Università degli Studi di Napoli Federico II

Dipartimento di Strutture per l'Ingegneria e l'Architettura

A methodological framework for the formulation of geometrically exact beam models

Ph.D. Dissertation

Massimo Paradiso

Dottorato di Ricerca in Ingegneria Strutturale, Geotecnica e Rischio Sismico - XXXIII ciclo

Tutor:

Prof. Ing. Luciano Rosati
Prof. Ing. Francesco Marmo

Coordinatore del corso:

Prof. Ing. Luciano Rosati

Contents

Chapter 1

Introduction

This dissertation has the purpose to describe the mathematical fundamentals and outline a methodological framework for formulating geometrically exact models for the analysis of beams undergoing large displacements.

The theme of high flexible beams has received growing attention during the last decades, what has been justified by the possibility of application in several fields of science and technology, from modeling of soft robotic manipulators (Grazioso et al., 2019a,b, Rucker et al., 2010, Rucker and Webster III, 2011, Sadati et al., 2017, Trivedi et al., 2008), of blades for wind turbines and helicopters (Cesnik and Shin, 2001, Faccio Júnior et al., 2019, Sabale and Gopal, 2019), even up to applications in biology (Balaeff et al., 2006, Coleman and Swigon, 2000, Swigon et al., 1998, Westcott et al., 1995).

In parallel, a number of beam models have been proposed in the last half century, based on several modeling strategies, aiming to capture the characteristic behavior of such a mechanical system.

1.1 Modeling of Beams undergoing Large Deformations

The response of a beam under large deflections reveals geometric non-linearity, that requires some numerical methodology to be managed.

With specific reference to the context of the finite element method, two fundamental strategies have been traditionally pursued to characterize the geometrically non-linear behavior of beams, which are the *corotational* approach and the *geometrical exact* one.

The main idea of the corotational approach relies on separating the rigid body component from the beam deformation, by defining a single element frame that continuously rotates with the element. Then, standard linear formulations are used with respect to the rotating frame and a non-linearity is introduced accounting for the finite rotation of such a frame.

Corotational formulations first appeared in the work of Belytschko and Hsieh (1973) and subsequent works (Belytschko and Glaum, 1979, Belytschko et al., 1977), as well as similar ideas were proposed by Oran (1973a,b) and Oran and Kassimali (1976). A comprehensive description of the first corotational approaches in beam modeling is provided by Crisfield (1990).

On the other side, the geometrically exact approach was developed exploiting a reduction process deriving the beam kinematics from the exact deformation analysis of a solid body.

The expression *geometrically exact* was first used by Simo and Fox (1989) to denote an approach for shell modeling, suitable for large scale computations, based on the analytical reduction of a solid model to a resultant form. The same expression was specifically applied to beam modeling two years later (Simo and Vu-Quoc, 1991), within the context of a series of papers started by Simo (1985) and Simo and Vu-Quoc (1986), describing a spatial beam formulation with the explicit intention to generalize the plane model proposed by Reissner (1972). The authors themselves considered their model as a reparameterization of the one proposed by Antman (1974), which extended the classical Kirchhoff-Love rod model (Love, 1944) by including finite extension and finite shearing.

From the perspective of Simo (1985) and coworkers, the beam model is derived by constraining the three-dimensional solid with the introduction of specific kinematic assumptions. The formulation leads to conceive the beam in terms of a three-dimensional orthogonal moving frame, with one of its axis remaining orthogonal to the beam cross-section in any configuration. This moving frame is also the reference system at which the resultant force and torque, acting on the typical cross-section, are evaluated. In this way, Simo (1985) goes back to the pioneering work of Cosserat and Cosserat (1909), also taken up by Ericksen and Truesdell (1957).

The 'exactness' of this approach relies on considering a rotation tensor, characterized as an element of the special orthogonal Lie group, denoted as $SO(3)$, to represent the three-dimensional spatial rotation of each beam cross-section. Then, the non-linearity of the algebraic space of the rotations is reflected in a non-linear beam model.

Meanwhile, Cardona and Geradin (1988) contributed to clarify the role of finite rotations in identifying the cross-section configuration, specifically focusing on a parameterization by the rotation vector and the relevant linearized variational representation. Further contributions about a vector parameterization of rotations and the consistent linearization procedure came from Ibrahimbegović et al. (1995), while an extension of the beam formulation including initially curved configurations was provided by Ibrahimbegović (1995).

A significant improvement in making clear how the features of the beam three-dimensional model are reflected on the resultant one-dimensional approximation came from Crisfield and Jelenić (1999). First, they pointed out that even if the theory of Simo (1985) and the derived models was often referred to as a 'geometrically exact finite-strain beam theory', an effective finite-strain formulation would require a consistent constitutive model, which is not straightforward to be implemented in conjunction with the usual assumption of the in-plane rigidity of cross-sections. For this reason, both in (Crisfield and Jelenić, 1999) and (Jelenić and Crisfield, 1998, 1999), the authors adopted the nomenclature of *geometrically exact* beam theory, dropping the term 'finite strain'.

A further clarification about how finite strains can be effectively included in a geometrically exact beam model is presented by Auricchio et al. (2008).

Beyond the formality about the appropriate nomenclature to use, the merit of Crisfield and Jelenić (1999) was to focus attention on the objectivity of the cross-section strain measures in finite element implementations based on geometrically exact beam theory. Actually, extending to rotations the conventional approach of directly interpolating the nodal kinematic parameters, or even their increments, leads to non-objectivity of the interpolated strain measures, even though the strain measures are themselves objective. In fact, being rotations non-additive quantities, regardless of the adopted parameterization, strains associated with a displacement-like interpolation of nodal parameters are not invariant under rigid rotations of the beam element's nodes.

Inspired by the previous work about the corotational beam formulation (Crisfield, 1990), Crisfield and Jelenić (1999) proposed an interpolation scheme based upon the relative rotation of the beam element's nodes, working on the local rotation vector representing the configuration of the final cross-section with respect to the initial one. A further refinement of such approach has been proposed by Magisano et al. (2020), as well as other strategies based on localizing the corotational decomposition have also been suggested (Garcea et al., 2012a,b). Objectivity of strain measures is also taken into account in recent isogeometric formulations of spatial beam models (Harsch et al., 2021).

Even if a significant improvement towards a better insight of beam modeling came from the work of Crisfield and Jelenić (1999), the beam axis representation was still based upon a classical polynomial interpolation, separately from the scheme adopted for the rotations of the cross-section local frame. Actually, two distinct configuration spaces were considered: the one relevant to the rotations was the Lie group $SO(3)$, along with a consistent non-linear interpolation scheme, while the classical Euclidean vector space $\mathbb{R}^3$ was assumed for representing the beam axis.

An alternative approach was proposed by Borri and Bottasso (1994), that considered helicoidal shape functions to describe the spatial configuration of the beam axis. Differently from polynomial interpolation, the resulting model is invariant with respect to the choice of the reference line, meaning that any other reference line has a helicoidal shape. Such an approach implies a coupled representation of beam axis displacements and cross-section rotations.

The necessity to consider coupled interpolation schemes for displacements and rotations, in favor of the accuracy of the formulation, was particularly pointed out by Zupan and Saje (2003a,b), considering a strain-based approach, rather than a position-based one.

A coupled representation of translation and rotation fields naturally arises from the approach proposed by Sonneville et al. (2014). Such a formulation, inspired by the interpolation methods of rigid body motions (Park, 1995, Žefran and Kumar, 1998), relies on considering the special Euclidean Lie group, namely $SE(3)$, as the configuration space, consisting of the spatial proper rigid motions. A similar beam model, formulated by explicitly representing the kinematic parameters in the physical space, is proposed by Santana et al. (2021).

Even if a rigid motion can be thought as a rotation of $SO(3)$ and a translation in $\mathbb{R}^3$, the composition rule characterizing $SE(3)$ takes into account the coupling between rotations and translations, so that, by means of an exponential interpolation method on $SE(3)$, a natural coupling of the position and the rotation variables follows.

1.2 Outline

This thesis adopts a formalism similar to the one proposed by Sonneville et al. (2014), describing the formulation of a beam model within the framework of the Lie group of Euclidean rigid transformations.

In particular, it will be shown how the Lie group formulation naturally arises from the geometric description of the beam kinematics, leading to a characterization of a beam cross-section configuration as an affine transformation within the physical space.

For this reason, in order to include the beam model within a framework as rigorous and consistent as possible, the algebraic structure of the group of affine transformations is deeply investigated, with particular attention towards the subgroup of proper Euclidean motions. The differential features of such a space are as well analyzed, and are presented as the specialization of more general properties characterizing smooth manifolds.

At the same time, the mechanical meaning of the algebraic and differential operations involved in the development of geometrically exact beam models is emphasized, showing how the mathematical formalism has an effective counterpart in the behavior of the beam as a mechanical system.

Hence, aim of this work is to provide a mathematical validation, as clear as possible, to some intuitive ideas which guide any engineer in formulating a mechanical model and, at once, to show how the properties of some algebraic structures, sometimes apparently counter-intuitive and misleading, are actually related to what we call 'reality'.

Since the formalism of differential geometry is crucial in providing a comprehensive mathematical framework for beam modeling, Chapter 2 is dedicated to review some essential notions about smooth manifolds, with specific focus on Lie groups. In Chapter 3 the basic features of affine spaces are briefly recalled, and particular attention is given to the algebraic structure characterizing the group of affine transformations. Chapter 4 is focused on Euclidean affine spaces and isometric transformations, also describing the differential structure of the group of rotations, $SO(3)$, and Euclidean motions, $SE(3)$.

In Chapter 5 the kinematics of a geometrically exact beam model is derived as the specialization of a solid body model, under some appropriate geometric and kinematic assumptions. It is also remarked the interpretation of the beam configuration as a curve on the Lie group $SE(3)$, that is actually assumed as the configuration space for the beam kinematics. This also implies the algebraic definition of internal forces and external loads as functionals for the virtual displacement space, the only one that effectively makes sense in a non-linear manifold context. At the same time, a proper definition of the cross-section strain measures is introduced on the basis of a local linearization of the beam motion. In this way, the infinitesimal variation of the beam configuration is treated as a linear model whose reference system is represented by the current beam configuration. Consequently, the cross-section stiffness matrix, considered as a linear operator between the infinitesimal variation of the strain parameters and the relevant variation of the internal forces, is recovered from the results of the cross-section analysis based on linear beam models.

Chapter 6 concerns a finite element beam model relying on specific assumptions about the beam deformation field, along with some introductory numerical tests. A brief discussion about some aspects to be specified in greater detail and further issues to be investigated is reported Chapter 7.

Finally, in order to make this work suitably self-contained, some common algebraic structures are briefly described in Appendix A, as well as a review of linear algebra is specifically reported in Appendix B.

Chapter 2

Smooth Manifolds and Lie Groups

2.1 Smooth Manifolds

A smooth manifold can be intended as a space that locally looks like some Euclidean space $\mathbb{R}^n$. It is a topological space enriched with an additional differentiable structure.

2.1.1 Topological Manifolds

A topological space is a mathematical space where the notion of limit, continuity and connectedness are introduced in the most general sense.

Here, some essential notions about topology and topological spaces are briefly summarized. For a more detailed description one can refer to Lee (2012).

Definition 2.1. Let X be a set. A *topology* on X is a collection τ of subsets of X, called *open subsets*, satisfying the following axioms:

- X and $\varnothing$ are in τ;
- the union of any family of open subsets is open;
- the intersection of any finite family of open subsets is open.

The set X along with a topology τ on it is called a *topological space*. In addition, we say that X is a *Hausdorff space* if for any pair of distinct points $p, q \in X$ there exist disjoint open subsets $U, V \subseteq X$ such that $p \in U$ and $q \in \mathcal{V}$. The open subsets U and V are *neighborhoods* of p and q, respectively.

A sequence of points $p_1, \ldots, p_\infty$ of X is said to *converge* to a point $p \in X$ if for every neighborhood U of p there exists an integer N such that $p_i \in U$ for all $i \geq N$. If the X is a Hausdorff space, the limit of any convergent sequence is unique.

Given two topological spaces X and Y, we say that a map $f : X \to Y$ is *continuous* if for any open subset V of Y the preimage $f^{-1}(V)$ is open in X. When the continuous map f is bijective and the inverse map is also continuous, it is called a *homeomorphism* and the topological spaces X and Y are said to be *homeomorphic*.

Given a topological space X, a family B of open subsets of X is a *basis* for the topology of X if every open subset of X is the union of some sub-family of B. Also, we say that X is *second-countable* if there exists a countable basis for its topology.

On the basis of these preliminary notions, we can introduce the definition of topological manifold.

Definition 2.2. Suppose M is a topological space. We say that M is a *topological manifold* of dimension n, or a *topological n-manifold*, if it is

- a Hausdorff space;
- second-countable;
- locally Euclidean of dimension n.

We remark that the requirement of being locally Euclidean means that every point $p \in M$ has a neighborhood U which is homeomorphic to an open subset of $\mathbb{R}^n$.

Definition 2.3. Let M be a topological n-manifold. A *coordinate chart*, or simply a *chart*, on M is a pair (U, φ) consisting of a open subset $U \subseteq M$ and a homeomorphism $\varphi\colon U \to \widehat{U} \subseteq \mathbb{R}^n$.

If $\varphi(p) = \mathbf{o}$, we say that the chart is *centered at* p. When $\varphi(p) \neq \mathbf{o}$, the chart can be centered at p by subtracting the constant vector $\varphi(p)$.

The set U of a chart (U, φ) is called a *coordinate domain*, or a *coordinate neighborhood*, of each of its points. The homeomorphism φ is the *(local) coordinate map* and consists of n functions $x^1, \dots, x^n$, which are called *local coordinates* on U, such that

$$\varphi(p) = \big(x^1(p), \dots, x^n(p)\big) \in \mathbb{R}^n\,, \quad \forall\, p \in U\,.$$

A collection of charts $A = \big\{(U_i, \varphi_i)\big\}_{i\in I}$ is called an *atlas* for M if the domains cover the whole topological manifold:

$$\bigcup_{i\in I} U_i = M\,.$$

2.1.2 Smooth Structure

The definition of a topological manifold allows one to extend to an arbitrary space, endowed with a topology, some notions characteristic of Euclidean spaces, such as the limit of a succession, continuity and connectedness.

However, one might ask whether further properties, such as differentiability and other related notions, could also apply.

To investigate such a possibility, a topological space needs to be endowed with an additional structure, which is the property of being smooth.

The definition of a smooth manifold relies on multivariable calculus, so we first recall the property of being smooth applied to a multivariable real function.

Definition 2.4. Let U and V be open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. A function $F\colon U \to V$ is *smooth*, also said C^∞, or *infinitely differentiable*, if each of its components functions has continuous partial derivatives of all orders.

In addition, if F is bijective with a smooth inverse map, it is called a *diffeomorphism*. Clearly, a diffeomorphism is also a homeomorphism.

Since smoothness is defined for real multivariable functions, and recalling that a manifold can be associated with a Euclidean space through coordinate charts, one could somehow combine such properties in order to extend the notion of differentiability to maps defined on manifolds.

At the same time, with the aim of properly define a smooth structure on a manifold, the features of such a structure should be independent of the choice of the chart.

Let us consider two overlapping charts of an n-dimensional topological manifold M, say (U, φ) and (V, ψ). The intersection of the domains U and V is a non-empty open set of M and provides a common subdomain of both the coordinate maps φ and ψ, so that the one can be composed with the inverse of the other:

$$\psi \circ \varphi^{-1} \colon \varphi(U \cap V) \to \psi(U \cap V)\,. \tag{2.1}$$

The map $\psi \circ \varphi^{-1}$ is called the *transition map* from φ to ψ and, being the composition of homeomorphisms, results an homeomorphism itself.

Noting that both $\varphi(U \cap V)$ and $\psi(U \cap V)$ are open subsets of $\mathbb{R}^n$, we say that the charts (U, φ) and (V, ψ) are *smoothly compatible* if either $U \cap V = \emptyset$ or the transition map $\psi \circ \varphi^{-1}$ is a diffeomorphism.

An atlas A for M is called a *smooth atlas* if any two charts in A are smoothly compatible.

Moreover, we say that a smooth atlas A is *maximal* if it is not properly contained in any larger smooth atlas. This means that any chart which is compatible with every chart in A is already in the atlas A.

The need to consider maximal atlases for a manifold M comes from the observation that adding a chart (U, φ) to a smooth atlas A provides a new atlas, say A', assuming (U, φ) is smoothly compatible with every chart already included in A. However, even if A and A' are formally distinct, de facto the latter atlas cannot give any additional information about M if compared with the former one.

For this reason we use maximal atlases to define a smooth structure on a manifold.

Definition 2.5. A *smooth manifold* is a pair (M, A) made by a topological manifold M and a maximal smooth atlas A on M.

We say that the maximal smooth atlas A is a *smooth structure*, also called a *differentiable structure* or a C^∞ *structure*, on M.

Furthermore, when the smooth structure is understood, it is usually omitted and M is itself referred to as a *smooth manifold*.

Remark. It is possible to prove that every smooth atlas for a manifold is contained in a unique maximal smooth atlas. Also, if the union of two smooth atlases is itself a smooth atlas on a manifold, then they are contained in the same maximal atlas, and vice-versa (see, e.g., Lee (2012)).

This means that a maximal smooth atlas induces an equivalence class, which consists of all the smooth atlases contained in it. Equivalently, such an equivalence class is defined by all the smooth atlases whose union is in turn a smooth atlas.

As a consequence, any smooth atlas on a manifold is representative of a single maximal atlas and determines a unique smooth structure.

Any chart (U, φ) contained in the given maximal smooth atlas is a *smooth* chart. The term 'smooth' also applies to the domain U and the coordinate map φ defining the chart.

Once a smooth chart on the manifold M has been chosen, the coordinate map $\varphi\colon U \to \widehat{U}$ provides an identification between U and $\widehat{U} \subseteq \mathbb{R}^n$. This means that the domain U can be thought as an open subset of M and, at the same time, as open subset of $\mathbb{R}^n$. Likewise, a point $p \in U$ can be identified with the n-tuple $(x^1, \dots, x^n) = \varphi(p)$.

We remark that the identification of a subset of M with a subset of $\mathbb{R}^n$ is only local and is strictly related with the choice of the coordinate chart. In order to emphasize such a characteristic, we say that the n-tuple $(x^1, \dots, x^n)$ is the *local coordinate representation* of p.

Smooth Functions and Maps

Let M be a smooth n-dimensional manifold. A *smooth function* is a map $f\colon M \to \mathbb{R}^k$ such that for every $p \in M$ there exists a smooth chart (U, φ) such that the composite function $f \circ \varphi^{-1}$ is smooth on $\widehat{U} = \varphi(U) \subseteq \mathbb{R}^k$.

The function $\widehat{f}\colon \varphi(U) \to \mathbb{R}^k$, defined by the property

$$\widehat{f}(\mathbf{x}) = (f \circ \varphi^{-1})(\mathbf{x}), \quad \forall\, \mathbf{x} \in \varphi(U),$$

is called the *coordinate representation* of f.

With specific reference to the case of $\mathbb{R}$ as codomain, the set of the real-valued functions $f\colon M \to \mathbb{R}$ is denoted as C^∞_M.

Let us now consider two smooth manifolds, say M and N, with dimensions n and k, respectively. A map $F\colon M \to N$ is a *smooth map* if for every point p of M there exist smooth charts (U, φ) and (V, ψ) containing p and $F(p)$, respectively, such that $F(U) \subseteq V$ and the composite map $\psi \circ F \circ \varphi^{-1}$, between $\varphi(U) \subseteq \mathbb{R}^n$ and $\psi(V) \subseteq \mathbb{R}^k$, is smooth.

Similarly to smooth functions, the *coordinate representation* of F is given by the function $\widehat{F}\colon \varphi(U) \to \psi(V)$ such that

$$\widehat{F}(\mathbf{x}) = (\psi \circ F \circ \varphi)(\mathbf{x}), \quad \forall\, \mathbf{x} \in \varphi(U).$$

Please observe that any smooth function $f : M \to \mathbb{R}^k$ can be seen as the specialization of a smooth map with the assumptions $N = V = \mathbb{R}^k$ and $\psi = \mathrm{id}_{\mathbb{R}^k}$.

In addition, a smooth map $F\colon M \to N$ is called a *diffeomorphism* from M to N if it is bijective with a smooth inverse. In such a case, we also say that M and N are *diffeomorphic*.

Clearly, since the coordinate representation of F is a diffeomorphism between open subsets of $\mathbb{R}^n$ and $\mathbb{R}^k$, respectively, the condition $n = k$ is necessary for M and N to be diffeomorphic.

2.2 Tangent Space

As in Euclidean spaces the crucial idea is the linear approximation, a similar notion can be defined on manifold by introducing the tangent space at a point, which can be seen as a sort of linearization of the manifold in a neighborhood of a point.

2.2.1 Geometric Tangent Vectors

Let us consider the n-dimensional real space $\mathbb{R}^n$. Each element $\mathbf{x}$ of $\mathbb{R}^n$ can be thought as a point in space, whose location is expressed by an n-tuple of coordinates $(x^1, \dots, x^n)$.

At the same time, the vector space structure of $\mathbb{R}^n$ leads one to visualize an element as a vector $\mathbf{v} = v^i \mathbf{e}_i$, where $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is the standard basis, whose geometric meaning is defined in terms of magnitude and direction, no matter where it is located.

In order to associate a vector with a point of the space $\mathbb{R}^n$, we introduce the following definition.

Definition 2.6. Given a point $\mathbf{a}$ of $\mathbb{R}^n$, we define the *geometric tangent space* of $\mathbb{R}^n$ at $\mathbf{a}$ as the set

$$\mathbb{R}^n_{\mathbf{a}} = \{\mathbf{a}\} \times \mathbb{R}^n = \{\, (\mathbf{a}, \mathbf{v}) \mid \mathbf{v} \in \mathbb{R}^n \,\} \,. \tag{2.2}$$

A *geometric tangent vector* at $\mathbf{a} \in \mathbb{R}^n$ is an element $(\mathbf{a}, \mathbf{v})$ of $\mathbb{R}^n_{\mathbf{a}}$, denoted as $\mathbf{v}_{\mathbf{a}}$ or also as $\mathbf{v}|_{\mathbf{a}}$.

The vector $\mathbf{v}_{\mathbf{a}}$ can be thought as a copy of $\mathbf{v}$ representing an oriented line segment with the initial point at $\mathbf{a}$. The geometric tangent space $\mathbb{R}^n_{\mathbf{a}}$ as a whole can be seen as a copy of the real vector space $\mathbb{R}^n$ with the null vector located at $\mathbf{a}$.

As such, $\mathbb{R}^n_{\mathbf{a}}$ satisfies the linear space properties, that is

$$(\mathbf{v} + \mathbf{w})_{\mathbf{a}} = \mathbf{v}_{\mathbf{a}} + \mathbf{w}_{\mathbf{a}} \,, \quad (c\mathbf{v})_{\mathbf{a}} = c\mathbf{v}_{\mathbf{a}} \,,$$

and, because of the natural isomorphism $\mathbb{R}^n_{\mathbf{a}} \cong \mathbb{R}^n$, it inherits all the property of $\mathbb{R}^n$.

Any geometric tangent vector $\mathbf{v}_{\mathbf{a}}$ yields a map $\mathrm{D}_{\mathbf{v}}|_{\mathbf{a}} : C^\infty_{\mathbb{R}^n} \to \mathbb{R}$, which takes the directional derivative with respect to $\mathbf{v}$ at $\mathbf{a}$:

$$\mathrm{D}_{\mathbf{v}}|_{\mathbf{a}}(f) = \mathrm{D}_{\mathbf{v}} f(\mathbf{a}) = \left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu=0} f(\mathbf{a} + \mu\mathbf{v}) \,.$$

Such an operator is linear over $\mathbb{R}$ and satisfies Leibniz' product rule:

$$\mathrm{D}_{\mathbf{v}}|_{\mathbf{a}}(fg) = f(\mathbf{a})\,\mathrm{D}_{\mathbf{v}}|_{\mathbf{a}}(g) + g(\mathbf{a})\,\mathrm{D}_{\mathbf{v}}|_{\mathbf{a}}(f)\,.$$

Moreover, considering the representation of the vector $\mathbf{v_a}$ with respect to the standard basis of $\mathbb{R}^n_{\mathbf{a}}$ as $\mathbf{v_a} = v^i\,\mathbf{e}_i|_{\mathbf{a}}$, the map $\mathrm{D}_{\mathbf{v}}|_{\mathbf{a}}$ applied at the function f reads

$$\mathrm{D}_{\mathbf{v}}|_{\mathbf{a}}(f) = v^i \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{a}}.$$

The construction obtained from the properties of directional derivative can be extended to a more general map by introducing the following definition.

Definition 2.7. Given a point $\mathbf{a} \in \mathbb{R}^n$, we say that a map $\mathbf{w}\colon C^\infty_{\mathbb{R}^n} \to \mathbb{R}$ is a *derivation* at $\mathbf{a}$ if it is linear over $\mathbb{R}$ and satisfies Leibniz' product rule:

- $\mathbf{w}(cf) = c\mathbf{w}(f)\,,\ \forall c \in \mathbb{R}\,,\ f \in C^\infty_{\mathbb{R}^n}$;
- $\mathbf{w}(fg) = f(\mathbf{a})\mathbf{w}(g) + g(\mathbf{a})\mathbf{w}(f)\,,\ \forall f, g \in C^\infty_{\mathbb{R}^n}$.

The set of all the derivations of $C^\infty_{\mathbb{R}^n}$ at $\mathbf{a}$ is denoted as $T_{\mathbf{a}}\mathbb{R}^n$. It is a vector space under addition and scalar multiplication.

Clearly, any map in the form $\mathrm{D}_{\mathbf{v}}|_{\mathbf{a}}$, providing the directional derivative at $\mathbf{a}$ relevant to the vector $\mathbf{v}$, is a derivation. In addition, it is possible to prove that the map associating $\mathrm{D}_{\mathbf{v}}|_{\mathbf{a}}$ with the geometric tangent vector $\mathbf{v_a}$ is an isomorphism (see, e.g. Lee (2012)).

As a consequence of the isomorphism $\mathbb{R}^n_a \cong T_{\mathbf{a}}\mathbb{R}^n$, a basis for $T_{\mathbf{a}}\mathbb{R}^n$ is given by the set

$$\left\{ \left.\frac{\partial}{\partial x^1}\right|_{\mathbf{a}}, \dots, \left.\frac{\partial}{\partial x^n}\right|_{\mathbf{a}} \right\}, \tag{2.3}$$

where $\partial/\partial x^i|_{\mathbf{a}} = \mathrm{D}_{\mathbf{e}_i}|_{\mathbf{a}}$ is the derivation such that

$$\left.\frac{\partial}{\partial x^i}\right|_{\mathbf{a}}(f) = \left.\frac{\partial f}{\partial x^i}\right|_{\mathbf{a}}.$$

2.2.2 Tangent Vectors on Manifolds

We introduce now the definition of tangent vectors on manifolds.

Definition 2.8. Let M be a smooth manifold and p a point of M. A linear map $\mathbf{v}\colon C^\infty_M \to \mathbb{R}$ is called a *derivation* at p if it satisfies Leibniz' product rule:

$$\mathbf{v}(fg) = f(p)\mathbf{v}(g) + g(p)\mathbf{v}(f)\,, \;\; \forall\, f, g \in C^\infty_M\,. \tag{2.4}$$

The derivation $\mathbf{v}$ is also called the *tangent vector* at p and the set of all the derivations at p, denoted as T_pM is called the *tangent space* to M at p.

Please notice that the sum of the derivations $\mathbf{v}, \mathbf{w} \in T_pM$ is defined as the map $\mathbf{v} + \mathbf{w}$ satisfying the point-wise addition, that is

$$(\mathbf{v} + \mathbf{w})(f) = \mathbf{v}(f) + \mathbf{w}(f)\,, \;\; \forall\, f \in C^\infty_M\,,$$

as well as the multiplication of $\mathbf{v} \in T_pM$ by a scalar $c \in \mathbb{R}$ is such that

$$(c\mathbf{v})(f) = c\,\mathbf{v}(f)\,, \;\; \forall\, f \in C^\infty_M\,.$$

One can easily verify that Leibniz' product rule is fulfilled for the addition, i.e.

$$\begin{aligned}
(\mathbf{v} + \mathbf{w})(fg) &= \mathbf{v}(fg) + \mathbf{w}(fg) \\
&= f(p)\mathbf{v}(g) + g(p)\mathbf{v}(f) + f(p)\mathbf{w}(g) + g(p)\mathbf{w}(f) \\
&= f(p)\big(\mathbf{v}(g) + \mathbf{w}(g)\big) + g(p)\big(\mathbf{w}(f) + \mathbf{w}(g)\big) \\
&= f(p)\big(\mathbf{v} + \mathbf{w}\big)(g) + g(p)\big(\mathbf{v} + \mathbf{w}\big)(f)\,,
\end{aligned}$$

and similarly for the scalar multiplication:

$$\begin{aligned}
(c\mathbf{v})(fg) &= c\,\mathbf{v}(fg) \\
&= c\big(f(p)\mathbf{v}(g) + g(p)\mathbf{v}(f)\big) \\
&= f(p)\big(c\mathbf{v}(g)\big) + g(p)\big(c\mathbf{v}(f)\big) \\
&= f(p)\big(c\mathbf{v}\big)(g) + g(p)\big(c\mathbf{v}\big)(f)\,.
\end{aligned}$$

Hence, it is confirmed that T_pM has a linear space structure.

2.2.3 The Differential of a Smooth Map

Definition 2.9. Let $F\colon M \to N$ be a smooth map from M to N. For every point $p \in M$, the *differential* of F at p is the map $\mathrm{d}F|_p : T_pM \to T_{F(p)}N$ such that

$$(\mathrm{d}F|_p\mathbf{v})(f) = \mathbf{v}(f \circ F)\,, \ \ \forall\, \mathbf{v} \in T_pM\,,\ f \in C^\infty_N\,, \tag{2.5}$$

where $\mathrm{d}F|_p\mathbf{v}$ is the imagine of $\mathbf{v}$ through $\mathrm{d}F|_p$.

To be sure that the above definition actually makes sense, please observe that $\mathrm{d}F|_p\mathbf{v}$ is a vector of the tangent space $T_{F(p)}N$ and, as a derivation, acts on a real-valued function f of C^∞_N. At the same time, the composition $f \circ F$ is in C^∞_M and the action of $\mathbf{v} \in T_pM$ on it is well-defined.

The linearity of $\mathrm{d}F|_p\mathbf{v}\colon C^\infty_N \to \mathbb{R}$ readily derives from the one of $\mathbf{v} \in T_pM$:

$$\begin{aligned}(\mathrm{d}F|_p\mathbf{v})(f+g) &= \mathbf{v}\big((f+g)\circ F\big) = \mathbf{v}(f\circ F) + \mathbf{v}(g \circ F)\\ &= (\mathrm{d}F|_p\mathbf{v})(f) + (\mathrm{d}F|_p\mathbf{v})(g)\,,\end{aligned}$$

as well as

$$(\mathrm{d}F|_p\mathbf{v})(cf) = \mathbf{v}\big((cf)\circ F\big) = c\mathbf{v}(f\circ F) = c(\mathrm{d}F|_p\mathbf{v})(f)\,.$$

In addition, since Leibniz' rule is fulfilled, $\mathrm{d}F|_p\mathbf{v}$ is actually a derivation on N:

$$\begin{aligned}(\mathrm{d}F|_p\mathbf{v})(fg) &= \mathbf{v}\big((fg)\circ F\big) = \mathbf{v}\big((f\circ F)(g\circ F)\big)\\ &= (f\circ F)(p)\,\mathbf{v}(g\circ F) + (g\circ F)(p)\,\mathbf{v}(f\circ F)\\ &= f\big(F(p)\big)\,(\mathrm{d}F|_p\mathbf{v})(g) + g\big(F(p)\big)\,(\mathrm{d}F|_p\mathbf{v})(f)\,.\end{aligned}$$

The main properties of the differentials of smooth maps are summarized by the following proposition.

Proposition 2.10. *Let M, N, P be smooth manifolds and let $F\colon M \to N$ and $G : N \to P$ be smooth maps. Then, for any $p \in M$, the following properties hold true.*

1. $\mathrm{d}F|_p\colon T_pM \to T_{F(p)}N$ *is linear.*

2. $\mathrm{d}(G \circ F)|_p = \mathrm{d}G|_{F(p)} \circ \mathrm{d}F|_p$.
3. $\mathrm{d}(\mathrm{id}_M)|_p = \mathrm{id}_{T_pM}$.
4. *If F is a diffeomorphism, then $\mathrm{d}F|_p : T_pM \to T_{F(p)}N$ is an isomorphism, and $(\mathrm{d}F|_p)^{-1} = \mathrm{d}(F^{-1})|_{F(p)}$.*

2.2.4 The Tangent Bundle

Given a smooth manifold M, the *tangent bundle* of M, denoted as TM, is the disjoint union of the tangent spaces at all points of M:

$$TM = \bigsqcup_{p \in M} T_pM\,. \tag{2.6}$$

An element of TM is a pair $(p, \mathbf{v})$, where $p \in M$ and $\mathbf{v} \in T_pM$, also simplified in $\mathbf{v}_p$.

There exists a natural projection map associating an element $(p, \mathbf{v})$ of TM with the point p of the manifold:

$$\begin{aligned} \pi : \quad & TM \to M \\ & (p, \mathbf{v}) \mapsto p\,. \end{aligned} \tag{2.7}$$

It can be proved that for any n-dimensional smooth manifold M, the tangent bundle has a natural topology and a smooth structure (see, e.g., Lee (2012)). Specifically TM is a $2n$-dimensional smooth manifold and the projection $\pi \colon TM \to M$ is also smooth.

When two manifolds M and N are considered, along with the relevant tangent bundles TM and TN, the following map is introduced:

$$\begin{aligned} \mathrm{d}F : \quad & TM \to TN \\ & (p, \mathbf{v}) \mapsto \left(F(p), \mathrm{d}F|_p\mathbf{v}\right), \end{aligned} \tag{2.8}$$

which is called the *global differential* of F.

The global differential $\mathrm{d}F$ is a smooth map. Such a property comes from the smoothness of its coordinate representation, along with the one of the map F.

Evidently, the differential $\mathrm{d}F|_p$ of F at p is the restriction of $\mathrm{d}F$ to the tangent space $T_pM \subseteq TM$. Consequently, the properties of $\mathrm{d}F|_p$ specified by Proposition 2.10 can be generalized to the global differential $\mathrm{d}F$.

Corollary 2.11. *Let M, N, P be smooth manifolds and let $F\colon M \to N$ and $G\colon N \to P$ be smooth maps. Then, the following properties apply.*

1. $\mathrm{d}(G \circ F) = \mathrm{d}G \circ \mathrm{d}F$.
2. $\mathrm{d}(\mathrm{id}_M) = \mathrm{id}_{TM}$.
3. *If F is a diffeomorphism, then $\mathrm{d}F : TM \to TN$ is also a diffeomorphism, and $(\mathrm{d}F)^{-1} = \mathrm{d}(F^{-1}) = \mathrm{d}F^{-1}$.*

2.2.5 Velocity Vectors of Curves

Definition 2.12. If M be a smooth manifold and $J \subseteq \mathbb{R}$ an interval. A *curve* on M is a continuous map $\gamma\colon J \to M$.

Since the interval J is a 1-dimensional manifold, the curve γ represents a smooth map, whose global differential is $\mathrm{d}\gamma\colon TJ \to TM$, where TJ and TM are the tangent bundles of J and M, respectively.

Moreover, given $\mu_0 \in J$, the differential of γ at μ_0 maps a vector in $T_{\mu_0}J$ to a vector $\mathbf{v}$ of $T_{\gamma(\mu_0)}M$. We say that the vector $\mathbf{v}$ is the *velocity* of the curve γ at μ_0:

$$\mathbf{v} = \left.\frac{\mathrm{d}\gamma}{\mathrm{d}\mu}\right|_{\mu_0} = \mathrm{d}\gamma|_{\mu_0} \left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu_0} \in T_{\gamma(\mu_0)}M\,, \tag{2.9}$$

where $\mathrm{d}/\mathrm{d}\mu|_{\mu_0}$ is the standard coordinate basis vector of $T_{\mu_0}J$.

The velocity is often denoted as $\gamma'(\mu_0)$, or also $\gamma'|_{\mu_0}$, just as one usually represents the derivative of a 1-parameter function in ordinary calculus.

At the same time, with the aim to point out the role of $\mathbf{v}$ as a tangent vector to the manifold M, we likewise say that it is the velocity of γ at p, meaning that $p = \gamma(\mu_0)$.

Furthermore, applying the definition (2.5) of the differential to $\mathrm{d}\gamma_{\mu_0}$, the velocity $\gamma'(\mu_0)$ acts on a function f of C^∞_M as follows:

$$\big(\gamma'(\mu_0)\big)(f) = \left(\mathrm{d}\gamma|_{\mu_0} \left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu_0}\right)(f) = \left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu_0}(f \circ \gamma) = (f \circ \gamma)'(\mu_0)\,. \tag{2.10}$$

It is worth noting that if μ_0 is an endpoint of J, the definition of the velocity still holds, provided that we intend the derivative with respect to μ as a one-side derivative.

