitian spinorial tensors in the space $W \otimes \overline{W}$. To this end fix a basis $\mathcal{B} := \{e_A\}_{A=1,2} \subset W$ and consider the basis $\{e_A \otimes \overline{e}_{B'}\}_{A,B'=1,2}$, canonically associated with the given one, in the space $W \otimes \overline{W}$. An element $\Xi^\dagger \in W \otimes \overline{W}$ is the **Hermitian conjugate** of a given $\Xi \in W \otimes \overline{W}$ **with respect to** $\mathcal{B}$, if the matrix of the components of $\Xi^\dagger$, referred to the base in $W \otimes \overline{W}$ canonically associated with $\mathcal{B}$, is the adjoint (i.e. Hermitian conjugate) of that of Ξ. In other words, if $\Xi = \Xi^{AB'} e_A \otimes \overline{e}_{B'}$ and $\Xi^\dagger = \Xi^{\dagger AB'} e_A \otimes \overline{e}_{B'}$, it has to be

$$\Xi^{\dagger AB'} := \overline{\Xi^{BA'}}\,.$$

Equivalently

$$_{\mathcal{B}}\Xi^\dagger := \overline{_{\mathcal{B}}\Xi^t}\,.$$

Proposition 11.6. *Given two basis $\mathcal{B} := \{e_A\}_{A=1,2} \subset W$ and $\widetilde{\mathcal{B}} := \{\widetilde{e}_A\}_{A=1,2} \subset W$, $\Xi^\dagger \in W \otimes \overline{W}$ is the hermitean conjugate of $\Xi \in W \otimes \overline{W}$ with respect to $\mathcal{B}$ if and only if it is the hermitean conjugate of Ξ with respect to $\widetilde{\mathcal{B}}$. Thus, the notion of Hermitian conjugate is* intrinsic.

Proof. Assume that $\Xi = \Xi^{AB'} e_A \otimes \overline{e}_{B'} = \widetilde{\Xi}^{AB'} \widetilde{e}_A \otimes \overline{\widetilde{e}}_{B'}$. If $e_B = L^A{}_B \widetilde{e}_A$ and L denotes the matrix whose elements are the coefficients $L^B{}_A$ one has

$$\widetilde{\Xi}^{AB'} = L^A{}_C \overline{L^{B'}{}_{C'}} \Xi^{CD'}\,.$$

Using a matricial notation, the found identity reads:

$$_{\widetilde{\mathcal{B}}}\Xi = L_{\mathcal{B}}\Xi L^{\dagger} .$$

Taking the Hermitian conjugate we achieve:

$$_{\widetilde{\mathcal{B}}}\Xi^{\dagger} = (L_{\mathcal{B}}\Xi L^{\dagger})^{\dagger} = (L_{\mathcal{B}}\Xi^{\dagger} L^{\dagger}) .$$

We have so found that

$$\widetilde{\Xi}^{\dagger AB'} = L^{A}{}_{C}\overline{L^{B'}{}_{C'}}\Xi^{\dagger CD'} .$$

That is just what we wanted to achieve. □

Definition 11.7. An element $\Xi \in W \otimes \overline{W}$ is said to be **real** if $\Xi = \Xi^{\dagger}$. $(W \otimes \overline{W})_{\mathbb{R}}$ denotes the space of real elements of $W \otimes \overline{W}$. ■

The space $(W \otimes \overline{W})_{\mathbb{R}}$ is a real four dimensional vector space. Indeed, we known (see the previous chapter) that the matrices σ^{μ}, $\mu = 0, 1, 2, 3$ defined as in (10.7), form a basis of the real space of complex 2×2 Hermitian matrix $\mathcal{H}(2, \mathbb{C})$. As a consequence, if $\mathcal{B} = \{e_A\}_{A=1,2}$ is an ϵ-orthonormal basis of W, $\Xi \in (W \otimes \overline{W})_{\mathbb{R}}$ if and only if there are four reals, ${}_{\mathcal{B}}t^0, {}_{\mathcal{B}}t^1, {}_{\mathcal{B}}t^2, {}_{\mathcal{B}}t^3$, bijectively defined by the matrix ${}_{\mathcal{B}}\Xi$ of the components of Ξ referred to the basis $\{e_A \otimes \overline{e}_{B'}\}_{A,B'=1,2}$, such that

$$_{\mathcal{B}}\Xi = {}_{\mathcal{B}}t^{\mu}\sigma_{\mu} . \tag{11.7}$$

The inverse relation can be obtained making use of the matrices ${\sigma'}^{\mu}$ (10.9) and the elementary result, presented in 3 in exercises 10.2:

$$_{\mathcal{B}}t^{\mu} = -tr({}_{\mathcal{B}}\Xi\, {\sigma'}^{\mu}) . \tag{11.8}$$