Proposition 2.13. *Let $\gamma: J \to M$ be an curve on a smooth manifold M and let $\tau\colon \tilde{J} \to J$ be a differentiable real-valued function whose domain is the interval $\tilde{J} = \{\, \mu \in \mathbb{R} \mid \tau(\mu) \in J \,\}$. Then, the map $\tilde{\gamma}: \tilde{J} \to M$, defined by $\tilde{\gamma}(\mu) = \gamma\big(\tau(\mu)\big)$, is a curve on M whose velocity vector is*

$$\tilde{\gamma}'(\mu) = \tau'(\mu)\,\gamma'\big(\tau(\mu)\big)\,. \tag{2.11}$$

Proof. Consider an arbitrary $f \in C^\infty_M$ and apply the property (2.10) to the velocity of $\tilde{\gamma}$:

$$\big(\tilde{\gamma}'(\mu)\big)(f) = (f \circ \tilde{\gamma})'(\mu) = (f \circ \gamma \circ \tau)'(\mu) = (f \circ \gamma)'\big(\tau(\mu)\big)\,\tau'(\mu)\,,$$

where the last equality comes from the chain rule applied to the composition of $f \circ \gamma$ and τ.

Hence, again by (2.10), one has

$$\big(\tilde{\gamma}'(\mu)\big)(f) = \tau'(\mu)\,\Big(\gamma'\big(\tau(\mu)\big)\Big)(f)\,,\ \ \forall f \in C^\infty_M\,,$$

whence, for the arbitrariness of f, one finally obtains (2.11). □

Proposition 2.14. *Let M be an n-dimensional smooth manifold and consider the tangent space T_pM at a point $p \in M$. Then, every $\mathbf{v} \in T_pM$ is the velocity of a smooth curve.*

Proof. Let (U, φ) be a smooth coordinate chart centered at p and let $v^1, \dots, v^n$ be the components of $\mathbf{v}$ with respect to the induced basis for T_pM:

$$\mathbf{v} = v^i \left.\frac{\partial}{\partial x^i}\right|_p .$$

For a sufficiently small $\epsilon > 0$, consider a curve $\gamma: (-\epsilon, \epsilon) \to U$, such that its coordinate representation $\widehat{\gamma}\colon (-\epsilon, \epsilon) \to \varphi(U)$ is

$$\widehat{\gamma}(\mu) = (\mu v^1, \dots, \mu v^n)\,.$$

Hence, γ is a smooth curve satisfying $\gamma(0) = \varphi^{-1}\big(\widehat{\gamma}(0)\big) = \varphi^{-1}(\mathbf{o}) = p$, so that evaluating the velocity in local coordinates easily reads

$$\gamma'(0) = v^i \left.\frac{\partial}{\partial x^i}\right|_{\gamma(0)} = \mathbf{v} \in T_pM\,. \qquad \square$$

Any curve of a manifold which is the domain of some smooth maps, by composition, induces a curve also on the codomain.

Proposition 2.15. *Let $F\colon M \to N$ be a smooth map between the smooth manifolds M and N, and let $\gamma\colon J \to M$ be a curve. Then, the velocity of the composite curve $F \circ \gamma\colon J \to N$ at any $\mu_0 \in J$ is given by*

$$(F \circ \gamma)'(\mu_0) = \mathrm{d}F|_{\gamma(\mu_0)}\gamma'(\mu_0)\,.$$

Proof. Applying (2.9), the velocity of $F \circ \gamma$ at μ_0 is

$$(F \circ \gamma)'(\mu_0) = \mathrm{d}(F \circ \gamma)|_{\mu_0} \left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu_0} = \mathrm{d}F|_{\gamma(\mu_0)}\left(\mathrm{d}\gamma|_{\mu_0}\left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu_0}\right),$$

that is

$$(F \circ \gamma)'(\mu_0) = \mathrm{d}F|_{\gamma(\mu_0)}\gamma'(\mu_0)\,. \qquad \square$$

The result of the Proposition 2.15 can be exploited to evaluate the differential of a smooth map.

Corollary 2.16. *Given the smooth map $F\colon M \to N$ between the manifolds M and N, the differential of F at a point $p \in M$ applied at a vector $\mathbf{v} \in T_pM$ is*

$$\mathrm{d}F|_p\mathbf{v} = (F \circ \gamma)'(0)\,,$$

for any smooth curve $\gamma : J \to M$, with $0 \in J$, such that $\gamma(0) = p$ and $\gamma'(0) = \mathbf{v}$.

Proof. The proof follows from Proposition 2.15 specialized at $\mu_0 = 0$. $\square$

2.3 Vector Fields

We have seen that, along with a smooth manifold M, the tangent bundle TM is also defined. In addition, the map $\pi\colon TM \to M$, projecting the pair $(p, \mathbf{v}) \in TM$ onto the manifold, provides the point $p \in M$.

A vector field on the manifold can be thought as an inverse operation of the projection, i.e. a map associating a tangent vector with each point of M, as formally specified through the following definition.

Definition 2.17. Let M be a smooth manifold and TM the tangent bundle. A *vector field* on M is a section of the projection map π, that is a map $\boldsymbol{X}\colon M \to TM$ satisfying

$$\pi \circ \boldsymbol{X} = \mathrm{id}_M\,,$$

where $\mathbf{id}_M$ is the identity map on M.

Explicitly, a vector field on M is a continuous map in the form

$$\begin{aligned} \boldsymbol{X}:\ & M \to TM \\ & p \mapsto \boldsymbol{X}(p) = \boldsymbol{X}_p\,, \end{aligned}$$

where $\boldsymbol{X}_p$ is clearly a vector of the tangent space $T_pM \subseteq TM$.

Please notice that since the tangent bundle has a natural smooth structure, smoothness also applies to vector fields.

If (U, φ) is a smooth chart with the local coordinates $x^1, \dots, x^n$, the value of $\boldsymbol{X}$ at a point $p \in U$ can be expressed in terms of coordinate basis vectors as

$$\boldsymbol{X}_p = X^i(p)\,\frac{\partial}{\partial x^i}\bigg|_p\,,$$

where n maps $X^i\colon U \to \mathbb{R}$ are the *component functions* of $\boldsymbol{X}$.

The set of all the smooth vector fields on M is denoted as $\mathfrak{X}(M)$. Such a set can be endowed with the addition and scalar multiplication by point-wise evaluation:

$$(a\boldsymbol{X} + b\boldsymbol{Y})_p = a\boldsymbol{X}_p + b\boldsymbol{Y}_p\,,\ \ \forall\, \boldsymbol{X}, \boldsymbol{Y} \in \mathfrak{X}(M)\,,\ a, b \in \mathbb{R}\,.$$

With these operations, $\mathfrak{X}(M)$ has a vector space structure. The null element is the vector field $\boldsymbol{O}\colon p \to \mathbf{o}_p$ which maps any point $p \in M$ to the null vector $\mathbf{o}_p$ of the tangent space T_pM.

Furthermore, the multiplication of a vector field $\boldsymbol{X} \in \mathfrak{X}(M)$ by a real-valued function $f \in C^\infty_M$ is the map $f\boldsymbol{X}\colon M \to TM$ such that

$$(f\boldsymbol{X})_p = f(p)\boldsymbol{X}_p\,, \;\; \forall\, p \in M\,. \tag{2.12}$$

It is worth noting that the point-wise operation here above is consistent with the vector space structure of T_pM, and therefore of the space $\mathfrak{X}(M)$. As a matter of fact, since f is in C^∞_M, evaluating the map $f\boldsymbol{X}$ at p through the relation (2.12) results in multiplying the vector $\boldsymbol{X}_p \in T_pM$ by the scalar $f(p)$. Consequently, $(f\boldsymbol{X})_p$ is itself an element of the tangent space T_pM and the map $f\boldsymbol{X}$ is effectively proven to be a vector field on M.

A smooth vector field $\boldsymbol{X} \in \mathfrak{X}(M)$ and a real-valued function $f \in C^\infty_M$ can be related in a way other than the multiplication.

In fact, $\boldsymbol{X}$ can be thought as an operator acting on f to give a new function $\boldsymbol{X}f\colon M \to \mathbb{R}$ defined by the following property:

$$(\boldsymbol{X}f)(p) = \boldsymbol{X}_p(f)\,, \;\; \forall\, p \in M\,. \tag{2.13}$$

We remark again that $f\boldsymbol{X} \in \mathfrak{X}(M)$ represents the vector field given by the multiplication of $\boldsymbol{X}$ by f. Instead, the notation $\boldsymbol{X}f \in C^\infty_M$ refers to the function resulting from applying $\boldsymbol{X}$ on f.

In summary, a smooth vector field $\boldsymbol{X}$ defines a map on C^∞_M as follows:

$$\begin{aligned} \boldsymbol{X}:\; & C^\infty_M \to C^\infty_M \\ & f \mapsto \boldsymbol{X}f\,. \end{aligned} \tag{2.14}$$

Such a map is clearly linear over $\mathbb{R}$, i.e. $\boldsymbol{X}(cf) = c(\boldsymbol{X}f)$ for an arbitrary scalar $c \in \mathbb{R}$. Actually, applying the property (2.13) at an arbitrary point $p \in M$, and exploiting the linearity of the tangent vector $\boldsymbol{X}_p$, one has

$$\boldsymbol{X}(cf)\,(p) = \boldsymbol{X}_p(cf) = c\boldsymbol{X}_p(f) = c(\boldsymbol{X}f)\,(p)\,.$$

Moreover, the map (2.14) satisfies Leibniz' rule:

$$\boldsymbol{X}(fg) = f(\boldsymbol{X}g) + g(\boldsymbol{X}f)\,, \quad \forall\, f, g \in C^\infty_M\,, \tag{2.15}$$

which can be proved by applying again the defining property (2.13) and recalling that $\boldsymbol{X}_p$ is a tangent vector fulfilling (2.4):

$$\begin{aligned} \boldsymbol{X}(fg)\,(p) &= \boldsymbol{X}_p(fg) = f(p)\boldsymbol{X}_p(g) + g(p)\boldsymbol{X}_p(f) \\ &= f(p)(\boldsymbol{X}g)(p) + g(p)(\boldsymbol{X}f)(p) \\ &= \big(f(\boldsymbol{X}g) + g(\boldsymbol{X}f)\big)(p)\,. \end{aligned}$$

More generally, any map in the form (2.14) which is linear over $\mathbb{R}$ and satisfying Leibniz' rule as in (2.15) is called a *derivation* of C^∞_M.

Recalling that C^∞_M has a vector space structure, the space of the derivations on C^∞_M, denoted as $\operatorname{Der} C^\infty_M$, is a subspace of the endomorphisms of C^∞_M:

$$\operatorname{Der} C^\infty_M = \{\, \boldsymbol{D} \in \operatorname{End} C^\infty_M \mid \boldsymbol{D}(fg) = f\boldsymbol{D}(g) + g\boldsymbol{D}(f)\,, \ \forall\, f, g \in C^\infty_M \,\}\,.$$

It is important to point out that the space $\operatorname{Der} C^\infty_M$ is formally distinct from $\mathfrak{X}(M)$. However, we have already shown that any smooth vector field $\boldsymbol{X}$ induces a derivation on C^∞_M, i.e. the map (2.14).

The converse also holds true, that is any derivation $\boldsymbol{D} \in \operatorname{Der} C^\infty_M$ induces a smooth vector field on M. Such a property easily comes out by considering the restriction of $\boldsymbol{D}$ at a point $p \in M$ as the map $\boldsymbol{D}_p \colon C^\infty_M \to \mathbb{R}$ satisfying

$$\boldsymbol{D}_p(f) = (\boldsymbol{D}f)(p)\,, \ \ \forall\, f \in C^\infty_M\,.$$

By exploiting Leibniz' product rule of $\boldsymbol{D} \in \operatorname{Der} C^\infty_M$, it is straightforward to verify that $\boldsymbol{D}_p$ is a derivation at p, in the sense of Definition 2.8. Then, $\boldsymbol{D}_p$ is a tangent vector at the point p of M and, as a map from M to TM, $\boldsymbol{D}$ is a vector field.

Because of this result, it is possible to identify the smooth vector fields on M with the derivations of C^∞_M, and $\boldsymbol{X}$ can be thought as an element of both $\mathfrak{X}(M)$ and $\operatorname{Der} C^\infty_M$.

2.3.1 Vector Fields and Smooth Maps

Suppose $F\colon M \to N$ is smooth map between the smooth manifolds M and N, and let $\mathrm{d}F$ be the differential.

Even if for every point $p \in M$ the specification $\mathrm{d}F|_p$ maps the tangent vector $\boldsymbol{X}_p \in T_pM$ to the vector $\mathrm{d}F|_p\boldsymbol{X}_p \in T_{F(p)}N$, such a map does not implies the definition of a vector field on N.

Specifically, if F is not injective, there may be distinct vectors $\mathrm{d}F|_p\boldsymbol{X}_p$ and $\mathrm{d}F|_q\boldsymbol{X}_q$, associated with distinct points p and q of M, at the same point $F(p) = F(q)$ of N. On the other hand, if F is not surjective, there is no way to assign the a tangent vector at the points of $N \setminus F(M)$.

The possibility to relate a vector field on M with a vector field on N through the differential of F leads to the following definition.

Definition 2.18. Let M and N be smooth manifolds and $F : M \to N$ a smooth map. The vector fields $\boldsymbol{X} \in \mathfrak{X}(M)$ and $\boldsymbol{Y} \in \mathfrak{X}(N)$ are said to be *F-related* if, for every point p of M, the tangent vector $\boldsymbol{X}_p$ is mapped to the tangent vector $\boldsymbol{Y}_{F(p)}$ through the differential of F at p:

$$\mathrm{d}F|_p\boldsymbol{X}_p = \boldsymbol{Y}_{F(p)}\,, \ \ \forall\, p \in M\,. \tag{2.16}$$

If f is a real-valued function on N, recalling the definition (2.5) and the property (2.13), the differential of F at an arbitrary point $p \in M$ is such that

$$(\mathrm{d}F|_p\boldsymbol{X}_p)(f) = \boldsymbol{X}_p(f \circ F) = \boldsymbol{X}(f \circ F)\,(p)\,,$$

and, at the same time, the vector field $\boldsymbol{Y} \in \mathfrak{X}(N)$ satisfies

$$\boldsymbol{Y}_{F(p)}(f) = (\boldsymbol{Y}f)\,\big(F(p)\big) = \big(\boldsymbol{Y}f \circ F\big)(p)\,,$$

By comparison, and because of the arbitrariness of $p \in M$, one infers that the condition $\mathrm{d}F|_p\boldsymbol{X}_p = \boldsymbol{Y}_{F(p)}$ is the same as

$$\boldsymbol{X}(f \circ F) = \boldsymbol{Y}f \circ F\,, \ \ \forall\, f \in C^\infty_N\,, \tag{2.17}$$

or, equivalently, we can say that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are F-related if, and only if, the relation here above is satisfied.

A further important property of F-related vector fields is expressed in the following proposition.

Proposition 2.19. *Let $F: M \to N$ be a smooth map between the smooth manifolds M and N. Then, if F is a diffeomorphism, for every $\boldsymbol{X} \in \mathfrak{X}(M)$ there exists a unique vector field $\boldsymbol{Y} \in \mathfrak{X}(N)$ which is F-related with $\boldsymbol{X}$.*

Proof. Let $(p, \boldsymbol{X}_p)$ be the element of the tangent bundle TM given by the a vector field $\boldsymbol{X} \in \mathfrak{X}(M)$ at a point $p \in M$. Applying the map (2.8), one obtains the image through the global differential $\mathrm{d}F$ as

$$\big(F(p), \mathrm{d}F|_p \boldsymbol{X}_p\big) = \mathrm{d}F(p, \boldsymbol{X}_p) \in TN\,.$$

Assuming F is a diffeomorphism, the differential $\mathrm{d}F: TM \to TN$ is a diffeomorphism itself. Consequently, $(p, \boldsymbol{X}_p)$ is the preimage of an element $(q, \boldsymbol{Y}_q)$ of the tangent bundle TN, that is

$$(q, \boldsymbol{Y}_q) = \big(F(p), \mathrm{d}F|_p \boldsymbol{X}_p\big) \in TN\,,$$

which, being $p = F^{-1}(q)$, can be concisely represented by the tangent vector at the point $q \in N$ as

$$\boldsymbol{Y}_q = \mathrm{d}F|_{F^{-1}(q)} \boldsymbol{X}_{F^{-1}(q)}\,.$$

The expression here above can be seen as the specialization at q of the vector field $\boldsymbol{Y}: N \to TN$ resulting from the following composition:

$$\boldsymbol{Y} = \mathrm{d}F \circ \boldsymbol{X} \circ F^{-1}\,.$$

Hence, the vector field $\boldsymbol{Y} \in \mathfrak{X}(N)$ is F-related to $\boldsymbol{X}$ by construction and, since F and $\mathrm{d}F$ are bijective, it is unique. □

In the case that $F: M \to N$ is a diffeomorphism, the unique vector field which is F-related to $\boldsymbol{X}$ is usually denoted as $F_* \boldsymbol{X}$ and is called the *pushforward* of $\boldsymbol{X}$ by F.

As emerged from the proof of Proposition 2.19, the pushforward $F_*\boldsymbol{X}$ is explicitly defined by the following relation:

$$(F_*\boldsymbol{X})_q = \mathrm{d}F|_{F^{-1}(q)}\boldsymbol{X}_{F^{-1}(q)}\,, \ \ \forall\, q \in N\,. \tag{2.18}$$

2.3.2 Lie Algebra of Vector Fields

Suppose two vector fields $\boldsymbol{X}$ and $\boldsymbol{Y}$ are considered on a smooth manifold M. Since applying $\boldsymbol{X}$ on a smooth function f gives a further smooth function $\boldsymbol{X}f$, it makes sense to apply, in turn, the vector field $\boldsymbol{Y}$. However, the resulting operation $\boldsymbol{Y}\boldsymbol{X} : f \mapsto \boldsymbol{Y}(\boldsymbol{X}f)$ does not fulfill Leibniz' product rule and the composition $\boldsymbol{Y}\boldsymbol{X}$ does not provide a vector field.

Even if $\boldsymbol{Y}\boldsymbol{X}$ is not a vector field, as well as the composition $\boldsymbol{X}\boldsymbol{Y}$, they can be combined to give a further operator on C^∞_M.

Definition 2.20. Let M be a smooth map and $\boldsymbol{X}$ and $\boldsymbol{Y}$ two vector fields. The *Lie bracket* of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is the operator $[\boldsymbol{X},\boldsymbol{Y}]\colon C^\infty_M \to C^\infty_M$ such that

$$[\boldsymbol{X},\boldsymbol{Y}]f = \boldsymbol{X}\boldsymbol{Y}f - \boldsymbol{Y}\boldsymbol{X}f\,, \ \ \forall\, f \in C^\infty_M\,. \tag{2.19}$$

The important feature of the Lie bracket is to provide a vector field on the manifold M, that is a linear map on C^∞_M fulfilling Leibniz' product rule.

Such a property can be easily proved by direct computation for two arbitrary functions $f, g \in C^\infty_M$:

$$\begin{aligned}
[\boldsymbol{X},\boldsymbol{Y}](fg) &= \boldsymbol{X}\big(\boldsymbol{Y}(fg)\big) - \boldsymbol{Y}\big(\boldsymbol{X}(fg)\big)\\
&= \boldsymbol{X}\big(f(\boldsymbol{Y}g) + g(\boldsymbol{Y}f)\big) - \boldsymbol{Y}\big(f(\boldsymbol{X}g) + g(\boldsymbol{X}f)\big)\\
&= f(\boldsymbol{X}\boldsymbol{Y}g) + (\boldsymbol{X}f)(\boldsymbol{Y}g) + g(\boldsymbol{X}\boldsymbol{Y}f) + (\boldsymbol{X}g)(\boldsymbol{Y}f)\\
&\quad - f(\boldsymbol{Y}\boldsymbol{X}g) - (\boldsymbol{Y}f)(\boldsymbol{X}g) - g(\boldsymbol{Y}\boldsymbol{X}f) - (\boldsymbol{Y}g)(\boldsymbol{X}f)\\
&= f(\boldsymbol{X}\boldsymbol{Y}g) - f(\boldsymbol{Y}\boldsymbol{X}g) + g(\boldsymbol{X}\boldsymbol{Y}f) - g(\boldsymbol{Y}\boldsymbol{X}f)\\
&= f[\boldsymbol{X},\boldsymbol{Y}]g + g[\boldsymbol{X},\boldsymbol{Y}]f\,.
\end{aligned}$$

The next proposition summarizes some useful properties of the Lie bracket.

Proposition 2.21. *Let* $\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z} \in \mathfrak{X}(M)$ *be vector fields on the smooth manifold* M*. Then, the Lie bracket fulfills the following properties.*

1. *Bilinearity:*

$$[a\boldsymbol{X} + b\boldsymbol{Y}, \boldsymbol{Z}] = a[\boldsymbol{X}, \boldsymbol{Z}] + b[\boldsymbol{Y}, \boldsymbol{Z}]\,, \ \ \forall\, a, b \in \mathbb{R}\,,$$
$$[\boldsymbol{Z}, a\boldsymbol{X} + b\boldsymbol{Y}] = a[\boldsymbol{Z}, \boldsymbol{X}] + b[\boldsymbol{Z}, \boldsymbol{Y}]\,, \ \ \forall\, a, b \in \mathbb{R}\,.$$

2. *Antisymmetry:*

$$[\boldsymbol{X}, \boldsymbol{Y}] = -[\boldsymbol{Y}, \boldsymbol{X}]\,.$$

3. *Jacobi identity:*

$$\big[\boldsymbol{X}, [\boldsymbol{Y}, \boldsymbol{Z}]\big] + \big[\boldsymbol{Y}, [\boldsymbol{Z}, \boldsymbol{X}]\big] + \big[\boldsymbol{Z}, [\boldsymbol{X}, \boldsymbol{Y}]\big] = \boldsymbol{O}\,.$$

Proof. The properties can be verified by applying the defining property (2.19) for an arbitrary $f \in C^\infty_M$. An explicit computation can be found, e.g., in Lee (2012). □

Given the smooth manifold M, it is trivial to see that the space of vector fields $\mathfrak{X}(M)$, along with the operation defined by (2.19), is a Lie algebra over the real field $\mathbb{R}$ consistent with Definition A.27.

In addition, the Lie bracket of vector fields does have some specific properties, as shown here below.

Proposition 2.22. *Let $F\colon M \to N$ be a smooth map between the manifolds M and N and let $\boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathfrak{X}(M)$ and $\boldsymbol{Y}_1, \boldsymbol{Y}_2 \in \mathfrak{X}(N)$ be vector fields such that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are F-related to $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, respectively. Then, $[\boldsymbol{X}_1, \boldsymbol{X}_2]$ is F-related to $[\boldsymbol{Y}_1, \boldsymbol{Y}_2]$.*

Proof. Consider an arbitrary real-valued function f on N and apply the property (2.17) of the F-related vector fields. Since the composition of maps distributes over addition, one has

$$\begin{aligned}(\boldsymbol{X}_1\boldsymbol{X}_2 - \boldsymbol{X}_2\boldsymbol{X}_1)(f \circ F) &= \boldsymbol{X}_1(\boldsymbol{Y}_2 f \circ F) - \boldsymbol{X}_2(\boldsymbol{Y}_1 f \circ F)\\ &= (\boldsymbol{Y}_1\boldsymbol{Y}_2 f) \circ F - (\boldsymbol{Y}_2\boldsymbol{Y}_1 f) \circ F\\ &= \big((\boldsymbol{Y}_1\boldsymbol{Y}_2 - \boldsymbol{Y}_2\boldsymbol{Y}_1) f\big) \circ F\,,\end{aligned}$$

which is concisely written as

$$[\boldsymbol{X}_1, \boldsymbol{X}_2](f \circ F) = \big([\boldsymbol{Y}_1, \boldsymbol{Y}_2]f\big) \circ F\,, \ \ \forall\, f \in C_N^\infty\,. \qquad \square$$

A straightforward corollary concerns the special case of a diffeomorphism.

Corollary 2.23. *Let $F\colon M \to N$ be a diffeomorphism between the smooth manifolds M and N. Then, the pushforward of the Lie bracket of vector fields is the Lie bracket of the pushforwards of the vector fields:*

$$F_*[\boldsymbol{X}_1, \boldsymbol{X}_2] = [F_*\boldsymbol{X}_1, F_*\boldsymbol{X}_2]\,, \ \ \forall\, \boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathfrak{X}(M)\,. \tag{2.20}$$

Proof. We recall that if F is a diffeomorphism, the pushforward of $\boldsymbol{X}_1$ by F is the only vector field which is F-related to $\boldsymbol{X}_1$, and the same applies to $\boldsymbol{X}_2$.

Then, the proof promptly comes from Proposition 2.22 with the conditions $\boldsymbol{Y}_1 = F_*\boldsymbol{X}_1$ and $\boldsymbol{Y}_2 = F_*\boldsymbol{X}_2$. $\square$

2.3.3 Integral Curves

When a smooth curve $\gamma : J \to M$ is given on a smooth manifold M, the velocity vector $\gamma'(\mu)$ is a vector of the tangent space $T_{\gamma(\mu)}M$.

Conversely, one may be interested in finding a curve whose velocity vector at each point is equal to an assigned tangent vector.

Definition 2.24. Let M a smooth vector field and $\boldsymbol{V}\colon M \to TM$ a vector field. An *integral curve* of $\boldsymbol{V}$ is a differentiable curve $\gamma : J \to M$ whose velocity at each point is equal to the value of $\boldsymbol{V}$ at that point:

$$\gamma'(\mu) = \boldsymbol{V}_{\gamma(\mu)}\,, \ \ \forall\, \mu \in J\,. \tag{2.21}$$

A vector field $\boldsymbol{V}$ is said to be *complete* when the domain of the integral curves can be chosen to be the whole real field, i.e. $J = \mathbb{R}$.

Please recall that on a smooth coordinate domain $U \subseteq M$, the curve γ is represented in local coordinates by the real-valued functions $\gamma^1, \ldots, \gamma^n$, so that the velocity vector is expressed as $\gamma'(\mu) = {\gamma^i}'(\mu)\,\partial_i|_{\gamma(\mu)}$. At the same time, the vector field $\boldsymbol{V}$ is represented by the component functions $V^1, \dots, V^n$ with respect to the basis $\{\partial_1, \dots, \partial_n\}$.

Then, the condition (2.21) is expressed in local representation as

$$\gamma^{i\prime}(\mu)\,\frac{\partial}{\partial x^i}\bigg|_{\gamma(\mu)} = V^i\big(\gamma(\mu)\big)\,\frac{\partial}{\partial x^i}\bigg|_{\gamma(\mu)},$$

and finding the integral curve of $\boldsymbol{V}$ actually means solving the following system of ordinary differential equations:

$$\begin{aligned} \gamma^{1\prime}(\mu) &= V^1\big(\gamma^1(\mu), \dots, \gamma^n(\mu)\big)\,, \\ &\vdots \\ \gamma^{n\prime}(\mu) &= V^n\big(\gamma^1(\mu), \dots, \gamma^n(\mu)\big)\,. \end{aligned}$$

Proposition 2.25. *Let $\boldsymbol{V} \colon M \to TM$ be a smooth vector field on a smooth manifold M and let $\gamma \colon J \to M$ be an integral curve of $\boldsymbol{V}$. If $\tau \colon \tilde{J} \to J$ is a differentiable real-valued function, with $\tilde{J} = \{\, \mu \in \mathbb{R} \mid \tau(\mu) \in J \,\}$, the curve $\tilde{\gamma} \colon \tilde{J} \to M$, defined by $\tilde{\gamma}(\mu) = \gamma\big(\tau(\mu)\big)$, is an integral curve of the vector field $\tau'\boldsymbol{V}$.*

Proof. The proof readily comes from Proposition 2.13 and applying the relation (2.11) at any $\mu \in \tilde{J}$:

$$\tilde{\gamma}'(\mu) = \tau'(\mu)\gamma'\big(\tau(\mu)\big) = \tau'(\mu)\boldsymbol{V}_{\tilde{\gamma}(\mu)}\,.$$

Since the relation holds true for all μ in $\tilde{J}$, we conclude that $\tilde{\gamma}$ is the integral curve of the vector field provided by the multiplication of $\boldsymbol{V}$ by the function τ'. □

The following proposition shows how the integral curves of related vector fields are mapped the one to the other.

Proposition 2.26. *Consider a smooth map $F \colon M \to N$ between the smooth manifolds M and N. Then, the vector fields $\boldsymbol{X} \in \mathfrak{X}(M)$ and $\boldsymbol{Y} \in \mathfrak{X}(N)$ are F-related if, and only if, F takes integral curves of $\boldsymbol{X}$ to integral curves of $\boldsymbol{Y}$.*

Proof. Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ are F-related and let $\gamma : J \to M$ be an integral curve of $\boldsymbol{X}$. The composition $F \circ \gamma \colon J \to N$ is a curve of N whose velocity satisfies

$$(F \circ \gamma)'(\mu) = \mathrm{d}F|_{\gamma(\mu)}\gamma'(\mu) = \mathrm{d}F|_{\gamma(\mu)}\boldsymbol{X}_{\gamma(\mu)} = \boldsymbol{Y}_{F(\gamma(\mu))} = \boldsymbol{Y}_{(F\circ\gamma)(\mu)}\,,$$

that is $F \circ \gamma$ is the integral curve of $\boldsymbol{Y}$.

Conversely, suppose now F takes integral curves of $\boldsymbol{X}$ to integral curves of $\boldsymbol{Y}$. For an arbitrary point $p \in M$, consider $\varepsilon > 0$ such that $\gamma\colon (-\varepsilon, \varepsilon) \to M$ is an integral curve of $\boldsymbol{X}$ starting at p.

Since $F \circ \gamma$ is an integral curve of $\boldsymbol{Y}$ starting at $F(p)$, one has:

$$\boldsymbol{Y}_{F(p)} = (F \circ \gamma)'(0) = \mathrm{d}F|_p\gamma'(0) = \mathrm{d}F|_p\boldsymbol{X}_p\,,$$

whence, for the arbitrariness of the point $p \in M$, the vector fields $\boldsymbol{X}$ and $\boldsymbol{Y}$ are F-related. □

2.3.4 Flows

Let $\boldsymbol{V} \in \mathfrak{X}(M)$ be a complete smooth vector field on M. When a point $p \in M$ is fixed on M, let us suppose $\theta^{(p)}\colon \mathbb{R} \to M$ is the integral curve of $\boldsymbol{V}$ starting at p.

For each $\mu \in \mathbb{R}$, the point $\theta^{(p)}(\mu)$ of such a curve can also be seen as the image of the following map:

$$\begin{aligned} \theta_\mu :\ & M \to M \\ & p \mapsto \theta_\mu(p) = \theta^{(p)}(\mu)\,. \end{aligned} \tag{2.22}$$

Let q be the point of the curve $\theta^{(p)}$ relevant to $\lambda \in \mathbb{R}$, that is $q = \theta^{(p)}(\lambda)$, and define a further map $\theta^{(q)}\colon \mathbb{R} \to M$ such that $\theta^{(q)}(\mu) = \theta^{(p)}(\mu+\lambda)$. Then, since the function $\mu \mapsto \mu + \lambda$ is a translation on $\mathbb{R}$, by Proposition 2.25 the map $\theta^{(q)}$ is the integral curve of the same vector field $\boldsymbol{V}$ as $\theta^{(p)}$.

As a matter of fact, $\theta^{(p)}$ and $\theta^{(q)}$ represent different parameterizations of the same curve on M, so that the map (2.22) applies as follows

$$\theta_{\mu+\lambda}(p) = \theta^{(p)}(\mu + \lambda) = \theta^{(q)}(\mu) = \theta_\mu(q) = \theta_\mu\big(\theta_\lambda(p)\big)\,,$$

that is

$$\theta_{\mu+\lambda}(p) = (\theta_\mu \circ \theta_\lambda)(p)\,.$$

The relation here above, along with the property $\theta_0(p) = p$, assures that θ_μ is the specialization at μ of the map $\theta\colon \mathbb{R} \times M \to M$ representing the action of the additive group $\mathbb{R}$ on M (cf. Appendix A.2.2), which allows one to introduce the following definition.

Definition 2.27. Let M be a smooth manifold. A *global flow* on M, also called a *one-parameter group action*, is a continuous left $\mathbb{R}$-action on M.

Explicitly, a global flow θ on M is the map

$$\begin{aligned} \theta:\ &\mathbb{R} \times M \to M \\ &(\mu, p) \mapsto \theta(\mu, p) = \theta_\mu(p)\,, \end{aligned} \tag{2.23}$$

where, for each $\mu \in \mathbb{R}$, the map $\theta_\mu\colon M \to M$ is a bijection of M which is consistent with the group structure of $\mathbb{R}$:

- $\theta_0 = \mathrm{id}_M$;
- $\theta_{\mu+\lambda} = \theta_\mu \circ \theta_\lambda$.

Please notice that if the global flow θ is a continuous group action, the induced map θ_μ is a homeomorphism. Also, when θ is smooth, the map θ_μ is a diffeomorphism.

At the same time, for each $p \in M$, the global flow θ induces a curve $\theta^{(p)}\colon \mathbb{R} \to M$ defined by $\theta^{(p)}(\mu) = \theta(\mu, p)$. The image of such a curve is the orbit of $p \in M$ under the group action.

If $\theta : \mathbb{R} \times M \to M$ is a smooth global flow, and $\theta^{(p)} : \mu \to M$ is the associated curve starting at p, the velocity of $\theta^{(p)}$ at $\mu = 0$ is a tangent vector $\boldsymbol{V}_p \in T_pM$.

Consequently, the map

$$\begin{aligned} \boldsymbol{V}:\ &M \to TM \\ &p \mapsto \boldsymbol{V}_p = {\theta^{(p)}}'(0)\,, \end{aligned} \tag{2.24}$$

is a vector field on M, which is called the *infinitesimal generator* of θ.

It is possible to verify that the infinitesimal generator $\boldsymbol{V}$ of a smooth global flow θ is a smooth vector field on M, and that each curve $\theta^{(p)}$ is an integral curve of $\boldsymbol{V}$ (see, e.g., Lee (2012)).

Remark. The notion of a smooth global flow has been introduced by referring to a complete smooth vector field. In practice, the global flow has been constructed assuming that the integral curves of the smooth vector field are defined for all $\mu \in \mathbb{R}$.

However, such an assumption does not hold true for any vector field on a manifold, so that it is worth introducing the notion of a flow which preserves, at least locally, the same properties as a global one.

To this end, we define a *flow domain* for M as an open subset $D \subseteq \mathbb{R} \times M$ such that, for each $p \in M$, the set $D^{(p)} = \{ \mu \in \mathbb{R} \mid (\mu, p) \in D \}$ is an open interval containing 0.

Therefore, a *local flow*, or simply a *flow*, on M is a map $\theta \colon D \to M$ with the same definition as in (2.23) restricted to a flow domain D.

With this restriction, the identification between the curves of a flow and the integral curves of a smooth vector field still holds true (see, e.g., Lee (2012)).

2.3.5 Lie Derivative of a Vector Field

The tangent vector $\mathbf{v}$ at a point of a smooth manifold M, introduced by Definition 2.8, provides a generalization of the notion of directional derivative of a real-valued function f. Moreover, as shown by (2.9), any tangent vector $\mathbf{v}$ can be seen as the velocity of a curve γ, so that the action of $\mathbf{v}$ on f gives the derivative of f along γ, in the ordinary sense.

In addition, when a vector field $\boldsymbol{X}$ is considered on M, the function $\boldsymbol{X}f$, defined at any point of p by virtue of the identification (2.13), is the derivative of f in the direction of $\boldsymbol{X}$.

On this basis, it is natural to ask if a similar generalization applies to the derivative of a vector field on M. Such a role is played by the Lie derivative.

Definition 2.28. Let M be a smooth manifold with a vector field $\boldsymbol{X} \in \mathfrak{X}(M)$. The *Lie derivative* of a vector field $\boldsymbol{Y} \in \mathfrak{X}(M)$ with respect to $\boldsymbol{X}$, at a point p, is defined as

$$(\mathscr{L}_{X}\boldsymbol{Y})_p = \lim_{\mu \to 0} \frac{(\theta_{-\mu})_*\big|_{\theta_\mu(p)} \boldsymbol{Y}_{\theta_\mu(p)} - \boldsymbol{Y}_p}{\mu}\,, \tag{2.25}$$

where θ is the flow of $\boldsymbol{X}$.

Please observe that the numerator of the limit in (2.25) represents the variation of the vector field $\boldsymbol{Y}$ in the direction of $\boldsymbol{X}$. However, in order to make completely clear the above definition, such a variation can be interpreted as follows.

First, the vector field $\boldsymbol{Y}$ is evaluated at $\theta_\mu(p) = \theta^{(p)}(\mu)$, which is the point of the integral curve of $\boldsymbol{X}$ starting at p relevant to μ. Since the vector $\boldsymbol{Y}_{\theta_\mu(p)}$ belongs to the tangent space $T_{\theta_\mu(p)}M$, it cannot be directly compared with $\boldsymbol{Y}_p$, but one should consider the pushforward by $\theta_{-\mu} = \theta_\mu^{-1}$.

Then, differential $(\theta_{-\mu})_*\big|_{\theta_\mu(p)} : T_{\theta_\mu(p)}M \to T_pM$ is applied to the vector $\boldsymbol{Y}_{\theta_\mu(p)}$, so that the variation with respect to $\boldsymbol{Y}_p$ is actually evaluated in T_pM.

Since the limit in (2.25) is formally performed in T_pM, the Lie derivative $(\mathscr{L}_X\boldsymbol{Y})_p$ is clearly a tangent vector at p. Consequently, it is induced a vector field on M, denoted as $\mathscr{L}_X\boldsymbol{Y} : M \to TM$, which provides the Lie derivative of $\boldsymbol{Y}$ with respect of $\boldsymbol{X}$ at each point of M.

Even if the limit in (2.25) is well-defined, it is not used to evaluate the Lie derivative. On the contrary, a useful and simple formula, which does not require to find the flow of $\boldsymbol{X}$, can be applied.

The derivation of such a formula evaluating the Lie derivative is based on the following lemma (see, e.g. Boothby (2002)).

Lemma 2.29. *Let M be a smooth manifold with $\boldsymbol{X} \in \mathfrak{X}(M)$ the infinitesimal generator of a flow θ. Given a function $f \in C^\infty_M$ and a scalar $\varepsilon > 0$, there exists a C^∞ function $g : (-\varepsilon, \varepsilon) \times M \to \mathbb{R}$ such that*

$$f\big(\theta_\mu(p)\big) = f(p) + \mu g(\mu, p)\,, \quad \boldsymbol{X}_p f = g(0, p)\,.$$

Proof. Consider a function $h : (\mu, p) \mapsto h(\mu, p) = f\big(\theta_\mu(p)\big) - f(p)$, which is smooth on $(-\varepsilon, \varepsilon) \times M$ and satisfies $h(0,p) = 0$. Then, denoting by h' the derivative of h with respect to μ, for each fixed $p \in M$, the required map is

$$g(\mu, p) = \int_0^1 h'(\mu\lambda, p)\mathrm{d}\lambda\,.$$

Actually, by the fundamental theorem of calculus (Rudin, 1986), one has

$$\begin{aligned}\mu g(\mu, p) &= \int_0^1 h'(\mu\lambda, p)\mu\mathrm{d}\lambda = \int_0^\mu h'(\mu\lambda, p)\mathrm{d}(\mu\lambda)\\ &= h(\mu, p) - h(0, p) = h(\mu, p)\,,\end{aligned}$$

whence

$$f\big(\theta_\mu(p)\big) = f(p) + \mu g(\mu, p)\,.$$

On the other hand, if the vector field $\boldsymbol{X}$ is the infinitesimal generator of θ, recalling (2.24) and then using (2.10), one has

$$\boldsymbol{X}_p f = \left(\frac{\mathrm{d}}{\mathrm{d}\mu}\bigg|_0 \theta^{(p)}\right)(f) = \big(f \circ \theta^{(p)}\big)'(0) = \lim_{\mu\to 0} \frac{f\big(\theta^{(p)}(\mu)\big) - f\big(\theta^{(p)}(0)\big)}{\mu}\,,$$

that is

$$\boldsymbol{X}_p f = \lim_{\mu\to 0} \frac{h(\mu, p)}{\mu} = \lim_{\mu\to 0} g(\mu, p) = g(0, \mu)\,. \qquad \square$$

It is now possible to introduce the formula evaluating the Lie derivative.

Proposition 2.30. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be smooth vector fields on the smooth manifold M. Then, the Lie derivative of $\boldsymbol{Y}$ with respect to $\boldsymbol{X}$ is given by the Lie bracket of $\boldsymbol{X}$ and $\boldsymbol{Y}$:*

$$\mathscr{L}_X \boldsymbol{Y} = [\boldsymbol{X}, \boldsymbol{Y}]\,. \tag{2.26}$$

Proof. Consider an arbitrary point $p \in M$ and a function $f \in C^\infty_M$. Applying the defining property (2.5) to the differential $(\theta_{-\mu})_*\big|_{\theta_\mu(p)}$, one has

$$\big((\theta_{-\mu})_*\big|_{\theta_\mu(p)} \boldsymbol{Y}_{\theta_\mu(p)}\big)(f) = \boldsymbol{Y}_{\theta_\mu(p)}(f \circ \theta_{-\mu})\,. \tag{2.27}$$

Moreover, for some $\varepsilon > 0$, using Lemma 2.29 with $-\mu$ in place of μ, there exists a function $g\colon (-\varepsilon,\varepsilon)\times M\to\mathbb{R}$ such that

$$\big(f\circ\theta_\mu\big)(p)=f(p)-\mu g_{-\mu}(p)\,,$$

with $0<|\mu|<\varepsilon$ and $g_\mu(p)=g(\mu,p)$.

Consequently, the evaluation of the pushforward in (2.27) becomes

$$\big((\theta_{-\mu})_*\big|_{\theta_\mu(p)}\boldsymbol{Y}_{\theta_\mu(p)}\big)(f)=\boldsymbol{Y}_{\theta_\mu(p)}(f-\mu g_{-\mu})=(\boldsymbol{Y}f)\big(\theta_\mu(p)\big)-\mu\boldsymbol{Y}_{\theta_\mu(p)}(g_{-\mu})\,,$$

whence the Lie derivative defined by (2.25), applied at the function f, reads

$$\begin{aligned}(\mathscr{L}_{\boldsymbol{X}}\boldsymbol{Y})_p(f)&=\lim_{\mu\to0}\frac{\big((\theta_{-\mu})_*\big|_{\theta_\mu(p)}\boldsymbol{Y}_{\theta_\mu(p)}-\boldsymbol{Y}_p\big)(f)}{\mu}\\&=\lim_{\mu\to0}\frac{(\boldsymbol{Y}f)\big(\theta_\mu(p)\big)-\mu\boldsymbol{Y}_{\theta_\mu(p)}(g_{-\mu})-(\boldsymbol{Y}f)(p)}{\mu}\\&=\lim_{\mu\to0}\frac{\big((\boldsymbol{Y}f)\circ\theta^{(p)}\big)(\mu)-\big((\boldsymbol{Y}f)\circ\theta^{(p)}\big)(0)}{\mu}-\lim_{\mu\to0}\boldsymbol{Y}_{\theta^{(p)}(\mu)}(g_{-\mu})\,,\end{aligned}\tag{2.28}$$

where the property $\theta_\mu(p)=\theta^{(p)}(\mu)$ has been used, along with the initial condition $\theta^{(p)}(0)=p$.

The first limit in the sum (2.28) is the derivative at 0 of the real-valued function $(\boldsymbol{Y}f)\circ\theta^{(p)}:(-\varepsilon,\varepsilon)\to\mathbb{R}$, which, by (2.10) and then using the fact that $\boldsymbol{X}$ is the infinitesimal generator of the flow θ at p, is evaluated as follows:

$$\big((\boldsymbol{Y}f)\circ\theta^{(p)}\big)'(0)=\big(\theta^{(p)\prime}(0)\big)(\boldsymbol{Y}f)=\boldsymbol{X}_p(\boldsymbol{Y}f)\,.$$

At the same time, since by Lemma 2.29 $g_0(p)=g(0,p)$ coincides with $\boldsymbol{X}_p(f)=(\boldsymbol{X}f)(p)$, the function $g_{-\mu}$ tends to $\boldsymbol{X}f$ as μ approaches to 0, obtaining

$$\lim_{\mu\to0}\boldsymbol{Y}_{\theta_\mu(p)}(g_{-\mu})=\boldsymbol{Y}_p(\boldsymbol{X}f)$$

Then, the sum in (2.28) results

$$(\mathscr{L}_X \boldsymbol{Y})_p(f) = \boldsymbol{X}_p(\boldsymbol{Y}f) - \boldsymbol{Y}_p(\boldsymbol{X}f) = (\boldsymbol{X}\boldsymbol{Y})_p(f) - (\boldsymbol{Y}\boldsymbol{X})_p(f)$$
$$= (\boldsymbol{X}\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X})_p(f)\,,$$

which holds true for each $p \in M$, so that

$$(\mathscr{L}_X \boldsymbol{Y})(f) = (\boldsymbol{X}\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X})(f)\,,$$

whence, by Definition 2.20 and for the arbitrariness of f, one finally has

$$\mathscr{L}_X \boldsymbol{Y} = [\boldsymbol{X}, \boldsymbol{Y}]\,.$$ □

2.4 Lie Groups and Lie Algebras

In Section A.2.2 we introduced the notion of a group in the algebraic sense. When a group is also endowed with a smooth structure, we say that it is a Lie group.

Definition 2.31. A *Lie group* is a smooth manifold G which is also a group in the algebraic sense, with the property that the group composition and the inversion are both smooth.

The algebraic characterization of a group is given in Section A.2.2 and, within the context of smooth manifolds, it is required that the group multiplication $(g, h) \mapsto gh \in G$ and the inversion $g \mapsto g^{-1} \in G$ are both smooth.

As a consequence, the left translation L_g by any $g \in G$, defined by (A.5), is a diffeomorphism of G. In fact, not only L_g is smooth, because the group multiplication is, but the inverse $L_g^{-1} = L_{g^{-1}}$ is also smooth. Clearly, the same applies to the right translation R_g by g.

In addition, the conjugation C_g by g, defined by (A.6), is a Lie group automorphism. Actually, it is a bijective map of G preserving the group structure and, intended as the composition of the smooth maps L_g and $R_{g^{-1}}$, is also smooth.

More generally, in accordance with the adopted nomenclature, we say that a map $F : G \to H$, between the Lie groups G and H, is a *Lie group* *homomorphism* if it is smooth and at the same time is a group homomorphism.

We recall that when a group G is considered along with a set M, the action of G on M is the map $\alpha : G \times M \to M$ specified in A.2.2. If G is a Lie group and M is smooth manifold, we say that α is a *smooth action* if it is a smooth map.

Please notice that the permutation $\sigma_g : M \to M$, associated with the smooth action of the element $g \in G$ on M, is a diffeomorphism of M. Such a property readily results from the smoothness of the inverse map $\sigma_{g^{-1}}$.

2.4.1 The Lie Algebra of a Lie Group

Let G be a Lie group and let us consider a vector field $\boldsymbol{X} \in \mathfrak{X}(G)$. We say that $\boldsymbol{X}$ is *left-invariant* if it is invariant under all left translations L_g, meaning that it is L_g-related to itself for any $g \in G$.

More explicitly, denoting as $\mathrm{d}L_g : TG \to TG$ the global differential of the left translation by g, and recalling Definition 2.18, the vector field $\boldsymbol{X}$ is left-invariant if it satisfies the following identity:

$$\mathrm{d}L_g\big|_h \boldsymbol{X}_h = \boldsymbol{X}_{gh}\,, \ \ \forall\, g, h \in G\,, \tag{2.29}$$

where it has been exploited the property $\boldsymbol{X}_{L_g(h)} = \boldsymbol{X}_{gh}$.

Since L_g is a diffeomorphism of G, we properly say that $\boldsymbol{X}$ is a left-invariant vector field if it coincides with its pushforward by L_g, for any $g \in G$:

$$(L_g)_* \boldsymbol{X} = \boldsymbol{X}\,, \ \ \forall\, g \in G\,. \tag{2.30}$$

The set of all the left-invariant vector fields on G is denoted as $\mathfrak{X}_L(G)$ and is a linear subspace of $\mathfrak{X}(G)$.

In fact, recalling that the differential is a linear map (cf. Property 1 of Proposition 2.10), any linear combination $a\boldsymbol{X} + b\boldsymbol{Y}$ of the vector fields $\boldsymbol{X}, \boldsymbol{Y} \in \mathfrak{X}_L(M)$ is left-invariant:

$$(L_g)_*(a\boldsymbol{X} + b\boldsymbol{Y}) = a(L_g)_*\boldsymbol{X} + b(L_g)_*\boldsymbol{Y} = a\boldsymbol{X} + b\boldsymbol{Y}\,, \ \ \forall\, g \in G\,.$$

In addition, left-invariance is preserved when the Lie bracket is applied to any pair of left-invariant vector fields.

Proposition 2.32. *Let G be a Lie group and let $\mathfrak{X}_L(G)$ be the space of the left-invariant vector fields. Then $\mathfrak{X}_L(G)$ is closed under Lie bracket.*

Proof. Consider two arbitrary left-invariant vector fields $\boldsymbol{X}$ and $\boldsymbol{Y}$. Since the left translation L_g is a diffeomorphism for any $g \in G$, Corollary 2.23 applies:

$$(L_g)_*[\boldsymbol{X},\boldsymbol{Y}] = \big[(L_g)_*\boldsymbol{X},(L_g)_*\boldsymbol{Y}\big] = [\boldsymbol{X},\boldsymbol{Y}]\,,\ \ \forall\, g \in G\,,$$

that is $[\boldsymbol{X},\boldsymbol{Y}]$ is left-invariant. □

As a consequence of the above proposition, the set $\mathfrak{X}_L(G)$, along with the Lie bracket, has a Lie algebra structure and then it is a subalgebra of $\mathfrak{X}(G)$.

Specifically, the Lie algebra of all the smooth left-invariant vector fields on a Lie group G is called the *Lie algebra* of G and is denoted as $\mathsf{Lie}(G)$.

Proposition 2.33. *Let G be a Lie group, with the Lie algebra $\mathsf{Lie}(G)$, and let T_eG be the tangent space at the identity element. Then, the map*

$$\begin{aligned} \varepsilon:\ \mathsf{Lie}(G) &\to T_eG \\ \boldsymbol{X} &\mapsto \boldsymbol{X}_e\,, \end{aligned} \tag{2.31}$$

is a vector space isomorphism.

Proof. The linearity of the map ε over $\mathbb{R}$ is straightforward:

$$\varepsilon(a\boldsymbol{X}+b\boldsymbol{Y}) = (a\boldsymbol{X}+b\boldsymbol{Y})_e = a\boldsymbol{X}_e + b\boldsymbol{Y}_e = a\varepsilon(\boldsymbol{X}) + b\varepsilon(\boldsymbol{Y})\,,$$
$$\forall\, \boldsymbol{X},\boldsymbol{Y} \in \mathsf{Lie}(G)\,,\ a,b \in \mathbb{R}\,.$$

Moreover, for each $g \in G$ there exists a unique left translation L_g such that $L_g(e) = g$. Hence, given a vector $\mathbf{v} \in T_eG$, the differential of L_g at e uniquely determines the following map

$$\begin{aligned} \boldsymbol{X}:\ G &\to TG \\ g &\mapsto \boldsymbol{X}_g = \mathrm{d}L_g\big|_e\mathbf{v}\,, \end{aligned} \tag{2.32}$$

which clearly satisfies $\boldsymbol{X}_e = \mathbf{v}$.

Consequently, $\boldsymbol{X}$ is the unique vector field induced by the differential of the left translation such that its specialization $\boldsymbol{X}_e$ at the identity e coincides with the assigned vector $\mathbf{v} \in T_eG$.

Moreover, recalling Property 2 in Proposition 2.10, the vector field $\boldsymbol{X}$ defined by (2.32) satisfies the following relation

$$\mathrm{d}L_g\big|_h \boldsymbol{X}_h = \mathrm{d}L_g\big|_h\big(\mathrm{d}L_h\big|_e\mathbf{v}\big) = \mathrm{d}\big(L_g \circ L_h\big)\big|_e\mathbf{v} = \mathrm{d}L_{gh}\big|_e\mathbf{v} = \boldsymbol{X}_{gh}\,,$$

that is the fulfillment of the defining property (2.29) implies the left-invariance of $\boldsymbol{X}$.

In addition, by introducing a chart (U, φ) including a neighborhood of the identity e, it is possible to prove that the components of $\boldsymbol{X}$ in the local representation are C^∞ functions of the coordinates, which implies the smoothness of $\boldsymbol{X}$ (one can refer, e.g., to Boothby (2002), Lee (2012), Spivak (1999) for the details).

In conclusion, $\boldsymbol{X}$ is in $\mathsf{Lie}(G)$ and the uniqueness of the map (2.32), for each $\mathbf{v} \in T_eG$, assures that ε is one-to-one. □

When a homomorphism is defined between two Lie groups, the relevant Lie algebras are also related.

Proposition 2.34. *Let G and H be Lie groups with $\mathfrak{g}$ and $\mathfrak{h}$ as Lie algebras, respectively. If $F : G \to H$ is a Lie group homomorphism, for any $\boldsymbol{X} \in \mathfrak{g}$ there is a unique vector field $\boldsymbol{Y} \in \mathfrak{h}$ which is F-related to $\boldsymbol{X}$.*

Proof. As a Lie group homomorphism, F is consistent with group laws of G and H, implying

$$F\big(L_g(h)\big) = F(gh) = F(g)F(h) = L_{F(g)}\big(F(h)\big) \ \ \forall\, g, h \in G\,,$$

whence, for the arbitrariness of h, one finds

$$F \circ L_g = L_{F(g)} \circ F\,,\ \ \forall\, g \in G\,,$$

and the global differentials of the maps on both sides, by Property 1 in

Proposition 2.11, result

$$\mathrm{d}F \circ \mathrm{d}L_g = \mathrm{d}L_{F(g)} \circ \mathrm{d}F\,, \quad \forall\, g \in G\,.$$

Then, supposing $\boldsymbol{Y} \in \mathfrak{h}$ is the vector field on H identified by the tangent vector $\boldsymbol{Y}_e = \mathrm{d}F|_e \boldsymbol{X}_e$ at the identity, one has

$$\begin{aligned} \boldsymbol{Y}_{F(g)} &= \mathrm{d}L_{F(g)}\big|_e \boldsymbol{Y}_e = \mathrm{d}L_{F(g)}\big|_e \big(\mathrm{d}F|_e \boldsymbol{X}_e\big) = \big(\mathrm{d}L_{F(g)} \circ \mathrm{d}F\big)\big|_e \boldsymbol{X}_e \\ &= \big(dF \circ \mathrm{d}L_g\big)\big|_e \boldsymbol{X}_e = \mathrm{d}F|_{L_g(e)} \big(\mathrm{d}L_g\big|_e \boldsymbol{X}_e\big) = \mathrm{d}F|_g \boldsymbol{X}_g\,, \end{aligned}$$

where, by virtue of the left-invariance of both $\boldsymbol{X} \in \mathfrak{g}$ and $\boldsymbol{Y} \in \mathfrak{h}$, the identities $\boldsymbol{X}_g = \mathrm{d}L_g\big|_e \boldsymbol{X}_e$ and $\boldsymbol{Y}_{F(g)} = \mathrm{d}L_{F(g)}\big|_e \boldsymbol{Y}_e$ have been used.

In conclusion, the identity $\boldsymbol{Y}_{F(g)} = \mathrm{d}F|_g \boldsymbol{X}_g$, which holds for any $g \in G$, assures that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are F-related. □

The above proposition implies that for any Lie group homomorphism $F\colon G \to H$ there exists an *induced Lie algebra homomorphism* defined as

$$\begin{aligned} F_* :\ & \mathfrak{g} \to \mathfrak{h} \\ & \boldsymbol{X} \mapsto F_* \boldsymbol{X}\,, \end{aligned} \tag{2.33}$$

where $F_*\boldsymbol{X}$ is the unique left-invariant vector field on H which is F-related with $\boldsymbol{X}$.

Remark. If G is a subgroup of H, the inclusion $\iota\colon G \hookrightarrow H$ is group homomorphism and $\iota_*\colon \mathfrak{g} \to \mathfrak{h}$ is the induced Lie algebra homomorphism.

Specifically, the image $\iota_*(\mathfrak{g})$ is the subalgebra of $\mathfrak{h}$ consisting of the left-invariant vector fields on H which are ι-related with the vector fields on G, and are defined by the value at the identity given by $\mathrm{d}\iota|_e \boldsymbol{X}_e$, for some $\boldsymbol{X}_e \in T_eG$.

Since the differential $\mathrm{d}\iota|_e\colon T_eG \hookrightarrow T_eH$ is the inclusion of T_eG in T_eH, the subalgebra $\iota_*(\mathfrak{g})$ is characterized as

$$\iota_*(\mathfrak{g}) = \{\, \boldsymbol{X} \in \mathfrak{h} \mid \boldsymbol{X}_e \in T_eG \,\}\,.$$

2.4.2 One-parameter Subgroups and Exponential Map

Let us consider a Lie group G. Since $\mathbb{R}$ has a group structure under addition, the map $\gamma : \mathbb{R} \to G$ is a group homomorphism, which is called a *one-parameter subgroup* of G.

It is possible to prove that any left-invariant vector field on a Lie group is complete (see, e.g., Boothby (2002), Lee (2012)). Such a property assures that the integral curves of left-invariant vector fields are exactly the one-parameter subgroups of the Lie group, as stated by the proposition here below.

Proposition 2.35. *The one-parameter subgroups of a Lie group G are the integral curves of left-invariant vector fields starting at the identity.*

Proof. Let γ be the integral curve of some left-invariant vector field $\boldsymbol{X}$ starting at the identity, i.e. $\gamma(0) = e$ and $\gamma'(0) = \boldsymbol{X}_e$. Since $\boldsymbol{X}$ is complete, γ is defined on all of $\mathbb{R}$.

Moreover, recalling that $\boldsymbol{X}$ is L_g-related to itself, by Proposition 2.26 any left translation L_g takes integral curves of $\boldsymbol{X}$ to integral curves of $\boldsymbol{X}$. Consequently, setting $g = \gamma(\lambda)$ for some $\lambda \in \mathbb{R}$, the curve $\mu \mapsto L_{\gamma(\lambda)}\big(\gamma(\mu)\big) = \gamma(\lambda)\gamma(\mu)$ is an integral curve of $\boldsymbol{X}$ with the initial point at $\gamma(\lambda)$.

On the other hand, by Proposition 2.13 the map $\mu \mapsto \gamma(\lambda + \mu)$, is itself an integral curve of $\boldsymbol{X}$ starting at $\gamma(\lambda)$, so that one has

$$\gamma(\lambda)\gamma(\mu) = \gamma(\lambda + \mu)\,,$$

which means that γ preserves the group structure of $\mathbb{R}$ and G, resulting a one-parameter subgroup of G.

Conversely, suppose $\gamma\colon \mathbb{R} \to G$ is a one parameter subgroup of G. Since γ is a Lie group homomorphism, its value at 0 is the identity of G, that is $\gamma(0) = e$.

In addition, treating $\mathrm{d}/\mathrm{d}\mu$ as a left-invariant vector field of $\mathbb{R}$, it is mapped by the differential $\mathrm{d}\gamma$ to a vector field $\boldsymbol{X} = \mathrm{d}\gamma(\mathrm{d}/\mathrm{d}\mu)$. Since $\boldsymbol{X}$ is the unique left-invariant vector field of G which is γ-related to $\mathrm{d}/\mathrm{d}\mu$ (cf.

Proposition 2.34), for any $\mu_0 \in \mathbb{R}$ one has

$$\gamma'(\mu_0) = \mathrm{d}\gamma|_{\mu_0} \frac{\mathrm{d}}{\mathrm{d}\mu}\bigg|_{\mu_0} = \boldsymbol{X}_{\gamma(\mu_0)} \,,$$

that is, consistently with Definition 2.24, γ is the integral curve of $\boldsymbol{X}$ starting at the identity e. □

By identifying a one-parameter subgroup $\gamma : \mathbb{R} \to G$ with the integral curve of a vector field $\boldsymbol{X} \in \mathsf{Lie}(G)$, it is established a one-to-one correspondence between the tangent space at the identity, the Lie algebra of G the set of the one-parameter subgroups of G.

In fact, any tangent vector $\mathbf{v} \in T_eG$ uniquely identifies a left-invariant vector-field $\boldsymbol{X} \in \mathsf{Lie}(G)$ such that $\boldsymbol{X}_e = \mathbf{v}$, as well as a one-parameter subgroup γ of G satisfying $\gamma'(0) = \mathbf{v}$.

For this reason, we say that the one-parameter subgroup representing the integral curve of $\boldsymbol{X}$ starting at the identity, is *generated* by $\boldsymbol{X}$.

In addition, if $\gamma : \mathbb{R} \to G$ is the one-parameter subgroup generated by the left-invariant vector field $\boldsymbol{X}$, the global flow $\theta : \mathbb{R} \times G \to G$ associated with $\boldsymbol{X}$ is defined by

$$\theta(\mu, g) = L_g\big(\gamma(\mu)\big) = R_{\gamma(\mu)}(g) = g\gamma(\mu)\,. \tag{2.34}$$

Actually, since $\boldsymbol{X}$ is L_g-related to itself, by Proposition 2.26 the integral curves of $\boldsymbol{X}$ are translated the ones into the others by L_g.

Specifically, noting that $\gamma(0) = e$, the map $\theta^{(g)} : \mathbb{R} \to G$, defined by $\theta^{(g)}(\mu) = \theta(\mu, g) = g\gamma(\mu)$, is a curve starting at g and its velocity is

$$\begin{aligned}\theta^{(g)\prime}(0) &= \mathrm{d}\theta^{(g)}\big|_0 \frac{\mathrm{d}}{\mathrm{d}\mu}\bigg|_0 = \mathrm{d}\big(L_g \circ \gamma\big)\big|_0 \frac{\mathrm{d}}{\mathrm{d}\mu}\bigg|_0 = \big(\mathrm{d}L_g\big|_{\gamma(0)} \circ \mathrm{d}\gamma|_0\big) \frac{\mathrm{d}}{\mathrm{d}\mu}\bigg|_0 \\ &= \mathrm{d}L_g\big|_e \gamma'(0) = \mathrm{d}L_g\big|_e \boldsymbol{X}_e = \boldsymbol{X}_g\,,\end{aligned}$$

that is $\theta^{(g)} = g\gamma$ is the integral curve of $\boldsymbol{X}$ starting at g.

The connection between the left-invariant vector fields and the one-parameter subgroups is the basis of the definition of the exponential map.

Definition 2.36. Let G be a Lie group and $\mathfrak{g}$ the relevant Lie algebra. The *exponential map* of G is the map $\exp\colon \mathfrak{g} \to G$ such that

$$\exp(\boldsymbol{X}) = \gamma(1)\,, \ \ \forall\, \boldsymbol{X} \in \mathfrak{g}\,, \tag{2.35}$$

where $\gamma\colon \mathbb{R} \to G$ is the one-parameter subgroup of G generated by $\boldsymbol{X}$.

The importance of the exponential map in characterizing a Lie group relies on the fact that it maps the straight line of $\mathfrak{g}$ passing through $\boldsymbol{X}$ to the one-parameter subgroup generated by $\boldsymbol{X}$ or, which is the same, to the integral curve of $\boldsymbol{X}$ passing through the identity e. Such a property is discussed in the following proposition.

Proposition 2.37. *Let $\mathfrak{g}$ be the Lie algebra of the Lie group G. For any $\boldsymbol{X} \in \mathfrak{g}$, the curve $\gamma : \mathbb{R} \to G$ defined by $\gamma(\lambda) = \exp(\lambda\boldsymbol{X})$ is the one-parameter subgroup generated by $\boldsymbol{X}$.*

Proof. Let $\gamma : \mathbb{R} \to G$ be one-parameter subgroup generated by $\boldsymbol{X}$ and consider the real-valued function $\tau\colon \mathbb{R} \to \mathbb{R}$ defined by $\tau(\mu) = \lambda\mu$, for any $\lambda \in \mathbb{R}$, so that the composition $\gamma \circ \tau\colon \mathbb{R} \to G$ is the integral curve of the vector field $\lambda\boldsymbol{X}$ (cf. Proposition 2.25).

Since $\mathfrak{g}$ has a vector space structure, the scalar multiplication $\lambda\boldsymbol{X}$ is itself a left-invariant vector field of G and the exponential map provides

$$\exp(\lambda\boldsymbol{X}) = \gamma(\tau(1)) = \gamma(\lambda)\,, \ \ \forall\, \lambda \in \mathbb{R}\,. \qquad \square$$

Since the exponential map defines a one-parameter subgroup of G, which preserves the group structure of $\mathbb{R}$ and G, the addition in $\mathbb{R}$ is reflected in the composition law of G:

$$\exp\big((\lambda + \mu)\boldsymbol{X}\big) = \exp(\lambda\boldsymbol{X})\exp(\mu\boldsymbol{X})\,, \ \ \forall\, \boldsymbol{X} \in \mathfrak{g}\,, \ \lambda, \mu \in \mathbb{R}\,, \tag{2.36}$$

as well as the opposite in $\mathbb{R}$ is consistent with the inverse in G:

$$\exp(-\boldsymbol{X}) = \big(\exp(\boldsymbol{X})\big)^{-1}\,, \ \ \forall\, \boldsymbol{X} \in \mathfrak{g}\,. \tag{2.37}$$

Finally, since the integral curve of $\boldsymbol{X}$ starting at e is expressed in terms of exponential map as $\mu \mapsto \exp(\mu\boldsymbol{X})$, the flow of $\boldsymbol{X}$ provided by (2.34) specializes to

$$\theta(\mu, g) = L_g\big(\exp(\mu \boldsymbol{X})\big) = R_{\exp(\mu \boldsymbol{X})}(g) = g \exp(\mu \boldsymbol{X}), \tag{2.38}$$

where $\mu \mapsto g \exp(\mu \boldsymbol{X})$ is the integral curve of $\boldsymbol{X}$ starting at g.

Let us further observe that if $\sigma \colon \mathbb{R} \mapsto \mathfrak{g}$ is the curve defined by $\sigma(t) = t\boldsymbol{X}$, satisfying $\sigma'(0) = \boldsymbol{X}$, the velocity of the composition $t \mapsto (\exp \circ \sigma)(t)$ at $t = 0$ results

$$(\exp \circ \sigma)'(0) = \mathrm{d}\exp|_{\sigma(0)} \sigma'(0) = \mathrm{d}\exp|_{O} \boldsymbol{X}.$$

At the same time, by Proposition (2.37), the map $t \mapsto \exp(t\boldsymbol{X})$ is also the one-parameter subgroup generated by $\boldsymbol{X}$, whose tangent vector at the identity is $\boldsymbol{X}_e \in T_e G$ (cf. Proposition 2.35). Then, one has

$$\boldsymbol{X}_e = \left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_0 \exp(t\boldsymbol{X}) = \mathrm{d}\exp|_{O} \boldsymbol{X},$$

and, under the canonical identification of both $T_e G$ and $T_O \mathfrak{g}$ with $\mathfrak{g}$, the differential $\mathrm{d}\exp|_O$ is the identity map:

$$\mathrm{d}\exp|_O = \mathrm{id}_{\mathfrak{g}}. \tag{2.39}$$

Remark. The consistency of the exponential map with the group operations only concerns the group homomorphism between $\mathbb{R}$ and G, and it does not apply to the operations of the Lie algebra $\mathfrak{g}$.

This means, for example, that the exponential of the sum $\boldsymbol{X} + \boldsymbol{Y}$ is not the same, in general, as the product of $\exp(\boldsymbol{X})$ and $\exp(\boldsymbol{Y})$:

$$\exp(\boldsymbol{X} + \boldsymbol{Y}) \neq \exp(\boldsymbol{X}) \exp(\boldsymbol{Y}).$$

2.5 Lie Group Representation

Please recall that if $\mathcal{V}$ is a vector space over $\mathbb{R}$, $GL(\mathcal{V})$ is the group of the linear transformations of $\mathcal{V}$ and, assuming $\dim \mathcal{V} = n$, it is isomorphic to $GL(n)$ (cf. Appendix B.4.3).

Definition 2.38. Let G be a Lie group. A *(finite-dimensional) representation* of G is a Lie group homomorphism from G to $GL(\mathcal{V})$ for some real vector space $\mathcal{V}$.

A representation $\rho\colon G \to GL(\mathcal{V})$ is *faithful* if it is injective, and then the Lie group G is isomorphic with the subgroup $\rho(G) \subseteq GL(\mathcal{V}) \cong GL(n)$.

There exists a connection between group representations and actions defined in Section A.2.2. Actually, if $\rho\colon G \to GL(\mathcal{V})$ is a representation of G, there exists a left action $\alpha : G \times \mathcal{V} \to \mathcal{V}$ defined by $\alpha(g, \mathbf{v}) = \rho(g)\mathbf{v}$. The action α is *linear*, in the sense that the associated map on $\mathcal{V}$ is a linear transformation.

Moreover, a notion of representation also applies to Lie algebras, which are associated with the Lie algebra $\mathfrak{gl}(\mathcal{V})$ of the linear maps of a vector space $\mathcal{V}$.

Definition 2.39. Let $\mathfrak{g}$ be a n-dimensional Lie algebra. A *(finite-dimensional) representation* of $\mathfrak{g}$ is a Lie algebra homomorphism $\varphi\colon \mathfrak{g} \to \mathfrak{gl}(\mathcal{V})$, for some real vector space $\mathcal{V}$.

Similarly to the group representations, the homomorphism $\varphi\colon \mathfrak{g} \to \mathfrak{gl}(\mathcal{V})$ is a *faithful* representation when it is injective. In such a case, $\mathfrak{g}$ is isomorphic to the Lie subalgebra $\varphi(\mathfrak{g}) \subseteq \mathfrak{gl}(\mathcal{V}) \cong \mathfrak{gl}(n)$.

The representations of Lie algebras are strictly related to the ones of Lie groups. Specifically, if G is a Lie group with the Lie algebra $\mathfrak{g}$ and the map $\rho : G \to GL(\mathcal{V})$ is any representation, the induced Lie algebra homomorphism $\rho_*\colon \mathfrak{g} \to \mathfrak{gl}(\mathcal{V})$, defined by (2.33), is a representation of $\mathfrak{g}$.

Moreover, any action of the Lie group G on its algebra $\mathfrak{g}$, considered as a vector space, provides a representation of G. Among these representations, a key role is played by the adjoint representation, which is induced by the conjugation map C_g defined by (A.6).

Definition 2.40. Let G be a Lie group and $\mathfrak{g}$ the relevant Lie algebra. The *adjoint representation* of G is the representation

$$\begin{aligned} \mathrm{Ad} :\ & G \to GL(\mathfrak{g}) \\ & g \mapsto \mathrm{Ad}_g = (C_g)_*\,, \end{aligned} \tag{2.40}$$

where $(C_g)_* : \mathfrak{g} \to \mathfrak{g}$ is the Lie algebra homomorphism induced by the conjugation C_g.

The characterization of $(C_g)_*$ as a Lie group homomorphism promptly comes from Proposition 2.34, since $C_g : G \to G$ is a Lie group homomorphism.

Moreover, recalling that $C_{g_1 g_2} = C_{g_1} \circ C_{g_2}$, the relevant Lie algebra homomorphisms do satisfy $(C_{g_1 g_2})_* = (C_{g_1})_* \circ (C_{g_2})_*$, whence

$$\mathrm{Ad}_{g_1 g_2} = \mathrm{Ad}_{g_1} \circ \mathrm{Ad}_{g_2}\,, \quad \forall\, g_1, g_2 \in G\,,$$

as well as the bijectivity of C_g implies the one of Ad_g, with the inverse given by

$$\mathrm{Ad}_g^{-1} = \mathrm{Ad}_{g^{-1}}\,, \quad \forall\, g \in G\,.$$

Consequently, Ad_g is actually an automorphism of $\mathfrak{g}$ for any g of G and the map $\mathrm{Ad}\colon G \to GL(\mathfrak{g})$ is effectively a representation of G.

In addition to the adjoint representation for a Lie group, it is also possible to define a representation for a Lie algebra $\mathfrak{g}$ considering the linear space $\mathfrak{gl}(\mathfrak{g})$ of the linear maps on $\mathfrak{g}$.

Definition 2.41. Let $\mathfrak{g}$ be a Lie algebra and $\mathfrak{gl}(\mathfrak{g})$ the space of its linear transformations. The *adjoint representation* of $\mathfrak{g}$ is the representation

$$\begin{aligned} \mathrm{ad}:\ \ & \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}) \\ & \boldsymbol{X} \mapsto \mathrm{ad}_{\boldsymbol{X}}\,, \end{aligned} \tag{2.41}$$

such that $\mathrm{ad}_{\boldsymbol{X}}\colon \mathfrak{g} \to \mathfrak{g}$ is defined by the Lie bracket $\mathrm{ad}_{\boldsymbol{X}}(\boldsymbol{Y}) = [\boldsymbol{X}, \boldsymbol{Y}]$.

Recalling Property 1 in Proposition 2.21, the Lie bracket operator is linear in its entries, so that $\mathrm{ad}_{\boldsymbol{X}}$ is actually a linear map of $\mathfrak{gl}(\mathfrak{g})$ for each $\boldsymbol{X} \in \mathfrak{g}$.

On the other hand, the linearity of the Lie bracket with respect to the first entry implies the linearity of the map ad:

$$\mathrm{ad}_{a\boldsymbol{X}_1 + b\boldsymbol{X}_2} = a\,\mathrm{ad}_{\boldsymbol{X}_1} + b\,\mathrm{ad}_{\boldsymbol{X}_2}\,.$$

Moreover, using the Jacobi identity and the antisymmetry of the Lie bracket, expressed by Properties 3 and 2 in Proposition 2.21, respectively, one finds

$$\begin{aligned}\mathsf{ad}_{[X_1,X_2]}\,\boldsymbol{Y} &= \big[[\boldsymbol{X}_1,\boldsymbol{X}_2],\boldsymbol{Y}\big] = \big[\boldsymbol{X}_1,[\boldsymbol{X}_2,\boldsymbol{Y}]\big] - \big[\boldsymbol{X}_2,[\boldsymbol{X}_1,\boldsymbol{Y}]\big] \\ &= \mathsf{ad}_{X_1}\left(\mathsf{ad}_{X_2}\,\boldsymbol{Y}\right) - \mathsf{ad}_{X_2}\left(\mathsf{ad}_{X_1}\,\boldsymbol{Y}\right) \\ &= \left(\mathsf{ad}_{X_1}\circ\mathsf{ad}_{X_2} - \mathsf{ad}_{X_2}\circ\mathsf{ad}_{X_1}\right)\boldsymbol{Y}\,,\end{aligned}$$

which, for the arbitrariness of $\boldsymbol{Y}\in\mathfrak{g}$, implies

$$\mathsf{ad}_{[X_1,X_2]} = [\mathsf{ad}_{X_1},\mathsf{ad}_{X_2}]\,.$$

The specifications here above show that the adjoint map is consistent with the Lie algebra structure of $\mathfrak{g}$ and $\mathfrak{gl}(\mathfrak{g})$ and provides a representation of $\mathfrak{g}$.