To go on, the idea is to intepret the four real numbers ${}_{\mathcal{B}}t^{\mu}$ as the components of a four-vector in the Minkowski spacetime $\mathbb{M}^4$, with respect to some $\mathbf{g}$-orthonormal frame $\{f_{\mu}\}_{\mu=0,1,2,3}$ of the space of the translations T^4 ($\mathbf{g}$ is the metric with segnature $(-1, +1, +1, +1)$).
The remarkable fact is the following. The found assignment of a vector t^{μ} in correspondence with a real spinorial tensor $\Xi \in W \otimes \overline{W}$ could seem to depend on the fixed basis $\mathcal{B} \subset W$. Actually this is not the case, in view of the covering homomorphism $\Pi : SL(2, \mathbb{C}) \to SO(1, 3)\uparrow$ as discussed in the previous chapter. The following theorem clarifies the relationship between real spinorial tensors of $(W \otimes \overline{W})_{\mathbb{R}}$ and four-vectors.

Theorem 11.8. *Consider a spinor space W, with metric spinor ϵ, and Minkowski spacetime $\mathbb{M}^4$ with metric $\mathbf{g}$ with signature $(-1, +1, +1, +1)$ and, as usual, define $\eta := diag(-1, 1, 1, 1)$. Finally, let $\Pi : SL(2, \mathbb{C}) \ni L \mapsto \Lambda_L \in SO(1, 3)\uparrow$ be the covering homomorphism discussed in theorem 10.10.*
Fix an ϵ-orthonormal basis $\mathcal{B} := \{e_A\}_{A=1,2}$ in W and a $\mathbf{g}$-orthonormal basis $\mathcal{P} := \{f_{\mu}\}_{\mu=0,1,2,3}$ in the space of four-vectors T^4 of $\mathbb{M}^4$. The following holds.

(a) There is a real-vector-space isomorphism

$$h_{\mathcal{B},\mathcal{P}} : (W \otimes \overline{W})_{\mathbb{R}} \ni \Xi \to t_\Xi \in T^4$$

between the space of real spinorial tensors of $W \otimes \overline{W}$ and T^4 which is defined by (11.8), with inverse given by (11.7), where ${}_{\mathcal{B}}t^\mu$ are the components of t_Ξ referred to the base $\{f_\mu\}_{\mu=0,1,2,3}$.

(b) The definition of $h_{\mathcal{B},\mathcal{P}}$ is independent from the fixed bases, modulo the action of Π. In other words, if $\widetilde{\mathcal{B}} = \{\widetilde{e}_A\}_{A=1,2}$ is another ϵ-orthonormal basis of W so that $e_B = L^A{}_B \widetilde{e}_A$, where the coefficients $L^A{}_B$ defines $L \in SL(2,\mathbb{C})$, and $\widetilde{\mathcal{P}} := \{\widetilde{f}_\nu\}_{\nu=0,1,2,3} \subset T^4$ is the $\mathbf{g}$-orthonormal basis with $f_\nu = (\Lambda_L)^\mu{}_\nu \widetilde{f}_\mu$, then

$$h_{\mathcal{B},\mathcal{P}} = h_{\widetilde{\mathcal{B}},\widetilde{\mathcal{P}}} .$$

(c) With the given notations, for every pair $\Xi, \Sigma \in (W \otimes \overline{W})_{\mathbb{R}}$, it holds

$$\Xi^{AB'}\Sigma^{CD'}\epsilon_{AC}\epsilon_{B'D'} = -\eta_{\mu\nu}\, t^\mu_\Xi\, t^\nu_\Sigma . \tag{11.9}$$

So that, in particular, the spinorial tensor,

$$\Gamma := -\epsilon \otimes \overline{\epsilon} \;\; \in W^* \otimes W^* \otimes \overline{W^*} \otimes \overline{W^*} , \tag{11.10}$$

defines a Lorentzian metric in $(W \otimes \overline{W})_{\mathbb{R}}$ with signature $(-1,+1,+1,+1)$.