Moreover, if $\mathfrak{g}$ is the Lie algebra of the Lie group G, the adjoint representation of $\mathfrak{g}$ is precisely the homomorphism induced by the adjoint representation of G.

Proposition 2.42. *Let G be a Lie group with the Lie algebra $\mathfrak{g}$ and let $\mathrm{Ad}\colon G\to GL(\mathfrak{g})$ be the adjoint representation. Then, the induced Lie algebra homomorphism $\mathrm{Ad}_*\colon\mathfrak{g}\to\mathfrak{gl}(\mathfrak{g})$ is the adjoint representation ad of $\mathfrak{g}$.*

Proof. Consider a vector field $\boldsymbol{X}\in\mathfrak{g}$ and evaluate the map $\mathrm{Ad}_*\,\boldsymbol{X}$ at an arbitrary $\boldsymbol{Y}\in\mathfrak{g}$.

Since any vector field of $\mathfrak{g}$ is left-invariant, it is completely determined by its value at the identity $e\in G$. The same applies to $\mathrm{Ad}_*\,\boldsymbol{X}\in\mathfrak{gl}(\mathfrak{g})$, which is evaluated at $\mathbf{id}_{\mathfrak{g}}$, so that the following evaluation is considered:

$$\big((\mathrm{Ad}_*\,\boldsymbol{X})(\boldsymbol{Y})\big)_e = (\mathrm{Ad}_*\,\boldsymbol{X})_{\mathrm{id}_{\mathfrak{g}}}(\boldsymbol{Y}_e) = \big(\mathrm{Ad}_*|_e\boldsymbol{X}_e\big)(\boldsymbol{Y}_e)\,.$$

Moreover, recalling that the curve $\gamma\colon\mu\mapsto\exp(\mu\boldsymbol{X})$ satisfies $\gamma(0)=e$ and $\gamma'(0)=\boldsymbol{X}_e$, Corollary 2.16 can be applied to evaluate the differential of the adjoint map, obtaining

$$\big((\mathrm{Ad}_*\,\boldsymbol{X})(\boldsymbol{Y})\big)_e = (\mathrm{Ad}\circ\gamma)'(0)\boldsymbol{Y}_e = \left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_0\Big(\mathrm{Ad}_{\gamma(\mu)}\,\boldsymbol{Y}\Big)_e = \left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_0(C_{\gamma(\mu)})_*|_e\boldsymbol{Y}_e\,,$$

where Definition 2.40 has been applied.

Furthermore, being $C_{\gamma(\mu)} = R_{\gamma(-\mu)} \circ L_{\gamma(\mu)}$, and using the left-invariance of $\boldsymbol{Y}$, the above relation becomes

$$\begin{aligned}\big((\mathrm{Ad}_* \boldsymbol{X})(\boldsymbol{Y})\big)_e &= \frac{\mathrm{d}}{\mathrm{d}\mu}\bigg|_0 \Big((R_{\gamma(-\mu)})_*\big|_{\gamma(\mu)} \circ (L_{\gamma(\mu)})_*\big|_e \Big) \boldsymbol{Y}_e \\ &= \frac{\mathrm{d}}{\mathrm{d}\mu}\bigg|_0 (R_{\gamma(-\mu)})_*\big|_{\gamma(\mu)} \boldsymbol{Y}_{\gamma(\mu)} \,,\end{aligned}$$

In addition, using (2.38), it is easy to recognize that $R_{\gamma(-\mu)}$ coincides with the map $\theta_{-\mu}$, where θ is the flow of $\boldsymbol{X}$, so that the derivative in the above relation can be expressed as

$$\big((\mathrm{Ad}_* \boldsymbol{X})(\boldsymbol{Y})\big)_e = \lim_{\mu \to 0} \frac{(\theta_{-\mu})_*\big|_{\theta_\mu(e)} \boldsymbol{Y}_{\theta_\mu(e)} - \boldsymbol{Y}_e}{\mu} \,,$$

which, by Definition 2.28, is the Lie derivative $(\mathscr{L}_{\boldsymbol{X}} \boldsymbol{Y})_e$.

Consequently, since any vector field in $\mathfrak{g}$ is determined by its value at e, and applying Proposition 2.30, one finally has

$$(\mathrm{Ad}_* \boldsymbol{X})(\boldsymbol{Y}) = \mathscr{L}_{\boldsymbol{X}} \boldsymbol{Y} = [\boldsymbol{X}, \boldsymbol{Y}] = \mathrm{ad}_{\boldsymbol{X}}(\boldsymbol{Y}) \,. \qquad \square$$

2.6 Matrix Lie Groups

Within the context of the Lie groups, a central role is played by the groups of matrices.

As a matter of fact, the law of composition for matrix groups consists in the standard matrix multiplication, so that the group operations are reduced to functions on the entries. Moreover, specifying the smooth manifold structure of the matrix groups, the differential operations proper of a Lie group take themselves the form of matrix operations.

In order to point out how some matrix operations represent smooth maps on manifolds, the exponential function is first introduced.

2.6.1 Exponential of Matrices

The exponential of a matrix is crucial in the theory of matrix Lie groups, since it specializes the exponential map connecting a Lie algebra and the corresponding Lie group.

Definition 2.43. Let $\mathbf{X}$ be an n-dimensional square matrix. The *exponential* of $\mathbf{X}$, denoted as $e^{\mathbf{X}}$ or $\exp(\mathbf{X})$, is the power series

$$e^{\mathbf{X}} = \exp(\mathbf{X}) = \sum_{k=0}^{\infty} \frac{\mathbf{X}^k}{k!}\,. \tag{2.42}$$

Each term $\mathbf{X}^k$ appearing in (2.42) the k-th matrix power, i.d. the repeated matrix multiplication of $\mathbf{X}$ with itself:

$$\mathbf{X}^k = \underbrace{\mathbf{X} \times \cdots \times \mathbf{X}}_{k \text{ times}},$$

with the convention $\mathbf{X}^0 = \mathbf{I}$ for any $\mathbf{X}$.

Please notice that the series (2.42) is, formally, the same series defining the exponential function in real analysis (Rudin, 1986) and similar features are expected. Specifically, it can be proved that the series converges for all $\mathbf{X} \in M_n$ and $\exp : \mathbf{X} \mapsto e^{\mathbf{X}}$ is a continuous function of $\mathbf{X}$ (see, e.g., Hall (2015)).

The main difference between the matrix exponential defined by (2.42) and the relevant real-valued function concerns the addition formula, since the matrix exponentials, in general, do not commute:

$$e^{(\mathbf{X}+\mathbf{Y})} \neq e^{\mathbf{X}} e^{\mathbf{Y}}\,.$$

However, when the matrices $\mathbf{X}$ and $\mathbf{Y}$ do commute, i.e. $\mathbf{XY} = \mathbf{YX}$, the relevant exponentials also commute and the addition formula applies. Such a property and other ones are summarized by the following proposition.

Proposition 2.44. *Let $\mathbf{X}$ and $\mathbf{Y}$ be n-by-n matrices. Then, the following properties are true.*

1. *If $\mathbf{XY} = \mathbf{YX}$, then $e^{(\mathbf{X}+\mathbf{Y})} = e^{\mathbf{X}} e^{\mathbf{Y}} = e^{\mathbf{Y}} e^{\mathbf{X}}$.*

2. $e^{\mathbf{0}} = \mathbf{I}$.
3. $e^{(a+b)\mathbf{X}} = e^{a\mathbf{X}} e^{b\mathbf{X}}\,, \;\; \forall\, a, b \in \mathbb{R}$.
4. $e^{\mathbf{X}}$ *is invertible and the inverse is* $(e^{\mathbf{X}})^{-1} = e^{-\mathbf{X}}$.

Proof. 1. Let us use (2.42) for both $e^{\mathbf{X}}$ and $e^{\mathbf{Y}}$ and apply the Cauchy product of these two series:

$$e^{\mathbf{X}} e^{\mathbf{Y}} = \sum_{k=0}^{\infty} \frac{\mathbf{X}^k}{k!} \sum_{h=0}^{\infty} \frac{\mathbf{Y}^h}{h!} = \sum_{k=0}^{\infty} \sum_{m=0}^{k} \frac{\mathbf{X}^k \mathbf{Y}^{(k-m)}}{k!(k-m)!}$$
$$= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{m=0}^{k} \frac{m!}{m!(k-m)!} \mathbf{X}^k \mathbf{Y}^{(k-m)} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{m=0}^{k} \binom{k}{m} \mathbf{X}^k \mathbf{Y}^{(k-m)}\,.$$

If $\mathbf{XY} = \mathbf{YX}$, the last summation is the power $(\mathbf{X}+\mathbf{Y})^k$, so that one obtains

$$e^{\mathbf{X}} e^{\mathbf{Y}} = \sum_{k=0}^{\infty} \frac{(\mathbf{X}+\mathbf{Y})^k}{k!} = e^{\mathbf{X}+\mathbf{Y}}\,.$$

2. Observe that any $\mathbf{X} \in M_n$ can be written as $\mathbf{X} + \mathbf{0}$, and clearly $\mathbf{X0} = \mathbf{0X} = \mathbf{0}$. Then, Property 1 of the proposition applies and one has

$$e^{\mathbf{X}} = e^{(\mathbf{X}+\mathbf{0})} = e^{\mathbf{X}} e^{\mathbf{0}}\,,$$

whence $e^{\mathbf{0}} = \mathbf{I}$.

3. The matrix $(a+b)\mathbf{X}$ can be written as $a\mathbf{X} + b\mathbf{X}$, and the matrices $a\mathbf{X}$ and $b\mathbf{X}$ do commute. Then, by Property 1 of the proposition, one has

$$e^{(a+b)\mathbf{X}} = e^{a\mathbf{X}+b\mathbf{X}} = e^{a\mathbf{X}} e^{b\mathbf{X}}\,.$$

4. Setting $a = 1$ and $b = -1$ in Property 3 of the proposition and then applying Property 2, the following relation holds true:

$$e^{\mathbf{X}} e^{-\mathbf{X}} = e^{\mathbf{X}-\mathbf{X}} = e^{\mathbf{0}} = \mathbf{I}\,,$$

whence $e^{-\mathbf{X}}$ is the inverse of $e^{\mathbf{X}}$. □

2.6.2 Smooth Structure of Matrix Groups

A matrix group G is a subset of the space of square matrices M_n over $\mathbb{R}$ which satisfies the group axioms given by Definition A.8. Specifically, G is a subset of invertible matrices, which is closed under the matrix multiplication and contains the identity.

Roughly speaking, the manifold structure of a matrix group G comes from the fact that the entries of its matrices are real variables. In addition, the smoothness arises from considering the group as the domain of some differentiable functions.

As a matter of fact, a matrix group G can be properly characterized as a *matrix Lie group* if it is a closed subgroup of the general linear group $GL(n)$. This means that for any sequence $\mathbf{A}_k$ of matrices in G, which converges to some $\mathbf{A} \in GL(n)$, the matrix $\mathbf{A}$ is actually in G.

A discussion about how the smoothness of $GL(n)$ is reflected on its closed subgroups can be found, among others, in Gallier and Quaintance (2020), Hall (2015). In the context of this work, it is enough to say that $GL(n)$ is assumed as the reference matrix Lie group and, as such, its differential structure is briefly described.

Please recall that the general linear group $GL(n)$ is introduced by Definition B.59 as the group of the invertible n-dimensional matrices and, in this context, the reference field is $\mathbb{R}$. Its smooth manifold structure comes from the fact of being an open subset of the real vector space $M_n(\mathbb{R})$ and, as such, its dimension is n^2.

Specifically, let us notice that the determinant function $\det : \mathbb{R}^{n^2} \to \mathbb{R}$ is differentiable because it is a polynomial in the n^2 entries. Moreover, since the determinant of invertible matrices is non-null, $GL(n)$ represents the preimage of the open subset $\mathbb{R}/\{0\}$:

$$GL(n) = \det^{-1}\big(\mathbb{R}/\{0\}\big)\,,$$

whence $GL(n)$ results an open subset of M_n and also inherits its smoothness.

Furthermore, the entries of the matrix $\mathbf{AB}$ are polynomials of the entries of $\mathbf{A}$ and $\mathbf{B}$, so that the matrix multiplication is smooth. Also, the inversion is itself smooth by Cramer's rule, concluding that $GL(n)$ is actually a Lie group (cf. Definition 2.31).

At the same time, let us also consider the Lie algebra over $\mathbb{R}$, introduced following Definition B.56, as the real vector space M_n along with the commutator bracket and here simply denoted as $\mathfrak{gl}(n)$.

Since $\mathfrak{gl}(n)$ has the same dimension n^2 as $GL(n)$, it is isomorphic, as a vector space, with the tangent space at the identity $\mathbf{I}$. In addition, by Proposition 2.33, the space $T_{\mathbf{I}}GL(n)$ is also isomorphic with the Lie algebra of $GL(n)$, concluding that there exists a vector space isomorphism between $\mathfrak{gl}(n)$ and the Lie algebra of $GL(n)$:

$$\mathfrak{gl}(n) \cong \mathsf{Lie}\left(GL(n)\right). \tag{2.43}$$

It is possible to prove (see, e.g., Lee (2012)) that the above relation also holds true in the sense of a Lie algebra isomorphism, which justifies the notation $\mathfrak{gl}(n)$ for the Lie algebra of the n-dimensional square matrices.

As a Lie group, $GL(n)$ is connected to its Lie algebra via the exponential map (cf. Definition 2.36). However, in this case, such a relation takes the explicit form of the matrix exponential defined by (2.42).

Actually, since by Property 4 in Proposition 2.44 the exponential of any matrix is invertible, $e^{\mathbf{X}}$ is a matrix of $GL(n)$ for each $\mathbf{X} \in M_n$.

Furthermore, considering the Lie algebra $\mathfrak{gl}(n)$ as M_n endowed with the commutator bracket, the matrix exponential is actually a map between $\mathfrak{gl}(n)$ and the Lie group $GL(n)$. Then, with the identification (2.43), the map $\exp\colon \mathfrak{gl}(n) \to GL(n)$ represents the specialization of the exponential map to the general linear group.

Such a result is more properly justified by the following proposition.

Proposition 2.45. *For any $\mathbf{X} \in \mathfrak{gl}(n)$, the one-parameter subgroup of $GL(n)$ generated by $\mathbf{X}$ is $\gamma(\mu) = e^{\mu\mathbf{X}}$.*

Proof. Recalling that any left-invariant vector field on a Lie group is induced by a tangent vector at the identity by (2.32), let $\boldsymbol{V} \in \mathsf{Lie}\left(GL(n)\right)$ be defined by $\mathbf{X} \in \mathfrak{gl}(n)$ through the condition $\boldsymbol{V}_{\mathbf{I}} = \mathbf{X}$.

By Proposition 2.35, the one-parameter subgroup of $GL(n)$ generated by $\mathbf{X}$ is the integral curve of the vector field $\boldsymbol{V}$ starting at $\mathbf{I}$, so that let us verify that such an integral curve is actually represented by γ.

Exploiting Property 2 in Proposition 2.44, it is easy to see that γ is a curve starting at $\mathbf{I}$:

$$\gamma(0) = e^{0\mathbf{X}} = e^{\mathbf{0}} = \mathbf{I}\,.$$

Moreover, expressing the exponential of $\mu\mathbf{X}$ through the series (2.42) and taking the derivative with respect to μ, the velocity of γ results

$$\gamma'(\mu) = \sum_{k=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}\mu}\left(\frac{\mu^k \mathbf{X}^k}{k!}\right) = \sum_{k=1}^{\infty} \frac{k\mu^{k-1}\mathbf{X}^k}{k!} = \sum_{k=1}^{\infty}\left(\frac{\mu^{k-1}\mathbf{X}^{k-1}}{(k-1)!}\right)\mathbf{X}\,.$$

Please observe that the differentiated series does converge because, a part for the factor $\mathbf{X}$, it is the same convergent series defining $e^{\mu\mathbf{X}}$.

Then, noting also that $\mathbf{X}$ commutes with $\mathbf{X}^{k-1}$, for any $k > 0$, the velocity of the curve is

$$\gamma'(\mu) = \frac{\mathrm{d}}{\mathrm{d}\mu} e^{\mu\mathbf{X}} = e^{\mu\mathbf{X}}\mathbf{X} = \mathbf{X}e^{\mu\mathbf{X}}\,, \tag{2.44}$$

and its value at the identity $\mathbf{I} \in GL(n)$ results $\gamma'(0) = \mathbf{X} = \boldsymbol{V}_{\mathbf{I}}$. □

As consequence of Proposition 2.45, the flow of the vector field $\boldsymbol{V}$, given by (2.38), is expressed as

$$\theta(\mu, \mathbf{A}) = L_{\mathbf{A}}(e^{\mu\mathbf{X}}) = R_{e^{\mu\mathbf{X}}}(\mathbf{A}) = \mathbf{A}e^{\mu\mathbf{X}}\,, \tag{2.45}$$

where the matrix $\mathbf{X} \in \mathfrak{gl}(n)$ uniquely identifies the left-invariant vector field $\boldsymbol{V} \in \mathsf{Lie}\left(GL(n)\right)$ by setting $\mathbf{X} = \boldsymbol{V}_{\mathbf{I}}$.

Moreover, the integral curve of $\boldsymbol{V}$ starting at $\mathbf{A} \in GL(n)$ is the map

$$\begin{aligned} \gamma_{\mathbf{A}} :\ & \mathbb{R} \to GL(n) \\ & \mu \mapsto \mathbf{A}e^{\mu\mathbf{X}}\,, \end{aligned} \tag{2.46}$$

whence the velocity at $\mu = 0$ results

$$\gamma'_{\mathbf{A}}(0) = \frac{\mathrm{d}}{\mathrm{d}\mu}(\mathbf{A}e^{\mu\mathbf{X}})\Big|_0 = \mathbf{A}e^{\mu\mathbf{X}}\mathbf{X}\big|_0 = \mathbf{AX}\,.$$

Since $\gamma'_{\mathbf{A}}(0)$ is also the value of the left-invariant vector field $\boldsymbol{V}$ at $\mathbf{A}$, recalling the defining property (2.29), one infers

$$\gamma'_{\mathbf{A}}(0) = \boldsymbol{V}_{\mathbf{A}} = \mathrm{d}L_{\mathbf{A}}\big|_{\mathbf{I}}\boldsymbol{V}_{\mathbf{I}} = \mathrm{d}L_{\mathbf{A}}\big|_{\mathbf{I}}\mathbf{X}\,,$$

so that the differential $\mathrm{d}L_{\mathbf{A}}\big|_{\mathbf{I}}$ coincides with the matrix $\mathbf{A}$, considered as a linear operator from $T_{\mathbf{I}}GL(n)$ to $T_{\mathbf{A}}GL(n)$, and the value of the vector field $\boldsymbol{V}$ at any point $\mathbf{A} \in GL(n)$ is simply given by the matrix multiplication:

$$\boldsymbol{V}_{\mathbf{A}} = \mathbf{A}\boldsymbol{V}_{\mathbf{I}} = \mathbf{AX}\,. \tag{2.47}$$

Please observe that if a further curve σ is defined by left translation as $\sigma = L_{\mathbf{B}} \circ \gamma_{\mathbf{A}}$, for some $\mathbf{B} \in GL(n)$, the starting point is $\sigma(0) = \mathbf{BA}$ and the relevant velocity vector is

$$\sigma'(0) = (L_{\mathbf{B}} \circ \gamma_{\mathbf{A}})'(0) = \mathrm{d}L_{\mathbf{B}}\big|_{\gamma_{\mathbf{A}}(0)}\gamma'_{\mathbf{A}}(0) = \mathrm{d}L_{\mathbf{B}}\big|_{\mathbf{A}}\boldsymbol{V}_{\mathbf{A}}\,.$$

In addition, since the explicit form of $\sigma(\mu) = \mathbf{BA}e^{\mu\mathbf{X}}$, and the velocity at the starting point is

$$\sigma'(0) = \frac{\mathrm{d}}{\mathrm{d}\mu}(\mathbf{BA}e^{\mu\mathbf{X}})\Big|_0 = \mathbf{BAX} = \mathbf{B}\boldsymbol{V}_{\mathbf{A}}\,,$$

whence, by comparison the differential of the left translation results

$$\mathrm{d}L_{\mathbf{B}}\big|_{\mathbf{A}} = \mathbf{B}\,, \tag{2.48}$$

whatever the matrices $\mathbf{A}$ and $\mathbf{B}$ are in the matrix group $GL(n)$.

2.6.3 Adjoint Representation of Matrix Groups

The adjoint representation of a Lie group G has been introduced by Definition 2.40 as the group homomorphism between G and the general linear group of the lie algebra of G.

With reference to $GL(n)$, the adjoint representation of $GL(n)$ specializes as follows:

$$\begin{aligned} \mathrm{Ad}:\ & GL(n) \to GL\big(\mathfrak{gl}(n)\big) \\ & \mathbf{A} \mapsto \mathrm{d}C_{\mathbf{A}}\big|_{\mathbf{I}}\,, \end{aligned} \tag{2.49}$$

where, with the usual identification $\mathsf{Lie}\,\big(GL(n)\big) \cong \mathfrak{gl}(n)$, the Lie algebra homomorphism $(C_{\mathbf{A}})_*: \mathsf{Lie}\,\big(GL(n)\big) \to \mathsf{Lie}\,\big(GL(n)\big)$ induced by the conjugation map $C_{\mathbf{A}}$ specializes to the differential $\mathrm{d}C_{\mathbf{A}}\big|_{\mathbf{I}}: \mathfrak{gl}(n) \to \mathfrak{gl}(n)$.

In order to provide an explicit form for the adjoint representation, let us consider $\gamma : \mu \mapsto e^{\mu\mathbf{X}}$ as the integral curve of the left-invariant field $\boldsymbol{V}$ starting at $\mathbf{I}$, with $\gamma'(0) = \boldsymbol{V}_{\mathbf{I}} = \mathbf{X}$. Hence, applying the conjugation $C_{\mathbf{A}}$ provides the curve $\sigma = C_{\mathbf{A}} \circ \gamma$, such that $\sigma(0) = \mathbf{I}$, with the velocity evaluated as

$$\sigma'(0) = (C_{\mathbf{A}} \circ \gamma)'(0) = \left(\mathrm{d}C_{\mathbf{A}} \circ \frac{\mathrm{d}\gamma}{\mathrm{d}\mu}'\right)\bigg|_0 = \mathrm{d}C_{\mathbf{A}}\big|_{\sigma(0)}\gamma'(0)\,,$$

that is $\sigma'(0) = \mathrm{d}C_{\mathbf{A}}\big|_{\mathbf{I}}\mathbf{X}$.

At the same time, the explicit expression $\sigma: \mu \mapsto \mathbf{A}e^{\mu\mathbf{X}}\mathbf{A}^{-1}$ gives

$$\sigma'(0) = \frac{\mathrm{d}}{\mathrm{d}\mu}\bigg|_0 \big(\mathbf{A}e^{\mu\mathbf{X}}\mathbf{A}^{-1}\big) = \big(\mathbf{A}e^{\mu\mathbf{X}}\mathbf{X}\mathbf{A}^{-1}\big)\big|_0 = \mathbf{A}\mathbf{X}\mathbf{A}^{-1}\,,$$

whence $\mathrm{d}C_{\mathbf{A}}\big|_{\mathbf{I}}\mathbf{X} = \mathbf{A}\mathbf{X}\mathbf{A}^{-1}$ and the differential of the map $C_{\mathbf{A}}$ coincides with the same conjugation $C_{\mathbf{A}}$, considered as an automorphism of $\mathfrak{gl}(n)$.

Then, the adjoint representation of $GL(n)$ is the group homomorphism $\mathrm{Ad}: GL(n) \to GL\big(\mathfrak{gl}(n)\big)$ and its value at $\mathbf{A}$ results

$$\begin{aligned} \mathrm{Ad}_{\mathbf{A}}:\ & \mathfrak{gl}(n) \to \mathfrak{gl}(n) \\ & \mathbf{X} \mapsto \mathrm{Ad}_{\mathbf{A}}(\mathbf{X}) = \mathbf{A}\mathbf{X}\mathbf{A}^{-1}\,. \end{aligned} \tag{2.50}$$

Regarding the Lie algebra $\mathfrak{gl}(n)$, the adjoint representation introduced by Definition 2.41 specializes to the following map:

$$\begin{aligned} \mathrm{ad}:\ & \mathfrak{gl}(n) \to \mathfrak{gl}\big(\mathfrak{gl}(n)\big) \\ & \mathbf{X} \mapsto \mathrm{ad}_{\mathbf{X}}\,, \end{aligned} \tag{2.51}$$

where the map $\mathsf{ad}_{\mathbf{X}} \colon \mathfrak{gl}(n) \to \mathfrak{gl}(n)$ is a homomorphism of $\mathfrak{gl}(n)$ defined by matrix commutator as $\mathsf{ad}_{\mathbf{X}}(\mathbf{Y}) = [\mathbf{X}, \mathbf{Y}]$.

In addition, it is useful to consider also the k-th power of the adjoint representation $\mathsf{ad}^k_{\mathbf{X}}$, defined as

$$\mathsf{ad}^k_{\mathbf{X}}(\mathbf{Y}) = \underbrace{\big[\mathbf{X}, \cdots \big[\mathbf{X}, [\mathbf{X}, \mathbf{Y}]\big]\big]}_{k \text{ times}} . \tag{2.52}$$

2.6.4 Differential of the Exponential Map

Assuming the identification $\mathsf{Lie}\big(GL(n)\big) \cong \mathfrak{gl}(n)$, the exponential map, introduced by Definition 2.36, specializes to

$$\begin{aligned} \exp \colon \; \mathfrak{gl}(n) &\to GL(n) \\ \mathbf{X} &\mapsto e^{\mathbf{X}} , \end{aligned}$$

with the matrix exponential $e^{\mathbf{X}}$ given by (2.42).

In addition, since $\mathfrak{gl}(n)$ coincides with M_n as a vector space, it has a natural smooth manifold structure and any tangent space is canonical isomorphic with $\mathfrak{gl}(n)$ (see, e.g., Lee (2012)).

Then, the exponential map can be regarded as a smooth map between smooth manifolds, and with the identification $T\mathfrak{gl}(n) \cong \mathfrak{gl}(n)$, its differential at a point $\mathbf{X} \in \mathfrak{gl}(n)$ is a map from $\mathfrak{gl}(n)$ to $T_{\exp(\mathbf{X})}GL(n)$:

$$\begin{aligned} \mathsf{dexp}|_{\mathbf{X}} \colon \; \mathfrak{gl}(n) &\to T_{\exp(\mathbf{X})}GL(n) \\ \mathbf{Y} &\mapsto \mathsf{dexp}|_{\mathbf{X}}\mathbf{Y} . \end{aligned} \tag{2.53}$$

The explicit expression of $\mathsf{dexp}|_{\mathbf{X}}$ is provided by the following proposition.

Proposition 2.46. *The differential of the exponential map* $\exp \colon \mathfrak{gl}(n) \to GL(n)$ *at a point* $\mathbf{X} \in \mathfrak{gl}(n)$ *is given by*

$$\mathsf{dexp}|_{\mathbf{X}} = e^{\mathbf{X}} \mathbf{D}_{\mathbf{X}} , \tag{2.54}$$

where

$$\mathbf{D}_{\mathbf{X}} = \sum_{k=0}^{\infty} (-1)^k \frac{\mathrm{ad}_{\mathbf{X}}^k}{(k+1)!} \,. \tag{2.55}$$

Proof. Consider a curve $\phi\colon \mu \mapsto \phi(\mu)$ in $\mathfrak{gl}(n)$ such that $\phi(\mu) = \mathbf{X}$, with the velocity $\phi'(\mu) = \mathbf{Y}$, for some $\mathbf{Y} \in \mathfrak{gl}(n)$. Such a curve is well-defined in $\mathfrak{gl}(n)$ because $\phi(\mu)$ can be thought as a matrix function of one parameter and, with the identification $T_{\phi(\mu)}\mathfrak{gl}(n) \cong \mathfrak{gl}(n)$, the velocity $\phi'(\mu)$ results itself an element of $\mathfrak{gl}(n)$.

The composition of the exponential map with ϕ is the curve of $GL(n)$ defined as $\gamma\colon \mu \mapsto e^{\phi(\mu)}$, whose velocity at μ is a vector of the tangent space $T_{\gamma(\mu)}GL(n)$, with $\gamma(\mu) = \exp(\mathbf{X}) \in GL(n)$, evaluated as

$$\gamma'(\mu) = (\exp \circ \phi)'(\mu) = \mathrm{dexp}|_{\phi(\mu)}\phi'(\mu) = \mathrm{dexp}|_{\mathbf{X}}\mathbf{Y} \,. \tag{2.56}$$

With the aim of finding an explicit expression of $\gamma'(\mu)$, let us define

$$\theta(\lambda, \mu) = e^{-\lambda\phi(\mu)} \frac{\partial}{\partial \mu} e^{\lambda\phi(\mu)} \,,$$

which satisfies $\theta(0, \mu) = \mathbf{0}$.

Recalling the property (2.44) of the matrix exponential, the derivative of θ with respect to λ reads:

$$\begin{aligned} \left.\frac{\partial\theta}{\partial\lambda}\right|_{(\lambda,\mu)} &= -\phi(\mu)e^{-\lambda\phi(\mu)} \frac{\partial}{\partial\mu} e^{\lambda\phi(\mu)} + e^{-\lambda\phi(\mu)} \frac{\partial}{\partial\mu}\left(\frac{\partial}{\partial\lambda} e^{\lambda\phi(\mu)}\right) \\ &= -\phi(\mu)\theta(\lambda, \mu) + e^{-\lambda\phi(\mu)} \frac{\partial}{\partial\mu}\left(e^{\lambda\phi(\mu)}\phi(\mu)\right) \\ &= -\phi(\mu)\theta(\lambda, \mu) + \theta(\lambda, \mu)\phi(\mu) + \phi'(\mu) \\ &= -\left[\phi(\mu), \theta(\lambda, \mu)\right] + \phi'(\mu) \,, \end{aligned}$$

where, since $\phi(\mu)$ and $\theta(\lambda, \mu)$ are both elements of $\mathfrak{gl}(n)$, the commutator bracket has been applied.

Moreover, since ϕ and ϕ' both depend solely on μ, evaluating again the derivative with respect to λ, one finds

$$\begin{aligned}
\left.\frac{\partial^2\theta}{\partial\lambda^2}\right|_{(\lambda,\mu)} &= -\left[\phi(\mu), \left.\frac{\partial\theta}{\partial\lambda}\right|_{(\lambda,\mu)}\right] \\
&= -\Big[\phi(\mu), -\big[\phi(\mu), \theta(\lambda,\mu)\big] + \phi'(\mu)\Big] \\
&= \Big[\phi(\mu), \big[\phi(\mu), \theta(\lambda,\mu)\big]\Big] - \big[\phi(\mu), \phi'(\mu)\big] \\
&= \mathrm{ad}^2_{\phi(\mu)}\left(\theta(\lambda,\mu)\right) - \mathrm{ad}_{\phi(\mu)}\left(\phi'(\mu)\right),
\end{aligned}$$

so that, by induction, the following relation can be derived:

$$\left.\frac{\partial^k\theta}{\partial\lambda^k}\right|_{(\lambda,\mu)} = (-1)^k\,\mathrm{ad}^k_{\phi(\mu)}\left(\theta(\lambda,\mu)\right) + (-1)^{k-1}\,\mathrm{ad}^{k-1}_{\phi(\mu)}\left(\phi'(\mu)\right),$$

which, setting $\lambda = 0$, along with the conditions $\phi(\mu) = \mathbf{X}$ and $\phi'(\mu) = \mathbf{Y}$, specializes to

$$\left.\frac{\partial^k\theta}{\partial\lambda^k}\right|_{(0,\mu)} = (-1)^{k-1}\,\mathrm{ad}^{k-1}_{\mathbf{X}}(\mathbf{Y}).$$

Recalling that $\theta(0,\mu) = 0$, the derivatives reported here above can be used to compute the Taylor series of $\theta(\lambda,\mu)$ around the point $(0,\mu)$:

$$\theta(\lambda,\mu) = \sum_{k=1}^{\infty}\left.\frac{\partial^k\theta}{\partial\lambda^k}\right|_{(0,\mu)}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty}(-1)^{k-1}\,\mathrm{ad}^{k-1}_{\mathbf{X}}(\mathbf{Y})\frac{\lambda^k}{k!},$$

and, setting $\lambda = 1$, one finally obtains

$$\theta(1,\mu) = e^{-\phi(\mu)}\frac{\partial}{\partial\mu}e^{\phi(\mu)} = \sum_{k=1}^{\infty}(-1)^{k-1}\frac{\mathrm{ad}^{k-1}_{\mathbf{X}}(\mathbf{Y})}{k!}.$$

The comparison with (2.56) allows one to recognize

$$e^{-\mathbf{X}}(\exp\circ\phi)'(\mu) = \sum_{k=1}^{\infty}(-1)^{k-1}\frac{\mathrm{ad}^{k-1}_{\mathbf{X}}(\mathbf{Y})}{k!},$$

that is

$$e^{-\mathbf{X}}\,\mathrm{dexp}|_{\mathbf{X}}\mathbf{Y} = \sum_{k=0}^{\infty}(-1)^{k}\frac{\mathrm{ad}^{k}_{\mathbf{X}}(\mathbf{Y})}{(k+1)!},$$

whence, considering the series expansion in the form (2.55), one finally finds

$$\mathrm{dexp}|_{\mathbf{X}}\mathbf{Y} = e^{\mathbf{X}}\mathbf{D}_{\mathbf{X}}\mathbf{Y}\,,$$

and, for the arbitrariness of $\mathbf{Y} \in \mathfrak{gl}(n)$, the expression (2.54) is proved. □

Please observe that the operator $\mathbf{D}_{\mathbf{X}}$ defined by (2.55) is usually denoted as

$$\mathbf{D}_{\mathbf{X}} = \frac{\mathbf{I} - e^{-\mathrm{ad}_{\mathbf{X}}}}{\mathrm{ad}_{\mathbf{X}}} = \sum_{k=0}^{\infty}(-1)^k \frac{\mathrm{ad}_{\mathbf{X}}^k}{(k+1)!}\,,$$

in accordance with the relevant scalar-valued function expressed by the same series of powers (see, e.g., Gallier and Quaintance (2020), Hall (2015)).

Moreover, given $\mathbf{Y} \in \mathfrak{gl}(n)$, the vector $\mathrm{dexp}|_{\mathbf{X}}\mathbf{Y}$ is an element of the tangent space $T_{\exp(\mathbf{X})}GL(n)$ and results from the left translation by the differential $\mathrm{d}L_{\exp(\mathbf{X})}\big|_{\mathbf{I}}$, as described by (2.29) for the general case.

Since for a matrix Lie group the left translation of tangent vectors takes the form of the matrix multiplication, writing $\mathrm{dexp}|_{\mathbf{X}}\mathbf{Y} = e^{\mathbf{X}}\mathbf{D}_{\mathbf{X}}\mathbf{Y}$ can be interpreted as the composition of the map $\mathbf{D}_{\mathbf{X}} \colon \mathfrak{gl}(n) \to T_{\mathbf{I}}GL(n)$ with the map $\mathrm{d}L_{\exp(\mathbf{X})}\big|_{\mathbf{I}} \colon T_{\mathbf{I}}GL(n) \to T_{\exp(\mathbf{X})}GL(n)$.

Then, with the identification $\mathfrak{gl}(n) \cong T_{\mathbf{I}}GL(n)$, the operator $\mathbf{D}_{\mathbf{X}}$ applies as an endomorphism of $\mathfrak{gl}(n)$:

$$\begin{aligned} \mathbf{D}_{\mathbf{X}} :\ & \mathfrak{gl}(n) \to \mathfrak{gl}(n) \\ & \mathbf{Y} \mapsto \mathbf{D}_{\mathbf{X}}\mathbf{Y} = \sum_{k=0}^{\infty}(-1)^k \frac{\mathrm{ad}_{\mathbf{X}}^k(\mathbf{Y})}{(k+1)!}\,. \end{aligned} \tag{2.57}$$

2.6.5 Logarithm of Matrices

Since Definition 2.43 is based on a power series, the exponential of a matrix can be seen, although with some limitations, as an extension of the ordinary exponential function to the space of matrices. Hence, one could ask if the same is also possible for the logarithm function, meant as the inverse of the exponential function.

With the aim to provide a map to be considered as the inverse of the exponential map, let us introduce the following definition.

Definition 2.47. The *logarithm* of a n-dimensional square matrix $\mathbf{A}$, denoted as $\log(\mathbf{A})$, is the power series

$$\log(\mathbf{A}) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{(\mathbf{A}-\mathbf{I})^k}{k}\,, \tag{2.58}$$

whenever the series converges.

Please observe again that the series (2.58) is formally the same as the power series providing the logarithm of a real scalar variable. Accordingly, the convergence of such a series is not assured for any $\mathbf{A} \in M_n$.

Specifically, it can be proved that the series (2.58) is convergent for all $\mathbf{A}$ such that $\|\mathbf{A}-\mathbf{I}\| < 1$, for a suitable matrix norm, and $\log\colon \mathbf{A} \mapsto \log(\mathbf{A})$ is well-defined as a continuous function of $\mathbf{A}$ (see, e.g., Hall (2015)).

Moreover, since $\log(\mathbf{A})$ is an n-by-n matrix, its exponential can be evaluated through (2.42), giving

$$e^{\log(\mathbf{A})} = \exp\big(\log(\mathbf{A})\big) = \mathbf{A}\,, \tag{2.59}$$

with the condition $\|\mathbf{A}-\mathbf{I}\| < 1$, and also

$$\log(e^{\mathbf{X}}) = \log\big(\exp(\mathbf{X})\big) = \mathbf{X}\,, \tag{2.60}$$

for all the square matrices $\mathbf{X}$ satisfying $\|\mathbf{X}-\mathbf{I}\| < \log 2$ (the necessity of this condition is discussed in Hall (2015)).

Even if the logarithm of a matrix appears to be formally as the inverse of the exponential function, its treatment requires greater attention.

Specifically, any invertible real matrix $\mathbf{A}$ can be expressed in the form $\mathbf{A} = e^{\mathbf{X}}$, but the matrix $\mathbf{X}$ is not unique (see, e.g., Hall (2015)). In addition, a matrix with real and positive entries might have negative or even complex eigenvalues, so that the logarithm of $\mathbf{A}$ might be, in general, a complex matrix.

However, if a matrix $\mathbf{A}$ has no negative real eigenvalues, there is a unique matrix $\mathbf{X}$, which is the logarithm of $\mathbf{A}$, with the eigenvalues lying in the strip

$\{ z \in \mathbb{C} \mid -\pi < \Im(z) < \pi \}$. In this conditions, $\mathbf{X}$ is referred to as the *principal logarithm* and, if $\mathbf{A}$ is a real matrix, it is also real (Higham, 2008).

On the basis of the above observations, for any $\mathbf{A} \in GL(n)$, the matrix denoted as $\mathbf{X} = \log(\mathbf{A})$ is intended as the unique real matrix $\mathbf{X} \in M_n$ which is the principal logarithm of $\mathbf{A}$.

Moreover, following Gallier and Quaintance (2020), the condition $\{ z \in \mathbb{C} \mid -\pi < \Im(z) < \pi \}$ for the spectrum of $\mathbf{X} = \log(\mathbf{A})$ defines a subset $\mathfrak{e}(n)$ of $\mathfrak{gl}(n)$ such that the differential $\mathrm{dexp}|_{\mathbf{X}}$ of the exponential map is non-singular. Consequently, that the restriction of $\exp$ to $\mathfrak{e}(n) \subset \mathfrak{gl}(n)$ is a diffeomorphism to $\exp\big(\mathfrak{e}(n)\big) \subset GL(n)$, and in particular it is bijective.

Such a result is summarized by the following proposition. It is reported without the proof, referring to Gallier and Quaintance (2020) and Higham (2008), among others, for a deeper discussion.

Theorem 2.48. *The restriction of the exponential map to $\mathfrak{e}(n) \subset \mathfrak{gl}(n)$ is a diffeomorphism $\exp : \mathfrak{e}(n) \to \exp\big(\mathfrak{e}(n)\big)$. Furthermore, the codomain $\exp\big(\mathfrak{e}(n)\big)$ is an open subset of $GL(n)$, which consists of all the invertible matrices with no eigenvalues in $\mathbb{R}^-$ and contains the open ball $B_1(\mathbf{I})$ such that $B_1(\mathbf{I}) = \{ \mathbf{A} \in GL(n) \mid \|\mathbf{A} - \mathbf{I}\| < 1 \}$, for every matrix norm on M_n.*

Chapter 3

Affine Spaces and Affine Transformations

An affine space is a geometric structure whose introduction is required when one tries to exploit the properties of a vector space to describe geometric entities.

To make clear such a necessity, let us make an example considering the real vector space $\mathbb{R}^3$. Suppose $P_{(0)}$ is the set of real triples $\mathbf{v}_{(0)}$ with 0 as last entry:

$$P_{(0)} = \left\{ \mathbf{v}_{(0)} = (v^1, v^2, 0) \mid v^1, v^2 \in \mathbb{R} \right\} \subset \mathbb{R}^3 .$$

Please notice that the set $P_{(0)}$ is a real vector space, with the operations inherited from $\mathbb{R}^3$ and $\mathbf{o} = (0, 0, 0)$ as null vector. Hence, $P_{(0)}$ is a linear subspace of $\mathbb{R}^3$.

Now, let us denote as $P_{(1)}$ the set of the real triples $\mathbf{v}_{(1)}$ with $v^3 = 1$ as third component (the same applies for any fixed $v^3 \neq 0 \in \mathbb{R}$):

$$P_{(1)} = \left\{ \mathbf{v}_{(1)} = (v^1, v^2, 1) \mid v^1, v^2 \in \mathbb{R} \right\} \subset \mathbb{R}^3 .$$

Even if $P_{(1)}$ is a subset of $\mathbb{R}^3$, it is not a vector space. Actually, one can easily verify that $\mathbf{o} = (0, 0, 0)$ is not in $P_{(1)}$ and that the component-wise addition is not a law of composition for $P_{(1)}$.

$$\mathbf{a}_{(1)} + \mathbf{b}_{(1)} = (a^1, a^2, 1) + (b^1, b^2, 1) = (a^1 + b^1, a^2 + b^2, 2) \notin P_{(1)} .$$

At the same time, from a geometric point of view the sets $P_{(0)}$ and $P_{(1)}$ represent two parallel planes in $\mathbb{R}^3$. Hence, one can expect that they have similar geometric properties, even if they do not have the same algebraic structure.

Such a contradiction depends on the role of the null vector $\mathbf{o}$ in $\mathbb{R}^3$ in identifying a special point of the geometric space, whence comes the necessity to generalize the structure of a vector space to an affine space. As reported in Berger (1987), an affine space is nothing more than a vector space whose origin we try to forget about, by adding translations to the linear maps.

The example here discussed shows that, in order to exploit the algebraic structure of a vector space to describe a geometric space, one should consider two different types of elements, i.e. points and vectors, of which only the latter ones form a linear space. So we can cite Tarrida (2011) in saying that an affine space represents a natural generalization of the concept of vector space but with a clear distinction between points and vectors.

3.1 Affine Spaces

According to Tarrida (2011), we define an affine space as follows.

Definition 3.1. Let $\mathbb{F}$ be a field and $\mathcal{V}$ be an $\mathbb{F}$-vector space. An *affine space over* $\mathcal{V}$ is a set $\mathcal{A}$ together with a map

$$\begin{aligned} \Phi : \mathcal{A} \times \mathcal{V} &\to \mathcal{A} \\ (\mathsf{P}, \mathbf{v}) &\mapsto \mathsf{P} + \mathbf{v} , \end{aligned} \tag{3.1}$$

such that

- $\mathsf{P} + \mathbf{o} = \mathsf{P} , \ \forall \mathsf{P} \in \mathcal{A}$, where $\mathbf{o}$ is the null element in $\mathcal{V}$;
- $\mathsf{P} + (\mathbf{v} + \mathbf{w}) = (\mathsf{P} + \mathbf{v}) + \mathbf{w} , \ \forall \mathsf{P} \in \mathcal{A} , \ \mathbf{v}, \mathbf{w} \in \mathcal{V}$;
- given $\mathsf{P}, \mathsf{Q} \in \mathcal{A}$, there exists a unique $\mathbf{v} \in \mathcal{V}$ such that $\mathsf{P} + \mathbf{v} = \mathsf{Q}$.

The unique vector determined by the points P and Q is denoted by $\overrightarrow{\mathsf{PQ}}$:

$$\mathsf{P} + \overrightarrow{\mathsf{PQ}} = \mathsf{Q}\,.$$

You can observe from the third condition that the symbol "+" is overloaded with multiple meanings, easily resolvable by context. Actually, it refers both to the addition operation between elements of the vector space $\mathcal{V}$ and the result of applying the map $\boldsymbol{\Phi}$.

The *dimension* of an affine space $\mathcal{A}$ is defined to be the dimension of its associated vector space $\mathcal{V}$ and we write $\dim \mathcal{A} = \dim \mathcal{V}$.

Here we summarize the basic properties of an affine space, which are are direct consequences of Definition 3.1. The explicit proof can be found in Tarrida (2011).

Proposition 3.2. *Let* $\mathsf{P}, \mathsf{Q}, \mathsf{R}, \mathsf{S} \in \mathcal{A}$ *and* $\mathbf{u}, \mathbf{v} \in \mathcal{V}$ *be arbitrary points and vectors. Then, the following properties hold:*

1. $\mathsf{P} + \mathbf{u} = \mathsf{P} + \mathbf{v}$ *implies* $\mathbf{u} = \mathbf{v}$*;*
2. $\mathsf{P} + \mathbf{u} = \mathsf{Q} + \mathbf{u}$ *implies* $\mathsf{P} = \mathsf{Q}$*;*
3. $\overrightarrow{\mathsf{PQ}} = \mathbf{o}$ *if, and only if,* $\mathsf{P} = \mathsf{Q}$*;*
4. $\overrightarrow{\mathsf{PQ}} = -\overrightarrow{\mathsf{QP}}$*;*
5. *given* $\mathsf{P} \in \mathcal{A}\,, \mathbf{v} \in \mathcal{V}$*, there exists a unique* $\mathsf{Q} \in \mathcal{A}$ *such that* $\overrightarrow{\mathsf{PQ}} = \mathbf{v}$*;*
6. $\overrightarrow{\mathsf{PQ}} + \overrightarrow{\mathsf{QR}} = \overrightarrow{\mathsf{PR}}$*;*
7. $\overrightarrow{\mathsf{PQ}} = \overrightarrow{\mathsf{PR}}$ *implies* $\mathsf{Q} = \mathsf{R}$*;*
8. $\overrightarrow{\mathsf{PQ}} = \overrightarrow{\mathsf{RS}}$ *implies* $\overrightarrow{\mathsf{PR}} = \overrightarrow{\mathsf{QS}}$*.*

In addition, the transitivity of the action of $\mathcal{V}$ on $\mathcal{A}$ makes the following bijection well-defined for all $\mathsf{P} \in \mathcal{A}$:

$$\begin{aligned} \mu_{\mathsf{P}} : \ & \mathcal{A} \to \mathcal{V} & \qquad \mu_{\mathsf{P}}^{-1} : \ & \mathcal{V} \to \mathcal{A} \\ & \mathsf{Q} \mapsto \mathbf{v} = \overrightarrow{\mathsf{PQ}}\,, & & \mathbf{v} \mapsto \mathsf{Q} = \mathsf{P} + \mathbf{v}\,. \end{aligned} \tag{3.2}$$

Please notice that, as a vector space, $\mathcal{V}$ is a commutative group with respect to the addition. Moreover, the requirements introduced in Definition 3.1 for the map $\boldsymbol{\Phi}$ do satisfy the properties characterizing a simply transitive action of $\mathcal{V}$ on $\mathcal{A}$ (see A.2.2). Explicitly, the first two properties correspond to the identity and compatibility axioms characterizing a group action, and the third one makes such an action simply transitive.

For this reason, some authors formally introduce the definition of affine space in terms of a simply transitive group action of a vector space on a set (see, e.g., Berger (1987)). The permutation representation of such an action is denoted by $t_{\mathbf{v}}$ and is derived from the map Φ by fixing $\mathbf{v} \in \mathcal{V}$:

$$\begin{aligned} t_{\mathbf{v}} :\ & \mathscr{A} \to \mathscr{A} \\ & \mathsf{P} \mapsto t_{\mathbf{v}}(\mathsf{P}) = \Phi(\mathsf{P}, \mathbf{v}) = \mathsf{P} + \mathbf{v}\,. \end{aligned} \tag{3.3}$$

Moreover, the action of $\mathcal{V}$ on $\mathscr{A}$ can also be intended in terms of the following group homomorphism:

$$\begin{aligned} \tau :\ & \mathcal{V} \to \mathrm{Perm}(\mathscr{A}) \\ & \mathbf{v} \mapsto \tau(\mathbf{v}) = t_{\mathbf{v}}\,, \end{aligned}$$

meaning that any vector of $\mathcal{V}$ induces a permutation of the set $\mathscr{A}$.

Definition 3.3. The map $t_{\mathbf{v}} \in \mathrm{Perm}(\mathscr{A})$ expressed by (3.3) is called the *translation* of $\mathscr{A}$ by the vector $\mathbf{v}$.

The image of the homomorphism τ is a subgroup of $\mathrm{Perm}(\mathscr{A})$. It is called the *group of translations* of $\mathscr{A}$ and is denoted as $T(\mathscr{A})$:

$$T(\mathscr{A}) = \mathrm{Im}(\tau) = \{\, t_{\mathbf{v}} \mid \mathbf{v} \in \mathcal{V} \,\} \subseteq \mathrm{Perm}(\mathscr{A})\,. \tag{3.4}$$

Furthermore, by means of (3.3), it is easy to verify that the homomorphism τ satisfies the following properties:

- $t_{\mathbf{o}} = \mathrm{id}_{\mathscr{A}}$;
- $t_{\mathbf{v}} \circ t_{\mathbf{w}} = t_{\mathbf{w}} \circ t_{\mathbf{v}} = t_{\mathbf{v}+\mathbf{w}}\,,\ \ \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{V}$;
- $t_{\mathbf{v}}^{-1} = t_{-\mathbf{v}}\,,\ \ \forall\, \mathbf{v} \in \mathcal{V}$.

In particular, since $t_{\mathbf{v}} = \mathrm{id}_{\mathscr{A}}$ if, and only if, $\mathbf{v} = \mathbf{o}$, the map t has a trivial kernel and it is actually a group monomorphism.

Consequently, by restricting the codomain of τ to $T(\mathscr{A}) \subseteq \mathrm{Perm}(\mathscr{A})$, the resulting homomorphism $\mathcal{V} \to T(\mathscr{A})$ is bijective, and the group of translations of $\mathscr{A}$ is isomorphic with the vector space $\mathcal{V}$:

$$T(\mathscr{A}) \cong \mathcal{V}\,. \tag{3.5}$$

3.1.1 Affine Frames

Definition 3.4. Let $\mathscr{A}$ be an affine space over the $\mathbb{F}$-vector space $\mathcal{V}$. An *affine frame*, or simply a *frame*, of $\mathscr{A}$ is a set $\mathscr{F} = \{\mathsf{O};\mathcal{B}\}$ consisting of a point $\mathsf{O} \in \mathscr{A}$, the *origin* of the frame $\mathscr{F}$, and a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of the associated vector space $\mathcal{V}$.

Please notice that each vector $\mathbf{b}_i \in \mathcal{B}$ defines a translation of the origin O to a point $\mathsf{P}_i \in \mathscr{A}$. Then, a frame can equivalently be defined as a set $\mathscr{F} = \{\mathsf{O};\mathsf{P}_1, \ldots, \mathsf{P}_n\}$ of $n+1$ points of $\mathscr{A}$ such that $\mathcal{B} = \{\overrightarrow{\mathsf{OP}_1}, \ldots, \overrightarrow{\mathsf{OP}_n}\}$ is a basis for the vector space $\mathcal{V}$, with the clear identification $\mathbf{b}_i = \overrightarrow{\mathsf{OP}_i}$.

In any case, if the point O is fixed as the origin of the frame $\mathscr{F}$, the bijective map (3.2), specialized to μ_{O}, allows one to identify each point P of $\mathscr{A}$ with a unique vector of $\mathcal{V}$, that is

$$\mu_{\mathsf{O}}(\mathsf{P}) = \mathbf{p} = \overrightarrow{\mathsf{OP}}\,,$$

and vice-versa

$$\mathsf{P} = \mu_{\mathsf{O}}^{-1}(\mathbf{p}) = \mu_{\mathsf{O}}^{-1}(\overrightarrow{\mathsf{OP}})\,.$$

Moreover, since the basis $\mathcal{B}$ of $\mathcal{V}$ is also introduced, the vector $\mathbf{p}$ can be associated with the n-tuple $\bar{\mathbf{p}} = (p^1, \ldots, p^n) \in \mathbb{F}^n$ through the map $\varphi_{\mathcal{B}}$ defined by (B.1):

$$\bar{\mathbf{p}} = (p^1, \ldots, p^n) = \varphi_{\mathcal{B}}(\mathbf{p}) = \varphi_{\mathcal{B}}(\overrightarrow{\mathsf{OP}})\,.$$

Consequently, the composition of the maps $\varphi_{\mathcal{B}}$ and μ_{O} provides a bijective map $\psi_{\mathscr{F}} = \varphi_{\mathscr{B}} \circ \mu_{\mathsf{O}}$ defined as follows

$$\begin{aligned} \psi_{\mathscr{F}} :\ & \mathscr{A} \to \mathbb{F}^n & \psi_{\mathscr{F}}^{-1} :\ & \mathbb{F}^n \to \mathscr{A} \\ & \mathsf{P} \mapsto \bar{\mathbf{p}} = (p^1, \dots, p^n)\,, & & \bar{\mathbf{p}} \mapsto \mathsf{P} = \mathsf{O} + p^i \mathbf{b}_i\,. \end{aligned} \tag{3.6}$$

Please observe that bijectivity of $\psi_{\mathscr{F}}$ is a consequence of the one of $\varphi_{\mathscr{B}}$ and μ_{O}, so that the inverse map $\psi_{\mathscr{F}}^{-1} = \mu_{\mathsf{O}}^{-1} \circ \varphi_{\mathscr{B}}^{-1}$ is well-defined.

The entries $p_i \in \mathbb{F}$ of $\bar{\mathbf{p}}$ the *affine coordinates*, or simply *coordinates*, of the point P with respect to $\mathscr{F}$.

3.2 Affine Maps

Let $\mathscr{A}_1$ and $\mathscr{A}_2$ be affine spaces overs the $\mathbb{F}$-vector spaces $\mathscr{V}_1$ and $\mathscr{V}_2$, respectively, and consider an arbitrarily fixed point $\mathsf{P} \in \mathscr{A}_1$. Then, every map $f\colon \mathscr{A}_1 \to \mathscr{A}_2$ induces a map between the underlying vector spaces:

$$\begin{aligned} \boldsymbol{f}_{\mathsf{P}} :\ & \mathscr{V}_1 \to \mathscr{V}_2 \\ & \overrightarrow{\mathsf{PQ}} \mapsto \overrightarrow{f(\mathsf{P})f(\mathsf{Q})}\,. \end{aligned} \tag{3.7}$$

Definition 3.5. The map $f\colon \mathscr{A}_1 \to \mathscr{A}_2$ is an *affinity*, also called an *affine map* or an *affine transformation*, if the induced map $\boldsymbol{f}_{\mathsf{P}}$ given by (3.7) is a linear transformation.

Proposition 3.6. *The linear map induced by the affinity $f\colon \mathscr{A}_1 \to \mathscr{A}_2$ does not depend on the point P.*

Proof. Let us apply the map $\boldsymbol{f}_{\mathsf{P}}$, defined by (3.7), to the vectors $\overrightarrow{\mathsf{PQ}}$ and $\overrightarrow{\mathsf{PR}}$, being Q and R arbitrary points of $\mathscr{A}_1$:

$$\begin{aligned} \boldsymbol{f}_{\mathsf{P}}(\overrightarrow{\mathsf{PQ}}) &= \overrightarrow{f(\mathsf{P})f(\mathsf{Q})}\,, \\ \boldsymbol{f}_{\mathsf{P}}(\overrightarrow{\mathsf{PR}}) &= \overrightarrow{f(\mathsf{P})f(\mathsf{R})}\,. \end{aligned}$$

Since the resulting vectors are both elements of $\mathscr{V}_2$, it is possible to add the first one with the opposite of the second one, obtaining

$$\boldsymbol{f}_{\mathsf{P}}(\overrightarrow{\mathsf{PQ}}) - \boldsymbol{f}_{\mathsf{P}}(\overrightarrow{\mathsf{PR}}) = \overrightarrow{f(\mathsf{P})f(\mathsf{Q})} - \overrightarrow{f(\mathsf{P})f(\mathsf{R})}\,,$$

which, for the linearity of $\boldsymbol{f}_{\mathsf{P}}$, results

$$\boldsymbol{f}_{\mathsf{P}}(\overrightarrow{\mathsf{PQ}}) - \boldsymbol{f}_{\mathsf{P}}(\overrightarrow{\mathsf{PR}}) = \boldsymbol{f}_{\mathsf{P}}(\overrightarrow{\mathsf{PQ}} - \overrightarrow{\mathsf{PR}}) \,,$$

so that, by means of Properties 4 and 6 in Proposition 3.2, the following relation is found:

$$\boldsymbol{f}_{\mathsf{P}}(\overrightarrow{\mathsf{RQ}}) = \overrightarrow{f(\mathsf{R})f(\mathsf{Q})} \,, \ \ \forall\, \mathsf{Q}, \mathsf{R} \in \mathscr{A}_1 \,.$$

Hence, the image vector $\overrightarrow{f(\mathsf{R})f(\mathsf{Q})}$ is a function only of the points Q and R, in addition to the map f, and the linear transformation $\boldsymbol{f}_{\mathsf{P}}$ is actually independent of the point P. □

Consistently with the above result, the linear map induced by f is simply denoted as $\boldsymbol{f}$.

The connection between the affinity f and the associated linear transformation $\boldsymbol{f}$ is reflected in the property stated in the following proposition, whose proof can be found, e.g., in Tarrida (2011).

Proposition 3.7. *Let $\mathscr{A}_1$ and $\mathscr{A}_2$ be affine spaces with $\mathcal{V}_1$ and $\mathcal{V}_2$ as associated vector spaces, respectively. The map $f\colon \mathscr{A}_1 \to \mathscr{A}_2$ is an affinity if, and only if, there exists a unique linear map $\boldsymbol{f}\colon \mathcal{V}_1 \to \mathcal{V}_2$ such that*

$$f(\mathsf{P} + \mathbf{v}) = f(\mathsf{P}) + \boldsymbol{f}(\mathbf{v}) \,, \ \ \forall\, \mathsf{P} \in \mathscr{A}_1 \,, \ \mathbf{v} \in \mathcal{V}_1 \,. \tag{3.8}$$

An immediate consequence of such a property is that the identity map $\mathrm{id}_{\mathscr{A}}$ over an affine space $\mathscr{A}$ is an affinity and the induced linear map is the identity $\mathbf{id}_{\mathcal{V}}$ over $\mathcal{V}$:

$$\mathrm{id}_{\mathscr{A}}(\mathsf{P} + \mathbf{v}) = \mathsf{P} + \mathbf{v} = \mathrm{id}_{\mathscr{A}}(\mathsf{P}) + \mathbf{id}_{\mathcal{V}}(\mathbf{v}) \,, \ \ \forall\, \mathsf{P} \in \mathscr{A} \,, \ \mathbf{v} \in \mathcal{V} \,.$$

Also, the linear map $\boldsymbol{f}$ associated with the affinity f inherits the property of being injective, surjective or bijective, and vice-versa (see, e.g., Tarrida (2011)).

Proposition 3.8. *Let f and g be affinities with the induced linear maps $\boldsymbol{f}$ and $\boldsymbol{g}$, respectively.*

1. *If f is composable with g, and $\boldsymbol{f}$ is composable with $\boldsymbol{g}$, then $g \circ f$ is an affinity with $\boldsymbol{gf}$ as associated linear map.*
2. *If f is bijective, the inverse map is itself an affinity and $\boldsymbol{f}^{-1}$ is the induced linear map.*

Proof. 1. Suppose $f : \mathscr{A}_1 \to \mathscr{A}_2$ and $g : \mathscr{A}_2 \to \mathscr{A}_3$, as well as $\boldsymbol{f} : \mathcal{V}_1 \to \mathcal{V}_2$ and $\boldsymbol{g} : \mathcal{V}_2 \to \mathcal{V}_3$. At an arbitrary point $\mathsf{P} \in \mathscr{A}_1$, and for any vector $\mathbf{v} \in \mathcal{V}_1$, the composed map $g \circ f$ gives

$$(g \circ f)(\mathsf{P} + \mathbf{v}) = g\big(f(\mathsf{P} + \mathbf{v})\big) = g\big(f(\mathsf{P}) + \boldsymbol{f}(\mathbf{v})\big) = g\big(f(\mathsf{P})\big) + \boldsymbol{g}\big(\boldsymbol{f}(\mathbf{v})\big)\,,$$

that is

$$(g \circ f)(\mathsf{P} + \mathbf{v}) = (g \circ f)(\mathsf{P}) + (\boldsymbol{gf})(\mathbf{v})\,, \ \ \forall\, \mathsf{P} \in \mathscr{A}_1\,, \ \mathbf{v} \in \mathcal{V}_1\,.$$

Hence, by Proposition 3.7, the composition $g \circ f : \mathscr{A}_1 \to \mathscr{A}_3$ is an affinity and the induced linear map is $\boldsymbol{gf} : \mathcal{V}_1 \to \mathcal{V}_3$.

2. In the same conditions as above, let us set $g = f^{-1}$ and $\boldsymbol{g} = \boldsymbol{f}^{-1}$, with $\mathscr{A}_3 = \mathscr{A}_1$ and $\mathcal{V}_3 = \mathcal{V}_1$, and consider the following identity:

$$f^{-1}\big(f(\mathsf{P} + \mathbf{v})\big) = \mathbf{id}_{\mathscr{A}_1}(\mathsf{P} + \mathbf{v}) = \mathsf{P} + \mathbf{v}\,, \ \ \forall\, \mathsf{P} \in \mathscr{A}_1\,, \ \mathbf{v} \in \mathcal{V}_1\,.$$

Being also $\mathsf{P} = \mathbf{id}_{\mathscr{A}_1}(\mathsf{P})$ and $\mathbf{v} = \mathbf{id}_{\mathcal{V}_1}(\mathbf{v})$, one infers

$$f^{-1}\big(f(\mathsf{P} + \mathbf{v})\big) = f^{-1}\big(f(\mathsf{P})\big) + \boldsymbol{f}^{-1}\big(\boldsymbol{f}(\mathbf{v})\big)\,, \ \ \forall\, f(\mathsf{P}) \in \mathscr{A}_2\,, \ \boldsymbol{f}(\mathbf{v}) \in \mathcal{V}_2\,,$$

and, again by Proposition 3.7, the inverse map $f^{-1} : \mathscr{A}_2 \to \mathscr{A}_1$ results an affinity with the associated linear transformation $\boldsymbol{f}^{-1} : \mathcal{V}_1 \to \mathcal{V}_2$. □

3.2.1 Change of Frame Map

It has been shown in Section 3.1.1 that introducing an affine frame, any point P of the affine space is represented by a set of affine coordinates provided by the map (3.6).

When two frames for an affine space $\mathscr{A}$ are considered, the relevant defining points are connected by an affine transformation of $\mathscr{A}$. The main feature of such a relation is presented in the following proposition.

Proposition 3.9. *Let $\mathscr{A}$ be an affine space on the $\mathbb{F}$-vector space $\mathcal{V}$ and consider two affine frames $\mathscr{F}_{\mathcal{A}} = \{\mathsf{P}_0; \mathsf{P}_1, \ldots, \mathsf{P}_n\}$ and $\mathscr{F}_{\mathcal{B}} = \{\mathsf{Q}_0; \mathsf{Q}_1, \ldots, \mathsf{Q}_n\}$. Then, there exists a unique bijective affine transformation $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}} : \mathscr{A} \to \mathscr{A}$ mapping each point of $\mathscr{F}_{\mathcal{A}}$ to the relevant point of $\mathscr{F}_{\mathcal{B}}$:*

$$\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}(\mathsf{P}_i) = \mathsf{Q}_i\,, \;\; \forall\, i = 1, \ldots, n\,. \tag{3.9}$$

Proof. Let us set $\mathbf{a}_i = \overrightarrow{\mathsf{P}_0\mathsf{P}_i}$ and $\mathbf{b}_i = \overrightarrow{\mathsf{Q}_0\mathsf{Q}_i}$, with $i = 1, \ldots, n$, so that the sets $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ are the bases of $\mathcal{V}$ associated with the frames $\mathscr{F}_{\mathcal{A}}$ and $\mathscr{F}_{\mathcal{B}}$, respectively.

With this specification, the maps $\psi_{\mathcal{A}}$ and $\psi_{\mathcal{B}}$, obtained by specializing of the coordinate map (3.6) to the frames $\mathscr{F}_{\mathcal{A}}$ and $\mathscr{F}_{\mathcal{B}}$, respectively, are characterized by the following property:

$$\psi_{\mathcal{A}}(\mathsf{P}_i) = \bar{\mathbf{e}}_i = (0, \ldots, 1, \ldots, 0) \in \mathbb{F}^n\,, \;\; \forall\, i = 1, \ldots, n\,,$$

as well as

$$\psi_{\mathcal{B}}(\mathsf{Q}_i) = \bar{\mathbf{e}}_i = (0, \ldots, 1, \ldots, 0) \in \mathbb{F}^n\,, \;\; \forall\, i = 1, \ldots, n\,,$$

Then, the map $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ can be obtained by composing $\psi_{\mathcal{B}}^{-1}$ with $\psi_{\mathcal{A}}$, i.e.

$$\mathcal{H}_{\mathcal{A}}^{\mathcal{B}} = \psi_{\mathcal{B}}^{-1} \circ \psi_{\mathcal{A}}\,, \tag{3.10}$$

which clearly satisfies the required property:

$$\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}(\mathsf{P}_i) = (\psi_{\mathcal{B}}^{-1} \circ \psi_{\mathcal{A}})(\mathsf{P}_i) = \psi_{\mathcal{B}}^{-1}\big(\psi_{\mathcal{A}}(\mathsf{P}_i)\big) = \psi_{\mathcal{B}}^{-1}(\bar{\mathbf{e}}_i) = \mathsf{Q}_i\,, \qquad \forall\, i = 1, \ldots, n\,.$$

Please observe that the bijectivity of $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ readily follows from the one of both $\psi_{\mathcal{A}}$ and $\psi_{\mathcal{B}}$.

In addition, since the affine coordinates of any point $\mathsf{P} \in \mathscr{A}$ with respect to a frame are unambiguously determined, both the maps $\psi_{\mathcal{A}}$ and $\psi_{\mathcal{B}}$ are uniquely defined one the frames $\mathscr{F}_{\mathcal{A}}$ and $\mathscr{F}_{\mathcal{B}}$ have been introduced. Consequently, the map $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ is unique.

Moreover, in order to prove that $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ is an affine map of $\mathscr{A}$, let us recall that any coordinate map $\psi_{\mathcal{F}}$ is the composition of the maps defined by (B.1) and (3.2), which specializes to $\psi_{\mathcal{A}} = \varphi_{\mathcal{A}} \circ \mu_{\mathsf{P}_0}$ and $\psi_{\mathcal{B}} = \varphi_{\mathcal{B}} \circ \mu_{\mathsf{Q}_0}$.

Hence, the composition (3.10) can be expressed as

$$\mathcal{H}_{\mathcal{A}}^{\mathcal{B}} = \mu_{\mathsf{Q}_0}^{-1} \circ \left(\varphi_{\mathcal{B}}^{-1} \circ \varphi_{\mathcal{A}}\right) \circ \mu_{\mathsf{P}_0} = \mu_{\mathsf{Q}_0}^{-1} \circ \boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}} \circ \mu_{\mathsf{P}_0} \,,$$

where, recalling (B.5), $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}$ is the change of basis map, from $\mathcal{A}$ to $\mathcal{B}$, for the vector space $\mathcal{V}$.

As a consequence, considering (3.2) specialized to μ_{P_0} and$\mu_{\mathsf{Q}_0}^{-1}$, one has

$$\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}(\mathsf{P}_i) = (\mu_{\mathsf{Q}_0}^{-1} \circ \boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}} \circ \mu_{\mathsf{P}_0})(\mathsf{P}_i) = \mu_{\mathsf{Q}_0}^{-1}\left(\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}(\mathbf{a}_i)\right) = \mathsf{Q}_0 + \boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}(\mathbf{a}_i) \,,$$

or, equivalently,

$$\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}(\mathsf{P}_0 + \mathbf{a}_i) = \mathcal{H}_{\mathcal{A}}^{\mathcal{B}}(\mathsf{P}_0) + \boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}(\mathbf{a}_i) \,, \tag{3.11}$$

whence, recalling Proposition 3.7, $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ is an affinity. □

3.3 The Affine Group

Let us denote as $\mathsf{Aff}(\mathscr{A}) \subseteq \mathsf{Perm}(\mathscr{A})$ the set of permutations of an affine space $\mathscr{A}$ which are also affinities from $\mathscr{A}$ to itself. Such a set, along with the map composition, has a group structure.

The group structure of $\mathsf{Aff}(\mathscr{A})$ comes from the properties of Proposition 3.8. Specifically, composing affinities represents a law of composition for $\mathsf{Aff}(\mathscr{A})$, whose identity element is $\mathbf{id}_{\mathscr{A}}$ and where the inverse affinity f^{-1} is the inverse element of $f \in \mathsf{Aff}(\mathscr{A})$.

Definition 3.10. Let $\mathscr{A}$ be an affine space over the $\mathbb{F}$-vector space $\mathcal{V}$. The set of the bijective affinities of $\mathscr{A}$ into itself is called the *affine group*, or the *group of affinities*, of $\mathscr{A}$ and is denoted as $GA(\mathscr{A})$.

Since any bijective affinity of $\mathscr{A}$ induces an automorphism of $\mathcal{V}$, the following group homomorphism is well-defined (Berger, 1987):

$$\begin{aligned} \lambda :\ GA(\mathscr{A}) &\to GL(\mathcal{V}) \\ f &\mapsto \boldsymbol{f} \,, \end{aligned} \tag{3.12}$$

where $GL(\mathcal{V})$ is the general linear group on the vector space $\mathcal{V}$ introduced by Definition B.57. Actually, the consistency of λ with the group structure of $GA(\mathcal{A})$ and $GL(\mathcal{V})$ is a direct consequence of the first statement in Proposition 3.8:

$$\lambda(g \circ f) = \boldsymbol{g}\boldsymbol{f} = \lambda(g)\lambda(f)\,, \;\; \forall\, f, g \in GA(\mathcal{A})\,.$$

In order to emphasize the role of the composition of affinities as the group law for $GA(\mathcal{A})$, the notation is simplified as follows:

$$gf = g \circ f\,, \;\; \forall\, f, g \in GA(\mathcal{A})\,. \tag{3.13}$$

3.3.1 Subgroups of the Affine Group

A subgroup of the affine group $GA(\mathcal{A})$ is represented exactly by the group of translations $T(\mathcal{A})$ defined by (3.4). This is a consequence of the proposition here below.

Proposition 3.11. *The kernel of the homomorphism* $\lambda\colon GA(\mathcal{A}) \to GL(\mathcal{V})$ *is the group* $T(\mathcal{A})$ *of translations of* $\mathcal{A}$*:*

$$\operatorname{Ker}(\lambda) = T(\mathcal{A})\,.$$

Proof. The kernel of λ is the set of $f \in GA(\mathcal{A})$ mapped to the identity $\mathbf{id}_{\mathcal{V}}$. Hence, setting $\boldsymbol{f} = \mathbf{id}_{\mathcal{V}}$, the characterizing property (3.8) specializes in

$$f(\mathsf{P} + \mathbf{w}) = f(\mathsf{P}) + \mathbf{w}\,, \;\; \forall\, \mathsf{P} \in \mathcal{A}\,, \; \mathbf{w} \in \mathcal{V}\,.$$

On the other hand, we recall that any translation $t_{\mathbf{v}} \in T(\mathcal{A})$, defined by (3.3), is such that

$$t_{\mathbf{v}}(\mathsf{P} + \mathbf{w}) = t_{\mathbf{v}}(t_{\mathbf{w}}(\mathsf{P})) = t_{\mathbf{w}}(t_{\mathbf{v}}(\mathsf{P})) = t_{\mathbf{v}}(\mathsf{P}) + \mathbf{w}\,, \;\; \forall\, \mathsf{P} \in \mathcal{A}\,, \; \mathbf{w} \in \mathcal{V}\,,$$

showing that the map $t_{\mathbf{v}}$ is actually an affinity and $\mathbf{id}_{\mathcal{V}}$ is the associated linear transformation.

In conclusion, the set of translations acting on the affine space $\mathcal{A}$ by the vectors in $\mathcal{V}$ represents precisely the kernel of the group homomorphism λ:

$$\mathrm{Ker}(\lambda) = T(\mathscr{A}) \leq GA(\mathscr{A})\,. \qquad \square$$

Along with the group of translations $T(\mathscr{A})$, it is possible to recognize a further subgroup of $GA(\mathscr{A})$. Such a subgroup is the stabilizer $GA_{\mathsf{P}}(\mathscr{A})$, associated with any point $\mathsf{P} \in \mathscr{A}$.

Specifically, recalling (A.7), the stabilizer of P is the set

$$GA_{\mathsf{P}}(\mathscr{A}) = \{\, f_{\mathsf{P}} \in GA(\mathscr{A}) \mid f_{\mathsf{P}}(\mathsf{P}) = \mathsf{P}\,\} \subset GA(\mathscr{A})\,,$$

whence the following property holds:

$$f_{\mathsf{P}}(\mathsf{P} + \mathbf{v}) = \mathsf{P} + \boldsymbol{f}(\mathbf{v})\,,\;\; \forall\, f_{\mathsf{P}} \in GA_{\mathsf{P}}(\mathscr{A})\,,\; \mathbf{v} \in \mathcal{V}\,, \tag{3.14}$$

being $\boldsymbol{f}$ a linear map in $GL(\mathcal{V})$.