Proof. (a) It has been proved immediately before the statement of the theorem.
(b) Adopting notation as in (11.8) and (11.7), the thesis is equivalent to

$$L_{\mathcal{B}}\, \Xi\, L^\dagger = (\Lambda_L)^\mu{}_\nu\, {}_{\mathcal{B}}t^\nu .$$

This is nothing but the result proved in proposition 10.7.
(c) With the given definitions and making use of 4 in exercises 10.2, we have:

$$\Xi^{AB'}\Sigma^{CD'}\epsilon_{AC}\epsilon_{B'D'} = t^\mu_\Xi t^\nu_\Sigma \sigma^{AB'}_\mu \sigma^{CD'}_\nu \epsilon_{AC}\epsilon_{B'D'} = t^\mu_\Xi t^\nu_\Sigma tr\left(\epsilon^t \sigma^t_\mu \epsilon \sigma_n u\right) = -t^\mu_\Xi t^\nu_\Sigma tr\left(\epsilon \sigma^t_\mu \epsilon \sigma_n u\right)$$

$$= t^\mu_\Xi t^\nu_\Sigma tr(\sigma'_\mu \sigma_n u) = -\eta_{\mu\nu} t^\mu_\Xi t^\nu_\Sigma ,$$

where, in the last passage we employed 3 in exercises 10.2. $\square$

Bibliography

[Lang] S. Lang, *Linear Algebra* (Undergraduate Texts in Mathematics) third edition, Springer, 1987.

[Lang2] S. Lang, *Algebra* (Graduate Texts in Mathematics) Revised third edition, Springer, 1995.

[KhMo15] I.Khavkine and V. Moretti, *Algebraic QFT in Curved Spacetime and quasifree Hadamard states: an introduction* Chapter 5 of *Advances in Algebraic Quantum Field Theory* Springer 2015 (Eds R. Brunetti, C. Dappiaggi, K. Fredenhagen, and J.Yngvason)

[KNS] A. Kirillov, *Éléments de la Théorie de représentations* Mir, Mosca, 1974;
M.A. Najmark and A.I. Stern, *Theory of group representations*, Mir, Moscow, 1984.

[Moretti-a] V. Moretti,*Spectral Theory and Quantum Mechanics: Mathematical Foundations of Quantum Theories, Symmetries and Introduction to the Algebraic Formulation*, 2nd Edition, Springer, 2017

[Moretti-b] V. Moretti, *Fundamental Mathematical Structures of Quantum Theory Spectral Theory: Foundational Issues, Symmetries, Algebraic Formulation*, Springer 2019

[Moretti-c] V. Moretti, *Analytical Mechanics: Classical, Lagrangian and Hamiltonian Mechanics, Stability Theory, Special Relativity*, Springer 2023

[Moretti-d] V. Moretti, *Teoria della Relatività Speciale: formulazione matematica (Con un'introduzione alla formulazione della Relatività generale)*, dispense università di Trento https://moretti.maths.unitn.it/dispense.html

[Rudin] W. Rudin, Real and Complex Analysis, McGraw-Hill, 1987.

[Ruhl] W. Ruhl, *The Lorentz group and harmonic analysis*, W. A. Benjamin. Inc., New York, 1970.

[Sernesi] E. Sernesi, *Geometria* I and II, Bollati Boringhieri, Torino, 1990.

[Streater-Wightman] R.F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* with revised preface and corrections, Princeton University Press, Princeton and Oxford, 2000.

[Wightman] A. S. Wightman, *L'invariance dans la Mécanique quantique relativste* in *Relations de dispersion et Particules élémentaires*, edited by C. De Witt, Hermann, Paris, 1960.

[Wald] R.M. Wald *General Relativity*, Chicago University Press, Chicago, 1984.