In order to reconstruct the full group $GA(\mathscr{A})$ from its subgroups, it is first shown that $T(\mathscr{A})$ and $GA_{\mathsf{P}}(\mathscr{A})$ are distinct subgroups.

Lemma 3.12. *Let $\mathscr{A}$ be an affine space over the $\mathbb{F}$-vector space $\mathcal{V}$. For any point $\mathsf{P} \in \mathscr{A}$, the stabilizer $GA_{\mathsf{P}}(\mathscr{A})$ and the group of translations $T(\mathscr{A})$ are distinct subgroups of $GA(\mathscr{A})$:*

$$T(\mathscr{A}) \cap GA_{\mathsf{P}}(\mathscr{A}) = \mathbf{id}_{\mathscr{A}}\;.$$

Proof. In order to characterize the intersection of $GA_{\mathsf{P}}(\mathscr{A})$ and $T(\mathscr{A})$, let us consider an affinity f in $T(\mathscr{A})$. Then, by (3.3), there exists a vector $\mathbf{v} \in \mathcal{V}$ such that

$$f(\mathsf{P}) = t_{\mathbf{v}}(\mathsf{P}) = \mathsf{P} + \mathbf{v}\,.$$

Moreover, since it is also required $f \in GA_{\mathsf{P}}(\mathscr{A})$, the same affinity should satisfy the condition $f(\mathsf{P}) = \mathsf{P}$. Hence, one finds

$$\mathsf{P} + \mathbf{v} = \mathsf{P}\,,$$

whence, by Definition 3.1, $\mathbf{v}$ is the null vector $\mathbf{o}$ and the affinity f is the identity map:

$$f = t_{\mathbf{o}} = \mathrm{id}_{\mathscr{A}} \,. \qquad \square$$

It is now possible to show how the affine group of an affine space is actually coincident with the multiplication of the stabilizer of an arbitrary point and the group of translations. This means that any affinity f can be thought as the composition of an affinity f_{P} in the stabilizer of a point P and a translation $t_{\mathbf{v}}$.

Proposition 3.13. *Let $\mathscr{A}$ be an affine space over the $\mathbb{F}$-vector space $\mathscr{V}$ and P an arbitrary point. The group of affinities $GA(\mathscr{A})$ of $\mathscr{A}$ is the group multiplication of the translations of $\mathscr{A}$ and the stabilizer of P:*

$$GA(\mathscr{A}) = T(\mathscr{A})GA_{\mathsf{P}}(\mathscr{A}) \,, \quad \forall\, \mathsf{P} \in \mathscr{A} \,.$$

Proof. Since both the group of translations and the stabilizer of P are subgroups of $GA(\mathscr{A})$, any $t_{\mathbf{v}} \in T(\mathscr{A})$ is in $GA(\mathscr{A})$, and the same applies to any $f_{\mathsf{P}} \in GA_{\mathsf{P}}(\mathscr{A})$. Then, the composition $t_{\mathbf{v}} f_{\mathsf{P}}$ is an element of $GA(\mathscr{A})$ and the following map is well-defined as a bijection:

$$\begin{aligned} \gamma :\ T(\mathscr{A}) \times GA_{\mathsf{P}}(\mathscr{A}) &\to GA(\mathscr{A}) \\ (t_{\mathbf{v}}, f_{\mathsf{P}}) &\mapsto f = t_{\mathbf{v}} f_{\mathsf{P}} \,. \end{aligned} \tag{3.15}$$

In order to prove the injectivity of γ let us consider two element $(t_{\mathbf{v}}, f_{\mathsf{P}})$ and $(t_{\mathbf{w}}, g_{\mathsf{P}})$ of $T(\mathscr{A}) \times GA_{\mathsf{P}}(\mathscr{A})$ and show that

$$\gamma(t_{\mathbf{v}}, f_{\mathsf{P}}) = \gamma(t_{\mathbf{w}}, g_{\mathsf{P}}) \ \Rightarrow \ (t_{\mathbf{v}}, f_{\mathsf{P}}) = (t_{\mathbf{w}}, g_{\mathsf{P}}) \,. \tag{3.16}$$

In fact, by applying (3.15) at both $(t_{\mathbf{v}}, f_{\mathsf{P}})$ and $(t_{\mathbf{w}}, g_{\mathsf{P}})$, one has

$$t_{\mathbf{v}} f_{\mathsf{P}} = t_{\mathbf{w}} g_{\mathsf{P}} \,, \tag{3.17}$$

so that the composition of $t_{\mathbf{v}}^{-1}$ with both members provides

$$f_{\mathsf{P}} = (t_{\mathbf{v}}^{-1} t_{\mathbf{w}}) g_{\mathsf{P}} \,.$$

Since f_{P} is in $GA_{\mathsf{P}}(\mathscr{A})$, it is required that the composition of the translation $t_{\mathbf{v}}^{-1}t_{\mathbf{w}}$ with g_{P} is also in the stabilizer of P. However, by Lemma 3.12, $T(\mathscr{A})$ and $GA_{\mathsf{P}}(\mathscr{A})$ do have no elements in common except the identity map. So the above identity only makes sense if $t_{\mathbf{v}}^{-1}t_{\mathbf{w}} = \mathbf{id}_{\mathscr{A}}$, i.e. $t_{\mathbf{v}} = t_{\mathbf{w}}$.

Similarly, composing both members of (3.17) with f_{P}^{-1}, one can find

$$t_{\mathbf{v}} = t_{\mathbf{w}}(g_{\mathsf{P}}f_{\mathsf{P}}^{-1})\,,$$

which only holds true if $g_{\mathsf{P}}f_{\mathsf{P}}^{-1} = \mathbf{id}_{\mathscr{A}}$, that is $f_{\mathsf{P}} = g_{\mathsf{P}}$.

Then, resulting both $t_{\mathbf{v}} = t_{\mathbf{w}}$ and $f_{\mathsf{P}} = g_{\mathsf{P}}$, the implication in (3.16) is satisfied and the injectivity of γ is proved.

To show the surjectivity of γ, let us fix an arbitrary element $f \in GA(\mathscr{A})$ and verify that there exist some $t \in T(\mathscr{A})$ and $f_{\mathsf{P}} \in GA_{\mathsf{P}}(\mathscr{A})$ providing f through the map (3.15).

Concretely, consider the translation $t_{\mathbf{v}}$ associated with a vector $\mathbf{v} \in \mathcal{V}$ and set $f = t_{\mathbf{v}}f_{\mathsf{P}}$. By composing on the left with $t_{\mathbf{v}}^{-1} = t_{-\mathbf{v}}$, one has

$$f_{\mathsf{P}} = t_{-\mathbf{v}}f \in GA(\mathscr{A})\,.$$

It is worth noting that, since the affine group is closed under map composition and $T(\mathscr{A})$ is a subgroup, the above composition makes sense in $GA(\mathscr{A})$. However, the specific aim is to show in which conditions the affinity f_{P} is in the stabilizer of P. Then, let us set

$$f_{\mathsf{P}}(\mathsf{P}) = (t_{-\mathbf{v}}f)(\mathsf{P}) = \mathsf{P}\,, \tag{3.18}$$

or, explicitly,

$$t_{-\mathbf{v}}\big(f(\mathsf{P})\big) = f(\mathsf{P}) - \mathbf{v} = \mathsf{P}\,.$$

The relation here above can be equivalently expressed as

$$f(\mathsf{P}) = \mathsf{P} + \mathbf{v}\,,$$

and, using the map μ_{P} defined by (3.2), the vector $\mathbf{v}$ defining the translation $t_{\mathbf{v}}$ results

$$\mathbf{v} = \mu_{\mathsf{P}}\big(f(\mathsf{P})\big) = \overrightarrow{\mathsf{P}f(\mathsf{P})} \in \mathcal{V}\,. \tag{3.19}$$

In summary, the existence of the vector $\mathbf{v}$ defined by (3.19) ensures that any affinity f in $GA(\mathscr{A})$ can be expressed as $t_{\mathsf{v}} f_{\mathsf{P}}$, where f_{P} belongs to $GA_{\mathsf{P}}(\mathscr{A})$. Such a decomposition proves the surjectivity of the map γ, and then its bijectivity. □

It is important to remark that the bijection γ introduced by (3.15) relates $T(\mathscr{A}) \times GA_{\mathsf{P}}(\mathscr{A})$ and $GA(\mathscr{A})$ as sets, while no reference has been made to the group structure. Then, by Definition A.10, one can say that $GA(\mathscr{A})$ coincides with the multiplication $T(\mathscr{A})GA_{\mathsf{P}}(\mathscr{A})$ as a set.

3.3.2 Semidirect Product Decomposition

Since the group of translations $T(\mathscr{A})$ is the kernel of the homomorphism λ defined by (3.12), by Proposition A.16 the subgroup $T(\mathscr{A})$ is normal in $GA(\mathscr{A})$:

$$T(\mathscr{A}) \trianglelefteq GA(\mathscr{A}) ,$$

whence, recalling Proposition A.13, the multiplication $T(\mathscr{A})GA_{\mathsf{P}}(\mathscr{A})$ is also a group and coincides with $GA(\mathscr{A})$.

Using the nomenclature of Definition A.24, the stabilizer $GA_{\mathsf{P}}(\mathscr{A})$ is a *complement* for the translations $T(\mathscr{A})$ in the affine group $GA(\mathscr{A})$, and vice-versa.

Moreover, by Theorem A.23, the group of affinities of $\mathscr{A}$ is the semidirect product of $GA_{\mathsf{P}}(\mathscr{A})$ and $T(\mathscr{A})$, up to an isomorphism:

$$GA(\mathscr{A}) \cong T(\mathscr{A}) \rtimes GA_{\mathsf{P}}(\mathscr{A}) . \tag{3.20}$$

Concretely, any map f of $GA(\mathscr{A})$ can be thought as an affinity f_{P} which fixes a point P of $\mathscr{A}$, followed by the a translation t_{v}:

$$f = t_{\mathsf{v}} f_{\mathsf{P}} . \tag{3.21}$$

It is worth noting that the affinities f and f_{P} do have the same associated linear map $\boldsymbol{f}$, that is, in terms of the group homomorphism defined by (3.12),

$$\lambda(f) = \lambda(f_{\mathsf{P}}) = \boldsymbol{f}\,.$$

Then, the decomposition (3.21) of the affinity $f \in GA(\mathscr{A})$ results from fixing a point $\mathsf{P} \in \mathscr{A}$ and obtaining f_{P} as the unique affinity in $GA_{\mathsf{P}}(\mathscr{A})$ with $\boldsymbol{f}$ as associated linear map. Also, using (3.19), the translation $t_{\mathbf{v}}$ is consistently provided by the vector $\mathbf{v} = \overrightarrow{\mathsf{P}f(\mathsf{P})}$.

Conversely, if the vector $\mathbf{v}$ is assigned, P is uniquely determined as the point satisfying $\mathsf{P} = f(\mathsf{P}) - \mathbf{v}$. Then, the affinity f_{P} is consistently obtained as the map in the stabilizer $GA_{\mathsf{P}}(\mathscr{A})$ associated with the linear transformation $\boldsymbol{f}$.

Recalling again Theorem A.23, the composition of two affinities $f = t_{\mathbf{v}} f_{\mathsf{P}}$ and $g = t_{\mathbf{w}} g_{\mathsf{Q}}$ accounts for the action by conjugation of g_{Q} on $t_{\mathbf{v}}$:

$$gf = (t_{\mathbf{w}} g_{\mathsf{Q}})(t_{\mathbf{v}} f_{\mathsf{P}}) = \big(t_{\mathbf{w}}(g_{\mathsf{Q}} t_{\mathbf{v}} g_{\mathsf{Q}}^{-1})\big)(g_{\mathsf{Q}} f_{\mathsf{P}})\,. \tag{3.22}$$

In order to provide a more convenient expression of the composition here above, it is useful to analyze how the elements in the stabilizer $GA_{\mathsf{P}}(\mathscr{A})$ act by conjugation on the translations of $T(\mathscr{A})$.

Proposition 3.14. *Given an affine space $\mathscr{A}$ on the vector space $\mathcal{V}$, let $f_{\mathsf{P}} \in GA_{\mathsf{P}}(\mathscr{A})$ be an affinity in the stabilizer of $\mathsf{P} \in \mathscr{A}$ and let $t_{\mathbf{v}} \in T(\mathscr{A})$ be the translation by the vector $\mathbf{v} \in \mathcal{V}$. Then, the action by conjugation of f_{P} on $t_{\mathbf{v}}$ reads*

$$f_{\mathsf{P}} t_{\mathbf{v}} f_{\mathsf{P}}^{-1} = t_{\boldsymbol{f}(\mathbf{v})}\,. \tag{3.23}$$

where $\boldsymbol{f} \in GL(\mathcal{V})$ is the automorphism associated with the affinity f_{P}.

Proof. Let us evaluate $f_{\mathsf{P}} t_{\mathbf{v}} f_{\mathsf{P}}^{-1}$ at an arbitrary point $\mathsf{Q} \in \mathscr{A}$ to be expressed as $\mathsf{Q} = \mathsf{P} + \mathbf{w}$ for a $\mathbf{w} \in \mathcal{V}$.

Since f_{P} is in the stabilizer of P, the inverse affinity f_{P}^{-1} is also in $GA_{\mathsf{P}}(\mathscr{A})$ and, recalling (3.14), it satisfies

$$f_{\mathsf{P}}^{-1}(\mathsf{Q}) = f_{\mathsf{P}}^{-1}(\mathsf{P} + \mathbf{w}) = \mathsf{P} + \boldsymbol{f}^{-1}(\mathbf{w})\,.$$

Then, the conjugation $f_{\mathsf{P}} t_{\mathbf{v}} f_{\mathsf{P}}^{-1}$ is evaluated as follows:

$$\begin{aligned}(f_{\mathsf{P}} t_{\mathbf{v}} f_{\mathsf{P}}^{-1})(\mathsf{Q}) &= (f_{\mathsf{P}} t_{\mathbf{v}})\big(f_{\mathsf{P}}^{-1}(\mathsf{Q})\big) = (f_{\mathsf{P}} t_{\mathbf{v}})\big(\mathsf{P} + \boldsymbol{f}^{-1}(\mathbf{w})\big) \\ &= f_{\mathsf{P}}\big(\mathsf{P} + \boldsymbol{f}^{-1}(\mathbf{w}) + \mathbf{v}\big) = \mathsf{P} + \boldsymbol{f}\big(\boldsymbol{f}^{-1}(\mathbf{w}) + \mathbf{v}\big) \\ &= \mathsf{P} + \mathbf{w} + \boldsymbol{f}(\mathbf{v}),\end{aligned}$$

where the linearity of $\boldsymbol{f} \in GL(\mathcal{V})$ has been exploited.

The above relation further simplify as

$$(f_{\mathsf{P}} t_{\mathbf{v}} f_{\mathsf{P}}^{-1})(\mathsf{Q}) = t_{\boldsymbol{f}(\mathbf{v})}\big(\mathsf{Q} + \boldsymbol{f}(\mathbf{v})\big) = t_{\boldsymbol{f}(\mathbf{v})}(\mathsf{Q}),$$

whence, for the arbitrariness of Q, one infers

$$(f_{\mathsf{P}} t_{\mathbf{v}} f_{\mathsf{P}}^{-1}) = t_{\boldsymbol{f}(\mathbf{v})}. \qquad \square$$

The result of Proposition 3.14 can be applied for a useful simplification of the composition rule (3.22), which can be expressed as

$$gf = \big(t_{\mathbf{w}} t_{\boldsymbol{g}(\mathbf{v})}\big)(g_{\mathsf{P}} f_{\mathsf{P}}) = t_{\mathbf{w}+\boldsymbol{g}(\mathbf{v})}(g_{\mathsf{P}} f_{\mathsf{P}}), \tag{3.24}$$

where, being $\boldsymbol{g}$ the linear map associated with the affinity g_{P}, the relation (3.23) has been exploited.

Along with the identification (3.5) of the group $T(\mathscr{A})$ with the vector space $\mathcal{V}$, a further useful identification relates the stabilizer of any point $\mathsf{P} \in \mathscr{A}$ with the general linear group of $\mathcal{V}$.

Proposition 3.15. *Let us fix an arbitrary point* $\mathsf{P} \in \mathscr{A}$. *The stabilizer* $GA_{\mathsf{P}}(\mathscr{A})$ *of* P *is isomorphic with* $GL(\mathcal{V})$*:*

$$GA_{\mathsf{P}}(\mathscr{A}) \cong GL(\mathcal{V}), \quad \forall\, \mathsf{P} \in \mathscr{A}. \tag{3.25}$$

Proof. Let $\lambda_{\mathsf{P}}\colon GA_{\mathsf{P}}(\mathscr{A}) \to GL(\mathcal{V})$ be the restriction of the group homomorphism (3.12) at the stabilizer $GA_{\mathsf{P}}(\mathscr{A})$.

In order to prove the surjectivity of λ_{P}, let us show that for any isomorphism in $GL(\mathcal{V})$ there exist some affinities in the stabilizer of P.

Consider a linear transformation $\boldsymbol{f} \in GL(\mathcal{V})$ and compose with the bijection μ_{P} defined by (3.2):

$$\begin{aligned} \boldsymbol{f} \circ \mu_{\mathsf{P}} :\ & \mathscr{A} \to \mathcal{V} \\ & \mathsf{Q} \mapsto \boldsymbol{f}(\overrightarrow{\mathsf{PQ}}) \,. \end{aligned}$$

The further composition with μ_{P}^{-1} provides the map $h = \mu_{\mathsf{P}}^{-1} \circ \boldsymbol{f} \circ \mu_{\mathsf{P}}$ from $\mathscr{A}$ to itself, with the following rule:

$$f(\mathsf{Q}) = \mathsf{P} + \boldsymbol{f}(\overrightarrow{\mathsf{PQ}}) \,, \ \ \forall\, \mathsf{Q} \in \mathscr{A} \,.$$

Moreover, since any point $\mathsf{Q} \in \mathscr{A}$ can be expressed as the translation $\mathsf{P} + \overrightarrow{\mathsf{PQ}}$, the map f satisfies the property

$$f(\mathsf{P} + \overrightarrow{\mathsf{PQ}}) = \mathsf{P} + \boldsymbol{f}(\overrightarrow{\mathsf{PQ}}) \,, \ \ \forall\, \overrightarrow{\mathsf{PQ}} \in \mathcal{V} \,,$$

whence, by Proposition 3.7, f is clearly an affinity in $GA(\mathscr{A})$ with $\boldsymbol{f}$ as associated linear map.

Moreover, since $f(\mathsf{P}) = \mathsf{P}$, the affinity f is in the stabilizer of P. Then, the preimage of $GL(\mathcal{V})$ under λ_{P} is exactly $GA_{\mathsf{P}}(\mathscr{A})$ and the map λ_{P} is surjective.

Please observe that λ_{P} is also injective because its kernel is trivial. In fact, by (3.14), the affinities mapped to $\mathbf{id}_{\mathcal{V}}$ by λ_{P} satisfy the condition

$$f(\mathsf{P} + \mathbf{v}) = \mathsf{P} + \mathbf{v} \,, \ \ \forall\, \mathbf{v} \in \mathcal{V} \,,$$

which holds true, for the arbitrariness of $\mathbf{v}$, only if $f = \mathrm{id}_{\mathscr{A}}$. □

As a consequence of the isomorphisms (3.5) and (3.25), and expressing $GA(\mathscr{A})$ as the semidirect product given by (3.20), the following isomorphism applies:

$$GA(\mathscr{A}) \cong \mathcal{V} \rtimes GL(\mathcal{V}) \,, \tag{3.26}$$

so that each affinity $f \in GA(\mathscr{A})$ can be represented as

$$f \cong (\mathbf{v}, \boldsymbol{f}) \,, \tag{3.27}$$

where $\mathbf{v} \in \mathcal{V}$ represents a translation $t_{\mathbf{v}}$ and $\boldsymbol{f} \in GL(\mathcal{V})$ is the linear map associated with f.

Moreover, consistently with the semidirect product in (3.26), and recalling (3.24), the following composition rule applies:

$$gf \cong (\mathbf{w}, \boldsymbol{g})(\mathbf{v}, \boldsymbol{f}) = \big(\mathbf{w} + \boldsymbol{g}(\mathbf{v}), \boldsymbol{gf}\big)\,. \tag{3.28}$$

3.4 Matrix Representation of Affinities

Please recall that introducing an frame $\mathcal{F}$ for an affine space $\mathcal{A}$ it is possible to represent each point $\mathsf{P} \in \mathcal{A}$ by an n-tuple $\bar{\mathbf{p}} \in \mathbb{F}^n$, given by the bijective map (3.6).

Moreover, by assembling the affine coordinates $(p^1, \ldots, p^n)$ of P in a column matrix $[\mathbf{p}]$, it is desirable that the transformation $f\colon \mathsf{P} \mapsto \mathsf{Q} = f(\mathsf{P})$ could be represented by a matrix multiplication in the form

$$[\mathbf{q}] = [\mathbf{L}][\mathbf{p}]\,,$$

just as linear maps do transform vectors, expressed in terms of coordinates, through (B.109).

However, even if each affine transformation $f \in GA(\mathcal{A})$ is associated with a linear map $\boldsymbol{f} \in GL(\mathcal{V})$, the matrix representation of f is not straightforward.

In fact, the group homomorphism $\lambda\colon GA(\mathcal{A}) \to GL(\mathcal{V})$, introduced by (3.12), is not an isomorphism. Specifically, λ is not injective, and its kernel is exactly the group of translations $T(\mathcal{A})$ (cf. Proposition 3.11).

Consequently, since there exists no linear map in $GL(\mathcal{V})$, other the identity $\mathbf{id}_{\mathcal{V}}$, to be associated with a translation, no affine transformation including a translation can be represented by an n-by-n matrix.

To clarify the observation here above, please observe that when the translation component of an affinity f vanishes, the representation (3.27) reads $f \cong (\mathbf{o}, \boldsymbol{f})$. Then, the transformation $\mathsf{Q} = f(\mathsf{P})$ can be expressed with respect to the frame $\mathscr{F} = \{\mathsf{O}; \mathcal{B}\}$ as

$$[\mathbf{q}] = [\mathbf{A}][\mathbf{p}]\,, \tag{3.29}$$

where $[\mathbf{p}] \in M_{n\times 1}$ and $[\mathbf{q}] \in M_{n\times 1}$ are the matrix representations of the vectors $\overrightarrow{\mathsf{OP}}$ and $\overrightarrow{\mathsf{OQ}}$, respectively, and $[\mathbf{A}] \in M_n$ is the matrix representation of the linear map $\boldsymbol{f}$. Formally, $[\mathbf{p}]$ and $[\mathbf{q}]$ are provided by the map $\theta_{\mathcal{B}}$ defined by (B.110), and $[\mathbf{A}] = \Psi_{(\mathcal{B},\mathcal{B})}(\boldsymbol{f})$ is obtained by specializing the map (B.107).

On the contrary, when the affinity f is a pure translation, the isomorphism (3.27) becomes $f \cong (\mathbf{v}, \boldsymbol{o})$, and the matrix representation of the transformation $\mathsf{Q} = f(\mathsf{P})$ results

$$[\mathbf{q}] = [\mathbf{p}] + [\mathbf{v}]\,, \tag{3.30}$$

where, in addition to $[\mathbf{p}]$ and $[\mathbf{q}]$, the matrix representation $[\mathbf{v}] \in M_{n\times 1}$ of the translation vector $\mathbf{v}$ has been considered.

Please observe that, as opposed to (3.29), the sum of column matrices in (3.30), representing the translation by $\mathbf{v}$, is not a linear transformation of $[\mathbf{p}]$.

3.4.1 Homogeneous Representation

With the aim to provide a matrix representation of a translation map consistent with a linear transformation, the homogeneous coordinates of affine points are traditionally introduced.

Definition 3.16. Let $\mathscr{A}$ be an affine space on the $\mathbb{F}$-vector space $\mathcal{V}$ and let $\mathscr{F} = \{\mathsf{O}; \mathcal{B}\}$ be an affine frame. The *homogeneous representation of a point* $\mathsf{P} \in \mathscr{A}$, with respect to the frame $\mathscr{F}$, is the column matrix $[\mathbf{p}]_{\mathsf{h}} \in M_{n+1\times 1}$ such that the first n entries are the affine coordinates $p^1, \ldots, p^n$ of the point P and the last one is 1:

$$[\mathbf{p}]_{\mathsf{h}} = [p^1 \ \cdots \ p^n \ 1]^{\mathsf{T}} = \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix}. \tag{3.31}$$

Please notice that, with a slight abuse of notation, in place of the matrix symbol $[\mathbf{p}]$, the vector $\mathbf{p}$ has been used to denote the n-by-1 submatrix of

$[\mathbf{p}]_{\mathsf{h}}$.

The benefit of expressing the coordinates of affine points in the form (3.31) is that the transformation (3.30) can be consistently represented as

$$\begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{v} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix}, \tag{3.32}$$

where $\mathbf{o}^\mathsf{T} = [\mathbf{o}]^\mathsf{T}$ is the 1-by-n submatrix with null entries, so that the translation map $(\mathbf{v}, \boldsymbol{o})$ takes the form of a linear transformation on the space of the $(n+1)$-by-1 matrices.

In addition, the matrix operation (3.29), representing the affine transformation $(\mathbf{o}, \boldsymbol{f})$, can itself be expressed with the formalism of the homogeneous coordinates as

$$\begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{o} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix}. \tag{3.33}$$

In conclusion, exploiting the homogeneous coordinate representation, both (3.32) and (3.33) can be combined in a unique linear transformation for a subset of $M_{n+1\times 1}$, representing the affine map $f \cong (\mathbf{v}, \boldsymbol{f})$.

Definition 3.17. Given the affine space $\mathscr{A}$ associated with the $\mathbb{F}$-vector space $\mathcal{V}$, let $f \cong (\mathbf{v}, \boldsymbol{f})$ be an affine map. The *homogeneous representation of the affine transformation* $f \in GA(\mathscr{A})$, with respect to a frame $\mathscr{F} = \{\mathsf{O}; \mathcal{B}\}$, is the square matrix $[\mathbf{L}]_{\mathsf{h}} \in M_{n+1}$ in the form

$$[\mathbf{L}]_{\mathsf{h}} = \begin{bmatrix} \mathbf{A} & \mathbf{v} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix}, \tag{3.34}$$

where the n-by-n submatrix $\mathbf{A} = [\mathbf{A}]$ is the matrix representation of the linear map $\boldsymbol{f} \in GL(\mathcal{V})$ associated with f.

3.4.2 Change of Frame Matrix

Let us consider two affine frames $\mathscr{F}_{\mathcal{A}}$ and $\mathscr{F}_{\mathcal{B}}$ for the affine space $\mathscr{A}$, which are assigned as $\mathscr{F}_{\mathcal{A}} = \{\mathsf{P}_0; \mathsf{P}_1, \ldots, \mathsf{P}_n\}$ and $\mathscr{F}_{\mathcal{B}} = \{\mathsf{Q}_0; \mathsf{Q}_1, \ldots, \mathsf{Q}_n\}$, respectively.

Equivalently, the affine frames can be defined as $\mathscr{F}_{\mathcal{A}} = \{\mathsf{P}_0; \mathcal{A}\}$ and $\mathscr{F}_{\mathcal{B}} =$

$\{\mathsf{Q}_0; \mathcal{B}\}$, where the bases of vectors $\mathcal{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ result from setting $\mathbf{a}_i = \overrightarrow{\mathsf{P}_0\mathsf{P}_i}$ and $\mathbf{b}_i = \overrightarrow{\mathsf{Q}_0\mathsf{Q}_i}$, with $i = 1, \dots, n$.

Recalling Proposition 3.9, the change of frame from $\mathscr{F}_{\mathcal{A}}$ to $\mathscr{F}_{\mathcal{B}}$ is represented by a map $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}} : \mathscr{A} \to \mathscr{A}$, which is a bijective affinity of $\mathscr{A}$ and hence is in $GA(\mathscr{A})$.

Then, the decomposition (3.21) can be applied and $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ can be intended as an affinity h_{P}, which stabilizes a point P, followed by a translation $t_{\mathbf{v}}$:

$$\mathcal{H}_{\mathcal{A}}^{\mathcal{B}} = t_{\mathbf{v}} h_{\mathsf{P}} \,, \tag{3.35}$$

where one between the vector $\mathbf{v}$ or the fixed point P should be chosen, while the other object is consistently derived.

Since the transformation $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ maps the origin of $\mathscr{F}_{\mathcal{A}}$ to the one of $\mathscr{F}_{\mathcal{B}}$, it is convenient to make the vector $\mathbf{v}$ represent the translation from P_0 to Q_0, defining

$$\mathbf{v} = \overrightarrow{\mathsf{P}_0\mathsf{Q}_0} = \mathbf{v}_{\mathcal{A}}^{\mathcal{B}} \,, \tag{3.36}$$

and the fixed point P is obtained as the point of $\mathscr{A}$ satisfying the condition $\mathsf{P} = \mathcal{H}_{\mathcal{A}}^{\mathcal{B}}(\mathsf{P}) - \mathbf{v}$, that is

$$\mathsf{P} = \mathcal{H}_{\mathcal{A}}^{\mathcal{B}}(\mathsf{P}) - \overrightarrow{\mathsf{P}_0\mathsf{Q}_0} = \mathsf{P}_0 \,. \tag{3.37}$$

In conclusion, the change of frame map $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ is decomposed as

$$\mathcal{H}_{\mathcal{A}}^{\mathcal{B}} = t_{\mathbf{v}_{\mathcal{A}}^{\mathcal{B}}} h_{\mathsf{P}_0} \,, \tag{3.38}$$

where $h_{\mathsf{P}_0} \in GA_{\mathsf{P}_0}(\mathscr{A})$ is the affinity of $\mathscr{A}$ which fixes the origin P_0 of $\mathscr{F}_{\mathcal{A}}$, and $t_{\mathbf{v}_{\mathcal{A}}^{\mathcal{B}}}$ is the translation map from P_0 to Q_0.

In addition, recalling (3.11), the linear map associated with $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ is $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}$, which is the change of basis map from $\mathcal{A}$ to $\mathcal{B}$, so that, by (3.27), the change of frame can be represented as

$$\mathcal{H}_{\mathcal{A}}^{\mathcal{B}} \cong (\mathbf{v}_{\mathcal{A}}^{\mathcal{B}}, \boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}) \,. \tag{3.39}$$

Consequently, applying Definition 3.17, the homogeneous representation of $\mathcal{H}_{\mathcal{A}}^{\mathcal{B}}$ is

$$[\mathbf{H}_{\mathcal{A}}^{\mathcal{B}}]_{\mathsf{h}} = \begin{bmatrix} \mathbf{C}_{\mathcal{A}}^{\mathcal{B}} & \mathbf{v}_{\mathcal{A}}^{\mathcal{B}} \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix}, \tag{3.40}$$

where $\mathbf{C}_{\mathcal{A}}^{\mathcal{B}} = [\mathbf{C}_{\mathcal{A}}^{\mathcal{B}}]$ is the active matrix of change of basis for $\mathcal{V}$, specified by (B.118). Accordingly, $[\mathbf{H}_{\mathcal{A}}^{\mathcal{B}}]_{\mathsf{h}}$ is the *active* matrix of change of frame from $\mathcal{F}_{\mathcal{A}}$ to $\mathcal{F}_{\mathcal{B}}$, with the role to describe how the affine coordinates of a point, relevant to $\mathcal{F}_{\mathcal{A}}$, do change when it is mapped to a point associated with $\mathcal{F}_{\mathcal{B}}$.

Now, consider a point $\mathsf{X} \in \mathscr{A}$ whose coordinate vectors, with respect to $\mathcal{F}_{\mathcal{A}}$ and $\mathcal{F}_{\mathcal{B}}$, are $\bar{\mathbf{x}}_{\mathcal{A}}$ and $\bar{\mathbf{x}}_{\mathcal{B}}$, respectively. Then, the coordinates of X are such that

$$\mathsf{X} = \mathsf{P}_0 + x_{\mathcal{A}}^i \mathbf{a}_i = \mathsf{Q}_0 + x_{\mathcal{B}}^k \mathbf{b}_k ,$$

whence, using (3.36) to express $\mathsf{Q}_0 = \mathsf{P}_0 + \mathbf{v}_{\mathcal{A}}^{\mathcal{B}}$, and also recalling (B.118), one finds

$$\mathsf{P}_0 + x_{\mathcal{A}}^i \mathbf{a}_i = \mathsf{P}_0 + x_{\mathcal{B}}^k (C_{\mathcal{A}}^{\mathcal{B}})^i{}_k \, \mathbf{a}_i + (v_{\mathcal{A}}^{\mathcal{B}})^i \mathbf{a}_i .$$

The identity here above implies $x_{\mathcal{A}}^i = (C_{\mathcal{A}}^{\mathcal{B}})^i{}_k \, x_{\mathcal{B}}^k + (v_{\mathcal{A}}^{\mathcal{B}})^i$, which can be expressed in matrix form, using the homogeneous representation, as

$$\begin{bmatrix} \mathbf{x}_{\mathcal{A}} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\mathcal{A}}^{\mathcal{B}} & \mathbf{v}_{\mathcal{A}}^{\mathcal{B}} \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\mathcal{B}} \\ 1 \end{bmatrix} \quad \Leftrightarrow \quad [\mathbf{x}_{\mathcal{A}}]_{\mathsf{h}} = [\mathbf{H}_{\mathcal{A}}^{\mathcal{B}}]_{\mathsf{h}} [\mathbf{x}_{\mathcal{B}}]_{\mathsf{h}} ,$$

or equivalently

$$[\mathbf{x}_{\mathcal{B}}]_{\mathsf{h}} = [\mathbf{H}_{\mathcal{B}}^{\mathcal{A}}]_{\mathsf{h}} [\mathbf{x}_{\mathcal{A}}]_{\mathsf{h}} . \tag{3.41}$$

The matrix $[\mathbf{H}_{\mathcal{B}}^{\mathcal{A}}]_{\mathsf{h}} = [\mathbf{H}_{\mathcal{A}}^{\mathcal{B}}]_{\mathsf{h}}^{-1}$ is the *passive* matrix of change of frame from $\mathcal{F}_{\mathcal{A}}$ to $\mathcal{F}_{\mathcal{B}}$, which transforms, for a fixed point, the homogeneous coordinates with respect to $\mathcal{F}_{\mathcal{A}}$ to the ones with respect to $\mathcal{F}_{\mathcal{B}}$.

Explicitly, the inversion of the matrix $[\mathbf{H}_{\mathcal{A}}^{\mathcal{B}}]_{\mathsf{h}}$ provides

$$[\mathbf{H}_{\mathcal{B}}^{\mathcal{A}}]_{\mathsf{h}} = \begin{bmatrix} \mathbf{C}_{\mathcal{B}}^{\mathcal{A}} & \mathbf{v}_{\mathcal{B}}^{\mathcal{A}} \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\mathcal{A}}^{\mathcal{B}\,-1} & -\mathbf{C}_{\mathcal{A}}^{\mathcal{B}\,-1} \mathbf{v}_{\mathcal{A}}^{\mathcal{B}} \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix}, \tag{3.42}$$

which represents both the passive change of frame from $\mathscr{F}_{\mathcal{A}}$ to $\mathscr{F}_{\mathcal{B}}$ and the active transformation from $\mathscr{F}_{\mathcal{B}}$ to $\mathscr{F}_{\mathcal{A}}$.

Chapter 4

Euclidean Spaces and Rigid Motions

A Euclidean affine space is associated with a Euclidean vector space. For this reason, it is worth recalling in which sense a Euclidean vector space is a specialization of general linear spaces, and how this reflects on the characterization of Euclidean affine spaces.

Remark (Notation)**.** By virtue of the identification (B.111) discussed in Appendix B.4.1, the same symbol $\mathbf{v}$ will be used to denote both an element of a vector space or its matrix representation $[\mathbf{v}]$ with respect to a fixed basis. Similarly, on the basis of (B.108), the symbol of a tensor $\mathbf{A}$ will be used to denote both the associated linear map $\boldsymbol{f}$ and its matrix representation $[\mathbf{A}]$, and an evaluation $\mathbf{w} = \boldsymbol{f}(\mathbf{v})$ will be denoted by juxtaposition as

$$\mathbf{w} = \mathbf{A}\mathbf{v} \quad \Leftrightarrow \quad [\mathbf{w}] = [\mathbf{A}][\mathbf{v}]\,.$$

4.1 Euclidean Vector Space

Definition 4.1. A vector space $\mathcal{E}$ is said a *Euclidean vector space* if it is defined over the real field $\mathbb{R}$ and is endowed with a positive definite scalar product.

Please notice that the standard dot product $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$, introduced by (B.84), is positive definite. Then, $\mathbb{R}^n$ is the reference Euclidean vector space.

In a Euclidean vector space $\mathcal{E}$, the associated scalar product naturally induces the following norm:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}\,, \ \ \forall\, \mathbf{v} \in \mathcal{E}\,, \tag{4.1}$$

which is called the *Euclidean norm*.

It is easy to recognize that the Euclidean norm is consistent with the general notion of a *norm* as introduced by Definition B.50. Actually, since the scalar product of a Euclidean vector space is positive definite, the positivity condition is trivially satisfied.

In addition, the linearity of a scalar product implies

$$\|c\mathbf{v}\| = \sqrt{\langle c\mathbf{v}, c\mathbf{v} \rangle} = \sqrt{c^2 \langle \mathbf{v}, \mathbf{v} \rangle} = |c|\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = |c|\,\|\mathbf{v}\|\,, \ \ \forall\, \mathbf{v} \in \mathcal{E}, c \in \mathbb{R}\,, \ \ ,$$

which corresponds to the second condition of Definition B.50.

Finally, since the scalar product is bilinear and symmetric, one can verify that also the triangle inequality is satisfied:

$$\begin{aligned} \|\mathbf{v} + \mathbf{v}'\| &= \sqrt{\langle \mathbf{v} + \mathbf{v}', \mathbf{v} + \mathbf{v}' \rangle} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{v}', \mathbf{v}' \rangle + 2\langle \mathbf{v}, \mathbf{v}' \rangle} \\ &\leq \sqrt{\|\mathbf{v}\|^2 + \|\mathbf{v}'\|^2 + 2\|\mathbf{v}\|\,\|\mathbf{v}'\|} \\ &= \|\mathbf{v}\| + \|\mathbf{v}'\|\,, \end{aligned}$$

where the Cauchy-Schwartz' inequality (B.93) has been applied.

4.1.1 Orthonormal Bases

The Euclidean norm (4.1) allows one to recognize a *unit vector* as a vector $\mathbf{u}$ such that $\|\mathbf{u}\| = 1$. It is straightforward to see that for any non-null vector $\mathbf{v}$ of $\mathcal{V}$, there exists an associated unit vector given by $\mathbf{v}/\|\mathbf{v}\|$.

Definition 4.2. Let $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for a Euclidean vector space $\mathcal{E}$. We say that $\mathcal{B}$ is an *orthonormal basis*, with respect to the associated scalar product, if it is orthogonal and any vector in $\mathcal{B}$ has a unitary norm.

The orthogonality condition is defined by (B.86) and implies $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ for any pair of distinct vectors of an basis. At the same time, by (4.1), any vector $\mathbf{e}_i$ is such that $\langle \mathbf{e}_i, \mathbf{e}_i \rangle = 1$. Both properties can be summarized by the following condition:

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}\,, \tag{4.2}$$

where δ_{ij} is the Kronecker symbol:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j\,, \\ 0 & \text{if } i \neq j\,. \end{cases} \tag{4.3}$$

Moreover, by the linearity of the scalar product, one also has

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle v^i \mathbf{e}_i, w^j \mathbf{e}_j \rangle = v^i w^j \langle \mathbf{e}_i, \mathbf{e}_j \rangle = v^i w^j \delta_{ij}\,, \ \ \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{E}\,, \tag{4.4}$$

or, explicitly,

$$\langle \mathbf{v}, \mathbf{w} \rangle = u^1 v^1 + \cdots + u^n v^n\,, \ \ \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{E}\,,$$

whence, by virtue of (B.84), the scalar product of $\mathbf{v}$ and $\mathbf{w}$ equals the dot product of the coordinate vectors $\bar{\mathbf{v}}$ and $\overline{\mathbf{w}}$:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \bar{\mathbf{v}}, \overline{\mathbf{w}} \rangle_{\mathbb{R}^n} = \bar{\mathbf{v}} \cdot \overline{\mathbf{w}}\,, \ \ \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{E}\,. \tag{4.5}$$

For this reason, one can simply write

$$\mathbf{v} \cdot \mathbf{w} = \langle \mathbf{v}, \mathbf{w} \rangle\,, \ \ \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{E}\,, \tag{4.6}$$

meaning that the dot product of $\mathbf{v}$ and $\mathbf{w}$ in a Euclidean vector space is coincident with the standard dot product of the coordinate vectors relevant to a orthonormal basis.

Remark. In identifying an endomorphism $\boldsymbol{f}$ with a tensor $\mathbf{A}$, the equivalence is given specializing (B.103) as

$$\boldsymbol{f}(\mathbf{v}) = \mathbf{A}\mathbf{v} = A^i{}_k \mathbf{e}^k(\mathbf{v}) \mathbf{e}_i = A^i{}_k v^k \mathbf{e}_i\,, \tag{4.7}$$

which is derived considering $\mathbf{A} = A^i{}_k \mathbf{e}_i \otimes \mathbf{e}^k$ a mixed tensor in $\mathcal{T}^{(1,1)}(\mathcal{E})$.

At the same time, supposing the vector $\mathbf{Av}$ is involved the scalar product with a vector $\mathbf{w}$, one has

$$\langle \mathbf{Av}, \mathbf{w} \rangle = A^i{}_k v^k w^j \langle \mathbf{e}_i, \mathbf{e}_j \rangle \,,$$

where, if the basis of $\mathcal{E}$ is orthonormal, the condition $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$ implies the simplification

$$\langle \mathbf{Av}, \mathbf{w} \rangle = A^i{}_k v^k w^j \delta_{ij} = A_{jk} v^k w^j \,.$$

that is $\mathbf{A} = A_{jk} \mathbf{e}^i \otimes \mathbf{e}^j$ should be intended as a covariant tensor in $\mathcal{T}^{(0,2)}(\mathcal{E})$.

The apparent contradiction is actually resolved by the self-duality of the Euclidean vector space $\mathcal{E}$ (cf. Appendix B.3.2), which allows the identification $\mathcal{E}^* \cong \mathcal{E}$ and also $\mathcal{E} \otimes \mathcal{E}^* \cong \mathcal{E}^* \otimes \mathcal{E}^*$.

In addition, if the basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal, the coefficients of $\mathbf{A}$ considered both as a mixed tensor or a covariant one do coincide:

$$A^i{}_j = A_{ij} \,.$$

4.1.2 Linear Isometries

Definition 4.3. Let $\mathcal{E}$ and $\mathcal{E}'$ Euclidean vector spaces and let $\mathbf{Q} \colon \mathcal{E} \to \mathcal{E}'$ be a map. We say that $\mathbf{Q}$ is an *isometry* if it takes the scalar product of $\mathcal{E}$ to the one of $\mathcal{E}'$:

$$\langle \mathbf{Qu}, \mathbf{Qv} \rangle_{\mathcal{E}'} = \langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{E}} \,, \ \ \forall\, \mathbf{v}, \mathbf{v}' \in \mathcal{E} \,. \tag{4.8}$$

Please observe that any map preserving the scalar product is linear, as stated in the following proposition.

Proposition 4.4. *Any isometry $\mathbf{Q} \colon \mathcal{E} \to \mathcal{E}'$ between the Euclidean vector spaces $\mathcal{E}$ and $\mathcal{E}'$ is a linear transformation.*

Proof. Consider arbitrary vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ of $\mathcal{E}$. Then, exploiting the linearity of the scalar product both in $\mathcal{E}$ and $\mathcal{E}'$, the isometry $\mathbf{Q}$ satisfies

$$\begin{aligned}\langle \mathbf{Q}\mathbf{u} + \mathbf{Q}\mathbf{v}, \mathbf{Q}\mathbf{w}\rangle_{\mathcal{E}'} &= \langle \mathbf{Q}\mathbf{u}, \mathbf{Q}\mathbf{w}\rangle_{\mathcal{E}'} + \langle \mathbf{Q}\mathbf{v}, \mathbf{Q}\mathbf{w}\rangle_{\mathcal{E}'} \\ &= \langle \mathbf{u}, \mathbf{w}\rangle_{\mathcal{E}} + \langle \mathbf{v}, \mathbf{w}\rangle_{\mathcal{E}} = \langle \mathbf{u} + \mathbf{v}, \mathbf{w}\rangle_{\mathcal{E}} \\ &= \langle \mathbf{Q}(\mathbf{u} + \mathbf{v}), \mathbf{Q}\mathbf{w}\rangle_{\mathcal{E}'} ,\end{aligned}$$

whence, for the arbitrariness of $\mathbf{w}$, one derives

$$\mathbf{Q}(\mathbf{u} + \mathbf{v}) = \mathbf{Q}\mathbf{u} + \mathbf{Q}\mathbf{v} \;\; \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{E} .$$

Similarly, considering the multiplication by an arbitrary scalar, one has

$$\begin{aligned}\langle c\mathbf{Q}\mathbf{v}, \mathbf{Q}\mathbf{w}\rangle_{\mathcal{E}'} &= c\langle \mathbf{Q}\mathbf{v}, \mathbf{Q}\mathbf{w}\rangle_{\mathcal{E}'} = c\langle \mathbf{v}, \mathbf{w}\rangle_{\mathcal{E}} = \langle c\mathbf{v}, \mathbf{w}\rangle_{\mathcal{E}} \\ &= \langle \mathbf{Q}(c\mathbf{v}), \mathbf{Q}\mathbf{w}\rangle_{\mathcal{E}'} ,\end{aligned}$$

obtaining

$$\mathbf{Q}(c\mathbf{v}) = c\mathbf{Q}\mathbf{v} \;\; \forall\, \mathbf{v} \in \mathcal{E} ,\; c \in \mathbb{R} .$$

Then, by Definition B.8, the isometry $\mathbf{Q}$ il linear. □

Please notice that if the map $\mathbf{Q}$ is an isometry, the orthogonality of vectors is preserved:

$$\langle \mathbf{Q}\mathbf{u}, \mathbf{Q}\mathbf{v}\rangle_{\mathcal{E}'} = \langle \mathbf{u}, \mathbf{v}\rangle_{\mathcal{E}} = 0 ,\;\; \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{E} :\; \mathbf{u} \perp \mathbf{v} ,$$

which is the reason why $\mathbf{Q}$ is also called *orthogonal* (see, e.g. Mac Lane and Birkhoff (1999)).

Moreover, since any scalar product in a Euclidean vector space induces a norm by (4.1), an isometry $\mathbf{Q}$ also satisfies

$$\|\mathbf{Q}\mathbf{v}\|_{\mathcal{E}'} = \|\mathbf{v}\|_{\mathcal{E}} ,\;\; \forall\, \mathbf{v} \in \mathcal{E} . \tag{4.9}$$

A consequence of preserving the norm of vectors is the bijectivity of isometries, as shown in the following proposition.

Proposition 4.5. *Let* $\mathbf{Q}\colon \mathcal{E} \to \mathcal{E}'$ *be an isometry between the* n*-dimensional Euclidean vector spaces* $\mathcal{E}$ *and* $\mathcal{E}'$*. Then,* $\mathbf{Q}$ *is bijective.*

Proof. As a consequence of (4.9), the norm of the vector $\mathbf{Qv} \in \mathcal{E}'$ is null if, and only if, the norm of the vector $\mathbf{v} \in \mathcal{E}$ vanishes.

Then, since in any Euclidean vector space the scalar product is non-degenerate, as is the associated Euclidean norm, the condition $\mathbf{Qv} = \mathbf{o}'$ implies $\mathbf{v} = \mathbf{o}$, meaning that the linear isometry $\mathbf{Q}$ between $\mathcal{E}$ and $\mathcal{E}'$ is injective (cf. Proposition B.11).

Also, by virtue of Corollary B.14, since the Euclidean vector spaces $\mathcal{E}$ and $\mathcal{E}'$ have the same dimension, the orthogonal transformation $\mathbf{Q}$ is an isomorphism. □

A relevant feature of the orthogonal maps concerns the relation between the transpose and the inverse map.

Proposition 4.6. *Any linear transformation $\mathbf{Q} : \mathcal{E} \to \mathcal{E}'$ between the n-dimensional Euclidean vector spaces $\mathcal{E}$ and $\mathcal{E}'$ is orthogonal if, and only if, the transpose is coincident with the inverse map:*

$$\mathbf{Q}^{\mathsf{T}} = \mathbf{Q}^{-1} . \tag{4.10}$$

Proof. Let us apply the isomorphism $\mathbf{Q}$ to an arbitrary vector $\mathbf{u} \in \mathcal{E}$ and evaluate the transpose map to the image $\mathbf{Qu} \in \mathcal{E}'$. Recalling the identity (B.92), one has

$$\langle \mathbf{Q}^{\mathsf{T}}(\mathbf{Qu}), \mathbf{v} \rangle_{\mathcal{E}} = \langle \mathbf{Qu}, \mathbf{Qv} \rangle_{\mathcal{E}'} , \quad \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{E} .$$

Assuming $\mathbf{Q}$ is orthogonal, by (4.8) one derives

$$\langle \mathbf{Q}^{\mathsf{T}}(\mathbf{Qu}), \mathbf{v} \rangle_{\mathcal{E}} = \langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{E}} , \quad \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{E} ,$$

whence, for the arbitrariness of the vectors $\mathbf{u}$ and $\mathbf{v}$, one obtains $\mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathrm{id}_{\mathcal{E}}$, that is $\mathbf{Q}^{\mathsf{T}}$ is the inverse of $\mathbf{Q}$.

Conversely, supposing that (4.10) holds true, one can write

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{E}} = \langle \mathbf{Q}^{\mathsf{T}}(\mathbf{Qu}), \mathbf{v} \rangle_{\mathcal{E}} = \langle \mathbf{Qu}, \mathbf{Qv} \rangle_{\mathcal{E}'} , \quad \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{E} ,$$

which proves the orthogonality of $\mathbf{Q}$. □

A further property of orthogonal transformations between Euclidean vector spaces is to preserve the orthonormality of the relevant bases.

Proposition 4.7. *Let $\mathcal{E}$ and $\mathcal{E}'$ be Euclidean vector spaces with the same dimension n. Then, the linear map $\mathbf{Q}\colon \mathcal{E} \to \mathcal{E}'$ is orthogonal if, and only if, any orthonormal basis $\mathcal{B} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for $\mathcal{E}$ is mapped to an orthonormal basis $\mathcal{B}' = \{\mathbf{e}'_1, \dots, \mathbf{e}'_n\}$ of $\mathcal{E}'$.*

Proof. Since each basis vector $\mathbf{e}_i \in \mathcal{B}$ is mapped to $\mathbf{e}'_i \in \mathcal{B}'$ by $\mathbf{Q}$, the following identity holds true:

$$\langle \mathbf{e}'_i, \mathbf{e}'_j \rangle_{\mathcal{E}'} = \langle \mathbf{Q}\mathbf{e}_i, \mathbf{Q}\mathbf{e}_j \rangle_{\mathcal{E}'} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle_{\mathcal{E}} = \delta_{ij}\,.$$

Then, if $\mathbf{Q}$ is orthogonal, the scalar product is preserved, implying the orthonormality of $\mathcal{B}'$:

$$\langle \mathbf{e}'_i, \mathbf{e}'_j \rangle_{\mathcal{E}'} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle_{\mathcal{E}} = \delta_{ij}\,.$$

Conversely, if $\mathcal{B}'$ is orthonormal, the identity $\langle \mathbf{e}'_i, \mathbf{e}'_j \rangle_{\mathcal{E}'} = \delta_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle_{\mathcal{E}}$ implies the orthogonality of $\mathbf{Q}$. □

The Euclidean vector spaces $\mathcal{E}$ and $\mathcal{E}'$ such that an isometry $\mathbf{Q}\colon \mathcal{E} \to \mathcal{E}'$ does exist, are called *isometric*.

Recalling the identification (4.5), it is clear that an n-dimensional Euclidean vector space is isometric to $\mathbb{R}^n$, which is assumed as the universal model for all the n-dimensional Euclidean vector spaces.

4.1.3 The Orthogonal Group

By virtue of Proposition 4.5, an isometry $\mathbf{Q} : \mathcal{E} \to \mathcal{E}$, defined from a Euclidean vector space $\mathcal{E}$ to itself, is an automorphism of $\mathcal{E}$.

Then, the set of all the isometries of $\mathcal{E}$ is a subset of $\mathrm{Aut}(\mathcal{E})$ and, along with the usual linear map composition, has a group structure. Such a group, which is actually a subgroup of the general linear group of $\mathcal{E}$ (cf. Definition B.57), can be characterized using the property (4.10).

Definition 4.8. Let $\mathcal{E}$ be a Euclidean vector space. The *orthogonal group* of $\mathcal{E}$, denoted as $O(\mathcal{E})$, is the group of the automorphisms of $\mathcal{E}$ which are isometries:

$$O(\mathcal{E}) = \{\, \mathbf{Q} \in GL(\mathcal{E}) \mid \mathbf{Q}\mathbf{Q}^{\mathsf{T}} = \mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathrm{id}_{\mathcal{E}} \,\}\,. \tag{4.11}$$

In addition to the properties shown in Section 4.1.2, since $\mathbf{Q} \in O(\mathcal{E})$ is an endomorphism of $\mathcal{E}$, a further characterization concerns its determinant.

Actually, applying the properties of determinants given by (B.79) and (B.80), and using (4.10), one has

$$\det(\mathrm{id}_{\mathcal{E}}) = \det(\mathbf{Q}^{\mathsf{T}}\mathbf{Q}) = \det(\mathbf{Q}^{\mathsf{T}})\det(\mathbf{Q}) = 1\,,$$

and, exploiting the result of Proposition B.46, one also infers

$$\det(\mathbf{Q}^{\mathsf{T}})\det(\mathbf{Q}) = \det(\mathbf{Q})^2 = 1\,,$$

whence $\det(\mathbf{Q}) = \pm 1$.

Please notice that, when a basis $\mathcal{B}$ for $\mathcal{E}$ is considered, any isometry $\mathbf{Q} \in O(\mathcal{E})$ can be seen as a change of basis map from $\mathcal{B}$ to another basis $\mathcal{B}'$ of the same vector space $\mathcal{E}$. Also, recalling Proposition 4.7, if $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal, the basis $\mathcal{B}' = \{\mathbf{e}'_1, \ldots, \mathbf{e}'_n\}$ of $\mathcal{E}$ is itself orthonormal.

Moreover, suppose an orientation $\mathcal{O}$ is considered for $\mathcal{E}$ (cf. Definition B.51). Recalling Proposition B.52, the basis $\mathcal{B}'$ resulting from applying $\mathbf{Q}$ is consistently oriented with $\mathcal{B}$ if, and only if, the determinant of $\mathbf{Q}$ is positive.

Explicitly, the elements $\mathbf{R} \in O(\mathcal{E})$ such that $\det(\mathbf{R}) = 1$ are the isometries of $\mathcal{E}$ preserving the orientation induced by a basis $\mathcal{B}$, a condition leading to the following definition.

Definition 4.9. Given a Euclidean vector space $\mathcal{E}$, the *special orthogonal group* of $\mathcal{E}$, denoted as $SO(\mathcal{E})$, is the group of isometries of $\mathcal{E}$ preserving orientations:

$$SO(\mathcal{E}) = \{\, \mathbf{R} \in O(\mathcal{E}) \mid \det(\mathbf{R}) = 1 \,\}\,. \tag{4.12}$$

and an isometry $\mathbf{R} \in SO(\mathcal{E})$ is called a *rotation* of $\mathcal{E}$.

In order to verify that the definition introduced here above actually makes sense, please observe that $SO(\mathcal{E})$ is closed under linear map composition, resulting

$$\det(\mathbf{R}_1\mathbf{R}_2) = \det(\mathbf{R}_1)\det(\mathbf{R}_2) = 1\,, \quad \forall\,\mathbf{R}_1, \mathbf{R}_2 \in SO(\mathcal{E})\,.$$

so that the composition of linear maps provides the group structure to $SO(\mathcal{E})$.

Please observe that the isometries $\mathbf{Q}$ of $\mathcal{E}$ with $\det(\mathbf{Q}) = -1$, named the *reflections* of $\mathcal{E}$, do not form a group. This trivially comes noting that $\det(\mathbf{id}_{\mathcal{E}}) = 1$, so that the identity $\mathbf{id}_{\mathcal{E}}$ is not a reflection and $O(\mathcal{E})\backslash SO(\mathcal{E})$ is not a group.

More generally, $O(\mathcal{E})\backslash SO(\mathcal{E})$ is not closed under the usual linear map composition:

$$\det(\mathbf{Q}_1\mathbf{Q}_2) = \det(\mathbf{Q}_1)\det(\mathbf{Q}_2) = 1 \neq -1\,, \quad \forall\,\mathbf{Q}_1, \mathbf{Q}_2 \in O(\mathcal{E})\backslash SO(\mathcal{E})\,.$$

Matrix Orthogonal Group

It is worth recalling that, by virtue of the identification (B.108), the features of isometries apply in a similar way to their matrix representation.

Specifically, a matrix $\mathbf{Q} \in M_n$ is *orthogonal* if satisfies $\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I}$, which also reads

$$Q^{\mathsf{T}\,i}{}_j Q^j{}_k = Q^j{}_i Q^j{}_k = \langle \mathbf{Q}_{(i)}, \mathbf{Q}_{(k)} \rangle_{M_{n\times 1}} = \delta_{ik}\,,$$

meaning that the columns of $\mathbf{Q}$ are orthonormal.

Then, the *orthogonal group* of degree n is the subgroup of $GL(n)$ whose elements are orthogonal matrices:

$$O(n) = \{\, \mathbf{Q} \in GL(n) \mid \mathbf{Q}\mathbf{Q}^\mathsf{T} = \mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I} \,\}\,, \tag{4.13}$$

as well as the *special orthogonal group* of degree n is the subgroup of $O(n)$ defined as

$$SO(n) = \{\, \mathbf{R} \in O(n) \mid \det(\mathbf{R}) = 1 \,\}\,, \tag{4.14}$$

where, similarly to (B.117), the following isomorphism holds true:

$$SO(\mathcal{E}) \cong SO(n)\,. \tag{4.15}$$

4.2 Euclidean Affine Spaces

Definition 4.10. A *Euclidean affine space*, or simply a *Euclidean space*, is an affine space $\mathscr{E}$ such that the underlying vector space $\mathcal{E}$ is Euclidean.

Recalling that the vector space $\mathcal{E}$ is endowed with the Euclidean norm given by (4.1), one can exploit such a norm to define the distance between points of the space $\mathscr{E}$ (see, e.g., Tarrida (2011)).

Definition 4.11. Let $\mathscr{E}$ be a Euclidean space and $\mathcal{E}$ the associated vector space. The *Euclidean distance* between the points P and Q of $\mathscr{E}$ is the map

$$\begin{aligned} d:\ & \mathscr{E} \times \mathscr{E} \to \mathbb{R}_0^+ \\ & (\mathsf{P},\mathsf{Q}) \mapsto \|\overrightarrow{\mathsf{PQ}}\|\,, \end{aligned} \tag{4.16}$$

where $\|\overrightarrow{\mathsf{PQ}}\|$ is the Euclidean norm of the vector $\overrightarrow{\mathsf{PQ}}$.

4.2.1 Orthonormal Frames

Please recall from Definition 3.4 that an affine frame $\mathscr{F} = \{\mathsf{O};\mathcal{B}\}$ for $\mathscr{E}$ consists a point $\mathsf{O} \in \mathscr{E}$ and a basis $\mathcal{B}$ for the associated vector space $\mathcal{E}$, so that the features of the basis $\mathcal{B}$ are also reflected in the frame $\mathscr{F}$.

Definition 4.12. An affine frame $\mathscr{F} = \{\mathsf{O};\mathcal{B}\}$ for a Euclidean affine space $\mathscr{E}$ is *orthonormal* if $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal basis for the associated vector space $\mathcal{E}$.

A further feature of the underlying vector space $\mathcal{E}$ which is reflected on the affine space $\mathscr{E}$ concerns the orientation.

Specifically, the affine Euclidean space $\mathscr{E}$ is *oriented* if an orientation $\mathcal{O}$ is fixed for the associated vector space $\mathcal{E}$.

Then, the orientation of the affine frame $\mathscr{F} = \{\mathsf{O};\mathcal{B}\}$ is the one induced by basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of $\mathcal{E}$.

4.2.2 Euclidean Motion

Definition 4.13. Let $\mathscr{E}$ and $\mathscr{E}'$ be Euclidean affine spaces. An *isometry* between $\mathscr{E}$ and $\mathscr{E}'$ is a map $f\colon \mathscr{E} \to \mathscr{E}'$ preserving the distance:

$$d\big(f(\mathsf{P}), f(\mathsf{Q})\big) = d(\mathsf{P}, \mathsf{Q})\,, \ \ \forall\, \mathsf{P}, \mathsf{Q} \in \mathscr{E}\,. \tag{4.17}$$

The isometries between Euclidean affine spaces are strictly related to the linear isometries between the relevant vector spaces, as shown in the following proposition.

Proposition 4.14. *Let $\mathscr{E}$ and $\mathscr{E}'$ be Euclidean affine spaces and let $\mathcal{E}$ and $\mathcal{E}'$ be the associated vector spaces. Then, a map $f\colon \mathscr{E} \to \mathscr{E}'$ is an isometry if, and only if, it is an affinity and the associated linear map $\mathbf{Q}\colon \mathcal{E} \to \mathcal{E}'$ is orthogonal.*

Proof. Suppose f is an isometry, so that the following relation is satisfied:

$$d\big(f(\mathsf{P}), f(\mathsf{Q})\big) = d(\mathsf{P}, \mathsf{Q}) = \|\overrightarrow{\mathsf{PQ}}\|\,.$$

At the same time, the distance function on $\mathscr{E}$ is such that

$$d\big(f(\mathsf{P}), f(\mathsf{Q})\big) = \|\overrightarrow{f(\mathsf{P})f(\mathsf{Q})}\| = \|\mathbf{Q}_{\mathsf{P}}(\overrightarrow{\mathsf{PQ}})\|\,,$$

where $\mathbf{Q}_{\mathsf{P}}\colon \mathcal{E} \to \mathcal{E}'$ is the map induced by f, as defined by (3.7).

By comparison, it is clear that the map $\mathbf{Q}_{\mathsf{P}}$ is distance-preserving and, by Proposition 4.4, it is a linear transformation.

Consequently, recalling Definition 3.5, the isometry f is an affinity and the associated linear map $\mathbf{Q} = \mathbf{Q}_{\mathsf{P}}$, which actually does not depend on P (cf. Proposition 3.6), is orthogonal.

The converse is readily verified. In fact, if f is an affinity and $\mathbf{Q}$ is orthogonal, the following relation holds true:

$$d\big(f(\mathsf{P}), f(\mathsf{Q})\big) = \|\overrightarrow{f(\mathsf{P})f(\mathsf{Q})}\| = \|\mathbf{Q}(\overrightarrow{\mathsf{PQ}})\| = \|\overrightarrow{\mathsf{PQ}}\| = d(\mathsf{P}, \mathsf{Q})\,,$$

which means that the affinity f is actually an isometry. □

4.2.3 The Euclidean Group

When an isometry is in the form $f : \mathscr{E} \to \mathscr{E}$, the associated linear map is an orthogonal transformation $\mathbf{Q} : \mathcal{E} \to \mathcal{E}$. Consequently, since $\mathbf{Q}$ is an automorphism of $\mathcal{E}$, the affinity f is also bijective (see, e.g., Tarrida (2011)).

Then, the map composition provides a group structure to the set of the isometries of $\mathscr{E}$, which is a subgroup of the affine group $GA(\mathscr{E})$ (cf. Definition 3.10), and the group of the induced linear maps is the orthogonal group of $\mathcal{E}$.

Definition 4.15. Let $\mathscr{E}$ be a Euclidean affine space with the associated vector space $\mathcal{E}$. The *Euclidean group*, or the *group Euclidean transformations*, of $\mathscr{E}$ is the group of the isometries of $\mathscr{E}$ and is denoted as $E(\mathscr{E})$:

$$E(\mathscr{E}) = \{\, f \in GA(\mathscr{E}) \mid \mathbf{Q} \in O(\mathcal{E}) \,\}. \tag{4.18}$$

Moreover, when a rotation $\mathbf{R} \in SO(\mathcal{E})$ is considered as associated linear map, the isometry f is called a *(proper) Euclidean motion*, or also a *(proper) rigid motion*. This kind of isometries form a subgroup of the Euclidean group of $\mathscr{E}$.

Definition 4.16. The *special Euclidean group* of a Euclidean affine space $\mathscr{E}$ is the group of the proper Euclidean motions of $\mathscr{E}$:

$$SE(\mathscr{E}) = \{\, \mathcal{H} \in E(\mathscr{E}) \mid \mathbf{R} \in SO(\mathcal{E}) \,\}. \tag{4.19}$$

Please recall from Definition 3.1 that, as an affine space, the Euclidean space $\mathscr{E}$ is characterized by the action of the associated vector space $\mathcal{E}$ on its points, which takes the form of translations (cf. Definition 3.3).

Specifically, any translation map $t_{\mathbf{v}}$ is an isometry for $\mathscr{E}$, resulting

$$d\big(t_{\mathbf{v}}(\mathsf{P}), t_{\mathbf{v}}(\mathsf{Q})\big) = d(\mathsf{P} + \mathbf{v}, \mathsf{Q} + \mathbf{v}) = d(\mathsf{P}, \mathsf{Q})\,, \quad \forall\, \mathsf{P}, \mathsf{Q} \in \mathscr{E}\,,$$

since the vector from $\mathsf{P} + \mathbf{v}$ to $\mathsf{Q} + \mathbf{v}$ is the same vector $\overrightarrow{\mathsf{PQ}}$ from P to Q.

In addition, as shown in Proposition 3.11, the translations of $\mathscr{E}$ are the affinities whose relevant linear map is the identity $\mathbf{id}_{\mathcal{E}}$, and the then group of translations $T(\mathscr{E})$ is normal in $SE(\mathscr{E})$.

Consequently, similarly to the decomposition applied to the affine group $GA(\mathscr{E})$ in Section 3.3.2, the Euclidean group of $\mathscr{E}$ can be expressed as

$$SE(\mathscr{E}) \cong T(\mathscr{E}) \rtimes SE_{\mathsf{P}}(\mathscr{E})\,, \tag{4.20}$$

where SE_{P} is the group of the proper motions of $\mathscr{E}$ which fix a point P and acts by conjugation on the translations of $\mathscr{E}$.

In practice, any proper motion $\mathcal{H} \in SE(\mathscr{E})$ can be thought as a proper rigid motion $\mathcal{H}_{\mathsf{P}} \in SE_{\mathsf{P}}(\mathscr{E})$ around a point P, followed by the translation $t_{\mathbf{v}}$ by a vector $\mathbf{v}$:

$$\mathcal{H} = t_{\mathbf{v}}\mathcal{H}_{\mathsf{P}}\,. \tag{4.21}$$

In addition, following the same approach shown in Proposition 3.15, one can verify that the stabilizer $SE_{\mathsf{P}}(\mathscr{E})$ is isomorphic with the special orthogonal group $SO(\mathcal{E})$.

Then, considering also that any translation $t_{\mathbf{v}}$ is associated with a vector $\mathbf{v} \in \mathcal{E}$, and vice-versa, the following representation of $SE(\mathscr{E})$ applies:

$$SE(\mathscr{E}) \cong \mathcal{E} \rtimes SO(\mathcal{E})\,, \tag{4.22}$$

and the proper Euclidean motion $\mathcal{H}$ can be represented as

$$\mathcal{H} \cong (\mathbf{v}, \mathbf{R})\,. \tag{4.23}$$

Furthermore, the composition rule consistent with the semidirect product (4.22) is

$$\mathcal{H}_2\mathcal{H}_1 \cong (\mathbf{v}_2, \mathbf{R}_2)(\mathbf{v}_1, \mathbf{R}_1) = (\mathbf{v}_2 + \mathbf{R}_2\mathbf{v}_1, \mathbf{R}_2\mathbf{R}_1)\,, \tag{4.24}$$

where the vector $\mathbf{R}_2\mathbf{v}_1$ results from the action by conjugation of $\mathcal{H}_{\mathsf{P}_2} \cong \mathbf{R}_2$ on $t_{\mathbf{v}_1} \cong \mathbf{v}_1$ (cf. Proposition 3.14).

Matrix Euclidean Group

Please recall that when a frame $\mathscr{F}$ is considered, an affinity f can be expressed in matrix form by the homogeneous representation (cf. Definition

3.17).

Referring to an isometry f of the Euclidean group $E(\mathscr{E})$, the homogeneous representation reads

$$f \cong \mathbf{L} = \begin{bmatrix} \mathbf{Q} & \mathbf{v} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix}, \tag{4.25}$$

with the condition $\mathbf{Q} \in O(n)$.

Then, the matrices expressed as in (4.25) form a group which is called the *Euclidean group* of degree n:

$$E(n) = \left\{ \begin{bmatrix} \mathbf{Q} & \mathbf{v} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} \in GL(n+1) \;\middle|\; \mathbf{Q} \in O(n) \right\}. \tag{4.26}$$

Similarly, the homogeneous representation of a proper Euclidean motion $\mathcal{H} \in SE(\mathscr{E})$ results

$$\mathcal{H} \cong \mathbf{H} = \begin{bmatrix} \mathbf{R} & \mathbf{v} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix}, \tag{4.27}$$

where now $\mathbf{R}$ is in $SO(n)$.

Consequently, the group of the matrices in the form (4.27) is called the *special Euclidean group* of degree n:

$$SE(n) = \left\{ \begin{bmatrix} \mathbf{R} & \mathbf{v} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} \in E(n) \;\middle|\; \mathbf{R} \in SO(n) \right\}. \tag{4.28}$$

It is easy to verify that the matrix multiplication of two matrices $\mathbf{H}_2$ and $\mathbf{H}_1$ in the form (4.27) is consistent with the composition of the motions $(\mathbf{v}_2, \mathbf{R}_2)$ and $(\mathbf{v}_1, \mathbf{R}_1)$ expressed by (4.24):

$$\mathbf{H}_2\mathbf{H}_1 = \begin{bmatrix} \mathbf{R}_2 & \mathbf{v}_2 \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{v}_1 \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_2\mathbf{R}_1 & \mathbf{v}_2 + \mathbf{R}_2\mathbf{v}_1 \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix}, \tag{4.29}$$

so that the groups $SE(\mathscr{E})$ and $SE(n)$ are isomorphic:

$$SE(\mathscr{E}) \cong SE(n)\,. \tag{4.30}$$

Moreover, the matrix product (4.29) also highlights the action of $\mathbf{R}_2 \in SO(n)$ on $\mathbf{v}_1 \in \mathbb{R}^n$ in accordance with the following semidirect product decomposition:

$$SE(n) \cong \mathbb{R}^n \rtimes SO(n)\,. \tag{4.31}$$

4.3 Rotations in 3-Dimensional Spaces

The structure of Euclidean vector spaces and Euclidean affine spaces introduced in the previous section is deeper investigated for the case of three-dimensional spaces.

Specifically, for the role that this kind of spaces plays in Mechanics, it appears appropriate to provide a more comprehensive description of rigid motions in three-dimensional spaces.

4.3.1 Cross Product and Alternating Tensors

Let $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be an orthonormal basis for the Euclidean vector space $\mathcal{E}$ and consider the space of the alternating tensors $\Lambda^3(\mathcal{E})$.

Please recall from Appendix B.2.3 that $\Lambda^3(\mathcal{E})$ is a one-dimensional vector space whose basis is given by the wedge product of the vectors of $\mathcal{B}$:

$$\iota = \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3\,. \tag{4.32}$$

Since $\Lambda^3(\mathcal{E})$ is a real one-dimensional vector space, it is isomorphic with $\mathbb{R}$ and, for this reason, any element $\omega \in \Lambda^3(\mathcal{E})$ is also called a *pseudoscalar*.

Please observe that the vectors $\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k$ do satisfy

$$\mathbf{e}_i \wedge \mathbf{e}_j \wedge \mathbf{e}_k = \epsilon_{ijk}\iota\,, \tag{4.33}$$

where ϵ_{ijk} is the Levi-Civita symbol:

$$\epsilon_{ijk} = \begin{cases} \mathrm{sgn}(\sigma) & \text{if } (i,j,k) = \big(\sigma(1), \sigma(2), \sigma(3)\big)\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{4.34}$$

with $\sigma \in S_3$ a permutation on $\{1, 2, 3\}$.

Also, the space $\Lambda^2(\mathcal{E})$ of alternating tensors is a three-dimensional vector space over $\mathbb{R}$ and is isomorphic with $\mathcal{E}$. Actually any tensor $\mathbf{a} \wedge \mathbf{b} \in \Lambda^2(\mathcal{E})$ can be associated with a unique vector $\mathbf{c} \in \mathcal{E}$ by means of the following construction:

$$\langle \mathbf{c}, \mathbf{v} \rangle \iota = \mathbf{v} \wedge \mathbf{a} \wedge \mathbf{b}\,, \ \ \forall\, \mathbf{v} \in \mathcal{E}\,, \tag{4.35}$$

and the map identifying $\mathbf{a} \wedge \mathbf{b}$ with $\mathbf{c} = *(\mathbf{a} \wedge \mathbf{b})$ is also called the *Hodge star* (see, e.g., Winitzki (2020)).

As an example, the Hodge star of $\mathbf{e}_2 \wedge \mathbf{e}_3$ is evaluated as

$$\langle \mathbf{c}, \mathbf{v} \rangle \iota = \mathbf{v} \wedge \mathbf{e}_3 \wedge \mathbf{e}_3 = v^1 \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3 = v^1 \iota\,, \ \ \forall\, \mathbf{v} \in \mathcal{E}\,,$$

where, for the arbitrariness $\mathbf{v}$, one infers $\mathbf{c} = *(\mathbf{e}_2 \wedge \mathbf{e}_3) = \mathbf{e}_1$, with the usual identification $\mathbf{e}^1(\mathbf{v}) = \langle \mathbf{e}_1, \mathbf{v} \rangle$. Similarly, one can evaluate $*(\mathbf{e}_3 \wedge \mathbf{e}_1) = \mathbf{e}_2$ and $*(\mathbf{e}_1 \wedge \mathbf{e}_2) = \mathbf{e}_3$.

More generally, noting that $\mathbf{v} \wedge \mathbf{a} \wedge \mathbf{b} = \epsilon_{ijk} v^i a^j b^k \iota$, and that $\langle \mathbf{a} \times \mathbf{b}, \mathbf{v} \rangle = \epsilon_{ijk} v^i a^j b^k$ one can easily recognize that the Hodge star of $\mathbf{a} \wedge \mathbf{b}$ is exactly the cross product of $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbf{a} \times \mathbf{b} = *(\mathbf{a} \wedge \mathbf{b})\,. \tag{4.36}$$

Moreover, since $\Lambda^2(\mathcal{E})$ is a subspace of $\mathcal{T}^{(2,0)}(\mathcal{E})$, for the self-duality of $\mathcal{E}$ and with the identification (4.7), any $\mathbf{W} \in \Lambda^2(\mathcal{E})$ results a linear map $\mathbf{W} \colon \mathcal{E} \to \mathcal{E}$ such that the property of being alternating applies as follows:

$$\langle \mathbf{W}\mathbf{v}, \mathbf{w} \rangle = -\langle \mathbf{v}, \mathbf{W}\mathbf{w} \rangle\,, \ \ \forall\, \mathbf{v}, \mathbf{w}, \in \mathcal{E}\,, \tag{4.37}$$

which, being $\langle \mathbf{v}, \mathbf{W}\mathbf{w} \rangle = \langle \mathbf{W}^{\mathsf{T}}\mathbf{v}, \mathbf{w} \rangle$, specializes in

$$\mathbf{W}^{\mathsf{T}} = -\mathbf{W}\,. \tag{4.38}$$

A further connection between the cross product and the space of alternating tensors is given by the following map:

$$\begin{aligned} (\cdot)^{\frown} : \mathcal{E} &\to \Lambda^2(\mathcal{E}) \\ \mathbf{a} &\mapsto \widehat{\mathbf{a}}\,, \end{aligned} \tag{4.39}$$

such that $\widehat{\mathbf{a}} \in \Lambda^2(\mathcal{E})$ is the alternating tensor satisfying

$$\widehat{\mathbf{a}}\mathbf{b} = \mathbf{a} \times \mathbf{b}\,, \quad \forall\, \mathbf{b} \in \mathcal{E}\,. \tag{4.40}$$

Moreover, since the cross product is linear, the map (4.39) applies as follows

$$\widehat{\mathbf{a}} = (a^i \mathbf{e}_i)^{\frown} = a^i \widehat{\mathbf{e}}_i\,,$$

which takes the explicit matrix form:

$$\mathbf{a} = \begin{bmatrix} a^1 \\ a^2 \\ a^3 \end{bmatrix} \mapsto \widehat{\mathbf{a}} = \begin{bmatrix} 0 & -a^3 & a^2 \\ a^3 & 0 & -a^1 \\ -a^2 & a^1 & 0 \end{bmatrix}. \tag{4.41}$$

The skew-symmetric matrices provide the matrix representation of the alternating tensors in $\Lambda^2(\mathcal{E})$ and can themselves be considered as alternating tensors for the vector space $\mathbb{R}^3$. Then, the following characterization is introduced:

$$\Lambda^2(\mathbb{R}^3) = \{\, \mathbf{W} \in M_3 \mid \mathbf{W}^\mathsf{T} = -\mathbf{W} \,\}\,. \tag{4.42}$$

4.3.2 3-Dimensional Rotations

The transformations of the special orthogonal group, introduced by Definition 4.8, do have the geometric meaning of rotations, intended as the circular movement around an axis.

Specifically, in the framework of a three-dimensional Euclidean vector space $\mathcal{E}$, let $\mathbf{k}$ be a unit vector identifying an axis of rotation and let $\mathbf{u}$ be a vector rotated by an angle θ around $\mathbf{k}$, according to the right hand rule.

The resulting vector $\mathbf{v}$ is expressed as

$$\mathbf{v} = \cos\theta\,\mathbf{u} + \sin\theta\,\mathbf{k}\times\mathbf{u} + (1-\cos\theta)(\mathbf{k}\cdot\mathbf{u})\mathbf{k}\,,$$

which represents the *Rodrigues' rotation formula* (Rodrigues, 1840), and can be rearranged as

$$\mathbf{v} = \mathbf{u} + \sin\theta\,\mathbf{k}\times\mathbf{u} + (1-\cos\theta)\mathbf{k}\times\mathbf{k}\times\mathbf{u}\,. \tag{4.43}$$

Moreover, applying the map (4.39) to the unit vector $\mathbf{k}$, such a transformation can be expressed as

$$\mathbf{v} = \mathbf{u} + \sin\theta\,\widehat{\mathbf{k}}\mathbf{u} + (1-\cos\theta)\widehat{\mathbf{k}}^2\mathbf{u}\,, \tag{4.44}$$

so that the rotation by θ around $\mathbf{k}$ takes the form of the linear transformation $\mathbf{R}_{(\theta,\mathbf{k})}\colon \mathcal{E}\to\mathcal{E}$ expressed as

$$\mathbf{R}_{(\theta,\mathbf{k})} = \mathbf{I} + \sin\theta\,\widehat{\mathbf{k}} + (1-\cos\theta)\widehat{\mathbf{k}}^2\,. \tag{4.45}$$

It is possible to verify by direct computation that $\mathbf{R}_{(\theta,\mathbf{k})}\mathbf{R}^{\mathsf{T}}_{(\theta,\mathbf{k})} = \mathbf{I}$, as well as $\det(\mathbf{R}_{(\theta,\mathbf{k})}) = 1$, the geometric transformations in the form (4.45) are exactly the proper orthogonal transformation of $\mathcal{E}$.

When the transformation $\mathbf{R}_{(\theta,\mathbf{k})}$ is applied to the vectors of an orthonormal basis $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, a new basis $\mathcal{B}' = \{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ is obtained.

Please observe that, by Proposition 4.7, the new basis is itself orthonormal and, being $\det(\mathbf{R}_{(\theta,\mathbf{k})}) = 1 > 0$, the orientation induced by $\mathcal{B}$ is also preserved.

Then, $\mathbf{R}_{(\theta,\mathbf{k})} = \mathbf{R}^{\mathcal{B}'}_{\mathcal{B}}$ represents the change of basis map from $\mathcal{B}$ to $\mathcal{B}'$ and, as such, the associated matrix has both the meaning of the active transformation from $\mathcal{B}$ to $\mathcal{B}'$ and the passive coordinate transformation from $\mathcal{B}'$ to $\mathcal{B}$ (cf. Appendix B.4.4).

In other words, the coordinates of a vector $\mathbf{u}$ attached to $\mathcal{B}$ transform as $\mathbf{u}' = \mathbf{R}^{\mathcal{B}'}_{\mathcal{B}}\mathbf{u}$ when $\mathcal{B}$ is rotated to the configuration $\mathcal{B}'$. At the same time, if $\mathbf{u}_{\mathcal{B}'}$ is the coordinate vector of $\mathbf{u}$ with respect to $\mathcal{B}'$, the same transformation $\mathbf{u}_{\mathcal{B}} = \mathbf{R}^{\mathcal{B}'}_{\mathcal{B}}\mathbf{u}_{\mathcal{B}'}$ provides the coordinate vector $\mathbf{u}_{\mathcal{B}}$ with respect to the basis $\mathcal{B}$.

In addition, the same rotation $\mathbf{R}^{\mathcal{B}'}_{\mathcal{B}}$ also plays to role to represent the configuration of $\mathcal{B}'$ with respect to $\mathcal{B}$. Actually, supposing the basis $\mathcal{B}$ is fixed and the new one $\mathcal{B}'$ is introduced, the relevant basis vectors do transform as

$$\mathbf{e}'_j = \mathbf{R}_{\mathcal{B}}^{\mathcal{B}'} \mathbf{e}_j \,,$$

whence, the i-th coordinate of the unit vector $\mathbf{e}'_j$ with respect to $\mathcal{B}$ is

$$(\mathbf{e}'_j)^i = \mathbf{e}'_j \cdot \mathbf{e}_i = (\mathbf{R}_{\mathcal{B}}^{\mathcal{B}'} \mathbf{e}_j) \cdot \mathbf{e}_i = (\mathbf{R}_{\mathcal{B}}^{\mathcal{B}'})_{ij} \,,$$

that is the j-th column of the orthogonal matrix $\mathbf{R}_{\mathcal{B}}^{\mathcal{B}'}$ gathers the coordinates of the new unit vector $\mathbf{e}'_j$ with respect to the basis $\mathcal{B}$.

4.3.3 Lie Group Structure of $SO(3)$

By virtue of the isomorphism (4.15), the matrix group $SO(3)$ is here analyzed as a reference for the group of proper rotations of $\mathcal{E}$.

Please recall that the group $SO(3)$ is the matrix Lie group characterized by the property $\mathbf{R}\mathbf{R}^\mathsf{T} = \mathbf{R}^\mathsf{T}\mathbf{R} = \mathbf{I}$, with $\det(\mathbf{R}) = 1$.

The defining property can be exploited to identify the Lie algebra of $SO(3)$ as follows.

Let us consider a parameterization $t \mapsto \mathbf{R}(t)$ such that $\mathbf{R}(0) = \mathbf{I}$, which represents a curve on $SO(3)$ starting at the identity $\mathbf{I}$. Moreover, for the orthogonality of any matrix of $SO(3)$, the following identity holds true

$$\mathbf{R}(t)\big(\mathbf{R}(t)\big)^\mathsf{T} = \mathbf{I} \,,$$

and the derivative with respect to the scalar t reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{R}\mathbf{R}^\mathsf{T} = \mathbf{R}'(t)\mathbf{R}^\mathsf{T}(t) + \mathbf{R}(t)\mathbf{R}'^\mathsf{T}(t) = \mathbf{0} \,,$$

where $\mathbf{R}'(t)$ denotes the derivative of $\mathbf{R}$ with respect to t, and the same applies to the transpose, while the derivative of the identity matrix $\mathbf{I}$ does vanish.

Then, setting $t = 0$, one finds

$$\mathbf{R}'(0)\mathbf{R}^{\mathsf{T}}(0) + \mathbf{R}(0)\mathbf{R}'^{\mathsf{T}}(0) = \mathbf{R}'(0) + \mathbf{R}'^{\mathsf{T}}(0) = \mathbf{0}\,,$$

whence the velocity $\mathbf{R}'(t)$ of any curve in the form $t \mapsto \mathbf{R}(t)$ at the identity turns out to be a skew-symmetric matrix

$$\mathbf{R}'(0) = -\mathbf{R}'^{\mathsf{T}}(0)\,.$$

Since the velocity of a curve at a point of a smooth manifold is an element of the tangent space at that point (cf. Section 2.2.5), the tangent space of $SO(3)$ at the identity $\mathbf{I}$ is the space three-dimensional skew-symmetric matrices:

$$T_{\mathbf{I}}SO(3) = \Lambda^2(\mathbb{R}^3)\,.$$

With the identification $\mathsf{Lie}\,\big(SO(3)\big) \cong T_{\mathbf{I}}SO(3)$ (cf. Proposition 2.33), one can conclude that the Lie algebra of $SO(3)$ is the space of the skew-symmetric matrices, endowed with the matrix commutator as Lie bracket.

Such a space is a subalgebra of the general linear Lie algebra $\mathfrak{gl}(3)$ and is denoted as $\mathfrak{so}(3)$.

Definition 4.17. The *special orthogonal Lie algebra* of order 3, denoted as $\mathfrak{so}(3)$, is the space of the three-dimensional skew-symmetric matrices, endowed with the matrix commutator:

$$\mathfrak{so}(3) = \{\, \mathbf{W} \in \mathfrak{gl}(3) \mid \mathbf{W}^{\mathsf{T}} = -\mathbf{W} \,\}\,. \tag{4.46}$$

Because of the connection of $\mathfrak{so}(3)$ with the Lie group $SO(3)$, the following identification applies:

$$\mathsf{Lie}\,\big(SO(3)\big) \cong \mathfrak{so}(3)\,. \tag{4.47}$$

4.3.4 Exponential Map of $SO(3)$

By virtue of the vector space isomorphism $\mathbb{R}^3 \to \Lambda^2(\mathbb{R}^3)$ specified by (4.41), let us consider the elements of the Lie algebra $\mathfrak{so}(3)$ in the form $\widehat{\boldsymbol{\theta}}$, with $\boldsymbol{\theta}$

a vector of $\mathbb{R}^3$.

Moreover, it is useful to consider the decomposition

$$\boldsymbol{\theta} = \theta \mathbf{k}\,, \tag{4.48}$$

with $\theta = \|\boldsymbol{\theta}\|$ and $\mathbf{k} = \boldsymbol{\theta}/\theta$.

With the decomposition (4.48), the vector $\boldsymbol{\theta} \in \mathbb{R}^3$ can represent a three-dimensional rotation by the angle θ around the axis identified by $\mathbf{k}$, and the relevant rotation matrix is given by the Rodrigues' formula (4.45).

Then, the exponential map for $SO(3)$ reads

$$\begin{aligned} \exp :\ \mathfrak{so}(3) &\to SO(3) \\ \widehat{\boldsymbol{\theta}} &\mapsto \exp(\widehat{\boldsymbol{\theta}}) = \mathbf{R}_{\boldsymbol{\theta}}\,, \end{aligned} \tag{4.49}$$

where $\mathbf{R}_{\boldsymbol{\theta}}$ is expressed as

$$\mathbf{R}_{\boldsymbol{\theta}} = \mathbf{I} + \sin\|\boldsymbol{\theta}\| \, \frac{\widehat{\boldsymbol{\theta}}}{\|\boldsymbol{\theta}\|} + \big(1 - \cos\|\boldsymbol{\theta}\|\big) \frac{\widehat{\boldsymbol{\theta}}^2}{\|\boldsymbol{\theta}\|^2}\,. \tag{4.50}$$

The explicit expression of $\exp(\widehat{\boldsymbol{\theta}})$ can also be derived recalling that the exponential map of a matrix Lie group is given by the matrix exponential (cf. Section 2.6.2).

Then, the power series defined by (2.42) specializes to

$$e^{\widehat{\boldsymbol{\theta}}} = \sum_{k=0}^{\infty} \frac{\widehat{\boldsymbol{\theta}}^k}{k!} = \mathbf{I} + \sum_{j=0}^{\infty} \frac{\theta^{2j+1}}{(2j+1)!} \widehat{\mathbf{k}}^{2j+1} + \sum_{j=1}^{\infty} \frac{\theta^{2j}}{(2j)!} \widehat{\mathbf{k}}^{2j}\,, \tag{4.51}$$

where the odds powers have been separated from the even ones and the the decomposition $\widehat{\boldsymbol{\theta}} = \theta \widehat{\mathbf{k}}$ has been used.

Moreover, please observe that for a unit vector $\widehat{\mathbf{k}}$ the following identities apply:

$$\widehat{\mathbf{k}}^{2j} = (-1)^j \widehat{\mathbf{k}}^2\,, \qquad \widehat{\mathbf{k}}^{2j+1} = (-1)^j \widehat{\mathbf{k}}\,,$$

so that one finds

$$e^{\widehat{\theta}} = \mathbf{I} + \left(\sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} \theta^{2j+1} \right) \widehat{\mathbf{k}} + \left(\sum_{j=1}^{\infty} \frac{(-1)^j}{(2j)!} \theta^{2j} \right) \widehat{\mathbf{k}}^2 .$$

It is easy to recognize that the two power series in the formula here above represent exactly $\sin\theta$ and $1 - \cos\theta$, respectively, in the form of the Maclaurin series expansion. Hence, the same expression as in (4.50) is found.

Differential of the Exponential Map of $SO(3)$

In addition to the exponential map, an explicit formula for the differential dexp, whose role is specified in Section 2.6.4, can also be provided.

To this aim, let us first observe that representing the skew-symmetric matrices of $\Lambda^2(\mathbb{R}^3)$ by means of the map (4.41), the matrix commutator applies as follows:

$$[\widehat{\mathbf{a}}, \widehat{\mathbf{b}}] = (\mathbf{a} \times \mathbf{b})^{\widehat{}} = (\widehat{\mathbf{a}}\mathbf{b})^{\widehat{}} , \tag{4.52}$$

which can be verified by applying the skew-symmetric matrix $[\widehat{\mathbf{a}}, \widehat{\mathbf{b}}]$ to an arbitrary $\mathbf{c} \in \mathbb{R}^3$.

Please recall also that the matrix commutator represents the Lie bracket of $\mathfrak{so}(3)$, which coincides with the adjoint representation of $\mathfrak{so}(3)$ (cf. Definition 2.41). Then, in identifying $\mathfrak{so}(3)$ with $\mathbb{R}^3$, the evaluation $\mathsf{ad}_{\widehat{\mathbf{a}}}(\widehat{\mathbf{b}}) = [\widehat{\mathbf{a}}, \widehat{\mathbf{b}}]$ of the adjoint representation becomes

$$\mathsf{ad}_{\widehat{\mathbf{a}}}(\widehat{\mathbf{b}}) \cong \mathsf{ad}_{\mathbf{a}}(\mathbf{b}) = \widehat{\mathbf{a}}\mathbf{b} , \tag{4.53}$$

and the k-th power of the adjoint representation $\mathsf{ad}_{\widehat{\mathbf{a}}}$ takes the following explicit form

$$\mathsf{ad}_{\widehat{\mathbf{a}}}^k \cong \mathsf{ad}_{\mathbf{a}}^k = \widehat{\mathbf{a}}^k .$$

With the specifications here above, the operator $\mathbf{D}_{\widehat{\theta}}$ defined by (2.57) and associated with the differential of the exponential map, takes the following form:

$$\mathbf{D}_{\widehat{\theta}} = \sum_{k=0}^{\infty} (-1)^k \frac{\mathrm{ad}_{\widehat{\theta}}^k}{(k+1)!} \cong \mathbf{T}_{\theta} = \sum_{k=0}^{\infty} (-1)^k \frac{\widehat{\theta}^k}{(k+1)!} \,. \tag{4.54}$$

The *tangent operator* $\mathbf{T}_\theta$ is a linear operator whose action on $\mathbb{R}^3$ is equivalent to the one of $\mathbf{D}_{\widehat{\theta}}$ on $\mathfrak{so}(3)$:

$$\begin{aligned} \mathbf{T}_\theta : \mathbb{R}^3 &\to \mathbb{R}^3 \\ \mathbf{a} &\mapsto \mathbf{T}_\theta \mathbf{a} = \sum_{k=0}^{\infty} (-1)^k \frac{\widehat{\theta}^k \mathbf{a}}{(k+1)!} \,, \end{aligned} \tag{4.55}$$

where, by virtue of the equivalence $\mathfrak{so}(3) \cong \mathbb{R}^3$ specified by (4.41), and using (4.53), the following identity applies:

$$(\mathbf{T}_\theta \mathbf{a})^\frown = \mathbf{D}_{\widehat{\theta}}(\widehat{\mathbf{a}}) \,. \tag{4.56}$$

Then, the differential of the exponential map defined by (2.53), at a vector $\theta \cong \widehat{\theta} \in \mathfrak{so}(3)$, specializes to $SO(3)$ as follows:

$$\begin{aligned} \mathrm{dexp}|_\theta : \mathbb{R}^3 &\to T_{\mathbf{R}_\theta} SO(3) \\ \mathbf{a} &\mapsto \mathrm{dexp}|_\theta(\mathbf{a}) = \mathbf{R}_\theta (\mathbf{T}_\theta \mathbf{a})^\frown \,. \end{aligned} \tag{4.57}$$

Furthermore, with a procedure similar to the one proving the equivalence of the exponential map in (4.49) with the Rodrigues' rotation formula, the following explicit expression for $\mathbf{T}_\theta$ can be verified:

$$\mathbf{T}_\theta = \mathbf{I} - \frac{1-\cos\|\theta\|}{\|\theta\|} \frac{\widehat{\theta}}{\|\theta\|} + \left(1 - \frac{\sin\|\theta\|}{\|\theta\|}\right) \frac{\widehat{\theta}^2}{\|\theta\|^2} \,. \tag{4.58}$$

It is worth observing that the operator $\mathbf{T}_\theta$ maps any element $\mathbf{a}$ of $\mathbb{R}^3$, meant as the tangent space of the smooth manifold $\mathbb{R}^3$ at θ, to the vector $\mathbf{T}_\theta \mathbf{a}$ belonging to $\mathbb{R}^3 \cong \mathfrak{so}(3)$, isomorphic with the tangent space at the identity of $SO(3)$. Then, exploiting the Lie group structure of $SO(3)$, the tangent vector $(\mathbf{T}_\theta \mathbf{a})^\frown$ is mapped to $\mathbf{R}_\theta (\mathbf{T}_\theta \mathbf{a})^\frown$, which is a tangent vector to $SO(3)$ at $\mathbf{R}_\theta$, by left translation.

4.3.5 Logarithm Map of $SO(3)$

Since any rotation matrix $\mathbf{R} \in SO(3)$ can be expressed in the form of the Rodrigues' formula (4.45), the associated rotation vector can be derived by noting that the only skew-symmetric term results

$$\sin\theta\,\widehat{\mathbf{k}} = \frac{\mathbf{R} - \mathbf{R}^{\mathsf{T}}}{2},$$

while the trace of $\mathbf{R}$, applied to the symmetric part, gives

$$\mathrm{tr}(\mathbf{R}) = 1 - 2\cos\theta\,.$$

Then, the definition of the logarithm map for $SO(3)$ is

$$\begin{aligned} \log:\ SO(3) &\to \mathfrak{so}(3) \\ \mathbf{R} &\mapsto \widehat{\boldsymbol{\theta}}\,, \end{aligned} \tag{4.59}$$

where, evaluating $\theta = \arccos\left(\frac{\mathrm{tr}(\mathbf{R})-1}{2}\right)$, one has

$$\widehat{\boldsymbol{\theta}} = \log(\mathbf{R}) = \frac{\theta}{2\sin\theta}(\mathbf{R} - \mathbf{R}^{\mathsf{T}})\,. \tag{4.60}$$

Please recall from Theorem 2.48 that the matrix exponential $e^{\mathbf{X}}$ is a diffeomorphism between the matrix Lie algebra and the relevant matrix Lie group if the eigenvalues λ_i of $\mathbf{X}$, with $i = 1, \ldots, n$, satisfy the condition $-\pi < \Im(\lambda_i) < \pi$.

With respect to the group $SO(3)$, it is easy to verify that any skew-symmetric matrix $\widehat{\boldsymbol{\theta}} \in \mathfrak{so}(3)$ has a null real eigenvalue, while the other two eigenvalues are conjugate purely imaginary numbers:

$$\lambda_1 = 0\,, \qquad \lambda_{2,3} = \pm\theta i\,,$$

where θ is exactly the norm of the vector $\boldsymbol{\theta}$ associated with $\widehat{\boldsymbol{\theta}}$ through the map (4.41).

Then, since the matrix exponential $e^{\widehat{\boldsymbol{\theta}}}$ coincides with the matrix $\mathbf{R}_{\boldsymbol{\theta}}$ of the rotation by θ, the exponential map $\exp\colon \mathfrak{so}(3) \to SO(3)$ is a diffeomorphism if the angle of rotation satisfies $-\pi < \theta < \pi$.

In these conditions, the logarithm map given by (4.59) is the inverse of the exponential map defined by (4.49).

Moreover, limiting the analysis at $-\pi < \theta < \pi$, the differential of the logarithm map is also well-defined as $\mathsf{dlog} = \mathsf{dexp}^{-1}$, with the role to map a tangent vector $\mathbf{X}$ at $\mathbf{R} \in SO(3)$, to a vector in $\mathbb{R}^3$.

In order to find an explicit form for dlog, please notice from (4.57) that $\mathbf{X} \in T_{\mathbf{R}}SO(3)$ can be expressed in the form $\mathbf{X} = \mathbf{R}\widehat{\mathbf{b}}$, for a $\widehat{\mathbf{b}} \in \mathfrak{so}(3)$. Then, applying the left translation by $\mathbf{R}^\mathsf{T}$, one has

$$\widehat{\mathbf{b}} = \mathbf{R}^\mathsf{T}\mathbf{X}\,,$$

where the skew-symmetric matrix $\widehat{\mathbf{b}}$, playing the role of a tangent vector in $T_{\mathbf{I}}SO(3) \cong \mathfrak{so}(3)$, is the image of a matrix $\widehat{\mathbf{a}}$ through the operator $\mathbf{D}_{\log(\mathbf{R})}$.

Then, assuming $\boldsymbol{\theta}$ is the vector in $\mathbb{R}^3$ such that $\widehat{\boldsymbol{\theta}} = \log(\mathbf{R})$, the equivalence (4.56) implies

$$\widehat{\mathbf{b}} = \mathbf{D}_{\log(\mathbf{R})}(\widehat{\mathbf{a}}) = (\mathbf{T}_{\boldsymbol{\theta}}\mathbf{a})^\wedge\,.$$

whence, in terms of a linear transformation in $\mathbb{R}^3 \cong \mathfrak{so}(3)$, one infers

$$\mathbf{a} = \mathbf{T}_{\boldsymbol{\theta}}^{-1}\mathbf{b}.$$

The existence of $\mathbf{a}$, with the role of a tangent vector to $\mathbb{R}$ at $\boldsymbol{\theta}$, is assured by the diffeomorphic relation between $\mathfrak{so}(3) \cong \mathbb{R}^3$ and $SO(3)$, within the limitation $-\pi < \theta < \pi$.

Consequently, $\mathbf{T}_{\boldsymbol{\theta}}^{-1}$ is well-defined as the *inverse tangent operator* at $\boldsymbol{\theta}$, and its explicit expression is

$$\mathbf{T}_{\boldsymbol{\theta}}^{-1} = \mathbf{I} + \frac{1}{2}\widehat{\boldsymbol{\theta}} + \left(1 - \frac{\|\boldsymbol{\theta}\|}{2}\cot\frac{\|\boldsymbol{\theta}\|}{2}\right)\frac{\widehat{\boldsymbol{\theta}}^2}{\|\boldsymbol{\theta}\|^2}\,. \tag{4.61}$$

In conclusion, the differential of the logarithm map can be formally defined as

$$\begin{aligned} \mathsf{dlog}|_{\mathbf{R}} :\ & T_{\mathbf{R}}SO(3) \to \mathbb{R}^3 \\ & \mathbf{X} \mapsto \mathsf{dlog}|_{\mathbf{R}}\mathbf{X} = \mathbf{T}_{\boldsymbol{\theta}}^{-1}\mathbf{b}\,, \end{aligned} \tag{4.62}$$

with $\widehat{\mathbf{b}} = \mathbf{R}^\mathsf{T}\mathbf{X}$ and $\widehat{\boldsymbol{\theta}} = \log(\mathbf{R})$.

4.4 Rigid Motions in 3-Dimensional Spaces

The special Euclidean group introduced by Definition 4.16 gathers the proper rigid motions of a Euclidean affine space $\mathscr{E}$.

When one refers to a three-dimensional Euclidean space $\mathscr{E}$, any rigid motion, except for pure translations, can be described as a rotation about an axis a of $\mathscr{E}$ and a translation along the same axis. Such a property is the statement of the *Chasles' theorem*.

Let us consider a point $\mathsf{P} \in \mathscr{E}$ and suppose that the axis of rotation a passes through P. By fixing P, any point $\mathsf{Q} \in \mathscr{E}$ is uniquely identified by a vector $\overrightarrow{\mathsf{PQ}}$ of the associated vector space $\mathcal{E}$ (cf. Section 3.1).

Then, a rotation by an angle θ around the axis a, identified by the unit vector $\mathbf{k}$, is described by the transformation $\mathbf{R}_{(\theta,\mathbf{k})} = \mathbf{R}_{\boldsymbol{\theta}}$, given by (4.45), while a translation along a is represented by a vector in the form $d\mathbf{k}$, for some scalar $d \in \mathbb{R}$.

Consequently, exploiting the representation of the Euclidean motions as in (4.23), this kind of transformation can be expressed as

$$\mathcal{H}_a \cong (d\mathbf{k}, \mathbf{R}_{\boldsymbol{\theta}})\,. \tag{4.63}$$

Please notice that, denoting the rotation around P as $\mathcal{H}_\mathsf{P} \cong (\mathbf{o}, \mathbf{R}_{\boldsymbol{\theta}})$ and the translation along $\mathbf{k}$ as $t_{d\mathbf{k}} \cong (d\mathbf{k}, \mathrm{id}_{\mathcal{E}})$, the following decomposition applies:

$$\mathcal{H}_a = t_{d\mathbf{k}}\mathcal{H}_\mathsf{P} = \mathcal{H}_\mathsf{P} t_{d\mathbf{k}}\,,$$

where the transformations $\mathcal{H}_\mathsf{P}$ and $t_{d\mathbf{k}}$ do commute because $\mathbf{k}$ is an eigenvector of the linear map $\mathbf{R}_{\boldsymbol{\theta}}$.

In order to describe an arbitrary rigid motion of the Euclidean space, please observe that the point P can be reached, starting from an arbitrary point $\mathsf{O} \in \mathscr{E}$, by means of a translation $t_\mathbf{p}$ by the vector $\mathbf{p} = \overrightarrow{\mathsf{OP}}$.

For this reason, a Euclidean transformation $\mathcal{H}$ can be obtained by first applying a translation $t_{-\mathbf{p}}$, which makes P coincident with O. Then, with the axis a passing through O, the transformation $\mathcal{H}_a$ is applied and finally the inverse translation $t_{\mathbf{p}}$ brings the axis a back to the original configuration:

$$\mathcal{H} = t_{\mathbf{p}}\mathcal{H}_a t_{-\mathbf{p}} \cong (-\mathbf{p}, \mathbf{I})(d\mathbf{k}, \mathbf{R}_\theta)(\mathbf{p}, \mathbf{I})\,, \tag{4.64}$$

which, using the composition rule (4.24), simplifies in

$$\mathcal{H} \cong (\mathbf{v}, \mathbf{R}) = \big((\mathbf{I} - \mathbf{R}_\theta)\mathbf{p} + d\mathbf{k}, \mathbf{R}_\theta\big)\,. \tag{4.65}$$

In conclusion, a proper rigid motion $\mathcal{H} \in SE(\mathscr{E})$, consisting of an orthogonal transformation $\mathbf{R} \in SO(\mathcal{E})$ and a translation by a vector $\mathbf{v} \in \mathcal{E}$, is equivalent to a rotation of an angle θ around an axis a, having the direction of $\mathbf{k}$ and passing through a point P at the location $\mathbf{p} = \overrightarrow{\mathsf{OP}}$, along with a translation by $d\mathbf{k}$ in the same direction of a.

The equivalence of these two characterization of the motion $\mathcal{H}$ is assured by setting $\mathbf{R} = \mathbf{R}_\theta$ and $\mathbf{v} = (\mathbf{I} - \mathbf{R}_\theta)\mathbf{p} + d\mathbf{k}$.

It is worth remarking that an analogous description of the motion $\mathcal{H}$ can be made by using the homogeneous representation given by (4.27).

Actually, the rigid motion with respect the axis a in (4.63) can be expressed as

$$\mathcal{H}_a \cong \mathbf{H}_a = \begin{bmatrix} \mathbf{R}_\theta & d\mathbf{k} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix}, \tag{4.66}$$

where, since the point P of a has been considered fixed in $\mathscr{E}$, the matrix $\mathbf{H}_a$ is the homogeneous representation of the affine transformation $\mathcal{H}_a$ with respect to a reference frame $\mathscr{F}_\mathsf{P} = (\mathsf{P}; \mathcal{B})$, having P as origin (cf. Definition 3.17).

Moreover, the composition in (4.64) reads

$$\mathcal{H} \cong \mathbf{H} = \begin{bmatrix} \mathbf{I} & \mathbf{p} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_\theta & d\mathbf{k} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{p} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix}, \tag{4.67}$$

which simplifies in

$$\mathcal{H} \cong \mathbf{H} = \begin{bmatrix} \mathbf{R}_\theta & \mathbf{v} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_\theta & (\mathbf{I} - \mathbf{R}_\theta)\mathbf{p} + d\mathbf{k} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix}. \tag{4.68}$$

Please observe that the composition (4.67) can be seen as the change of frame from $\mathscr{F}_\mathsf{O} = (\mathsf{O};\mathcal{B})$ to $\mathscr{F}_\mathsf{P} = (\mathsf{P};\mathcal{B})$, followed by the rigid transformation $\mathbf{H}_a$ expressed in $\mathscr{F}_\mathsf{P}$ and a further change of frame to $\mathscr{F}_\mathsf{O}$.

Screw Motion

Let us consider the translation vector $\mathbf{v}$ of the Euclidean motion $\mathcal{H} \cong (\mathbf{v}, \mathbf{R}_\theta)$, expressed in the form $\mathbf{v} = (\mathbf{I} - \mathbf{R}_\theta)\mathbf{p} + d\mathbf{k}$.

Using the identity (4.113), one finds

$$\mathbf{v} = -\mathbf{T}_\theta^\mathsf{T}\widehat{\theta}\mathbf{p} + d\mathbf{k} = -\mathbf{T}_\theta^\mathsf{T}(\theta \times \mathbf{p}) + d\mathbf{k} = \mathbf{T}_\theta^\mathsf{T}(\mathbf{p} \times \theta) + d\mathbf{k}\,,$$

which, since $\mathbf{k}$ is an eigenvector of $\mathbf{T}_\theta$, can be written as

$$\mathbf{v} = \mathbf{T}_\theta^\mathsf{T}(d\mathbf{k} + \mathbf{p} \times \theta) = \mathbf{T}_\theta^\mathsf{T}\mathbf{u}\,, \tag{4.69}$$

where it has been set

$$\mathbf{u} = d\mathbf{k} + \mathbf{p} \times \theta\,. \tag{4.70}$$

Then, the Euclidean motion $\mathcal{H}$ can be represented as

$$\mathcal{H} \cong (\mathbf{T}_\theta^\mathsf{T}\mathbf{u}, \mathbf{R}_\theta)\,. \tag{4.71}$$

whose homogeneous representation results

$$\mathcal{H} \cong \mathbf{H} = \begin{bmatrix} \mathbf{R}_\theta & \mathbf{T}_\theta^\mathsf{T}\mathbf{u} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix}. \tag{4.72}$$

Please notice that, since both the rotation $\mathbf{R}_\theta$ and the tangent operator $\mathbf{T}_\theta$ depend on θ, once the rotation vector has been fixed, the Euclidean motion is a function only of the vector $\mathbf{u}$ expressed by (4.70).

The pair $(\theta, \mathbf{u})$, which is called a *screw*, completely identifies the rigid motion in (4.71), or equivalently the homogeneous representation in (4.72), which is called a *screw motion*.

It is worth observing that the rotation component of the screw $(\boldsymbol{\theta}, \mathbf{u})$ has the meaning of defining the angle of rotation $\theta = \|\boldsymbol{\theta}\|$ and the axis of rotation $\mathbf{k} = \boldsymbol{\theta}/\theta$, while the role of the translation component $\mathbf{u}$ follows from (4.70).

Actually, the projection of $\mathbf{u}$ on $\mathbf{k}$ is

$$d = \mathbf{u} \cdot \mathbf{k}\,, \tag{4.73}$$

which represents the pure translation contribution to the rigid motion parallel to the axis of the screw, while the other addend in (4.70) is

$$\mathbf{p} \times \boldsymbol{\theta} = \mathbf{u} - d\mathbf{k}\,. \tag{4.74}$$

which represents the moment of the rotation vector $\boldsymbol{\theta}$ with respect to the reference point O.

Furthermore, a screw $(\boldsymbol{\theta}, \mathbf{u})$ can also be intended as a vector $\mathbf{s} \in \mathbb{R}^6$, that is

$$(\boldsymbol{\theta}, \mathbf{u}) \cong \mathbf{s} = \begin{bmatrix} \mathbf{s}_{\mathsf{r}} \\ \mathbf{s}_{\mathsf{t}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{u} \end{bmatrix}, \tag{4.75}$$

with the clear meaning that $\mathbf{s}_{\mathsf{r}} = \boldsymbol{\theta}$ is the rotation vector, associated with the matrix $\mathbf{R}$, and $\mathbf{s}_{\mathsf{t}} = \mathbf{u}$ is the translation vector providing $\mathbf{v} = \mathbf{T}_{\boldsymbol{\theta}}^{\mathsf{T}}\mathbf{u}$.

4.4.1 Lie Group Structure of $SE(3)$

Similarly to the special orthogonal group analyzed in Section 4.3.3, the group isomorphism (4.30) allows one to consider the matrix Lie group $SE(3)$ as a reference for the group $SE(\mathscr{E})$ of the proper rigid transformations of Euclidean space $\mathscr{E}$.

Please recall from (4.28) that any element of $SE(3)$ is in the form

$$\mathbf{H} = \begin{bmatrix} \mathbf{R} & \mathbf{v} \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix},$$

with $\mathbf{R}$ a rotation matrix in $SO(3)$ and $\mathbf{v}$ a vector in $\mathbb{R}$.

In order to identify the Lie algebra of $SE(3)$, let us consider an arbitrary curve on the group, starting at the identity $\mathbf{I}$, by introducing a parameterization $t \mapsto \mathbf{H}(t)$ such that $\mathbf{H}(0) = \mathbf{I}$, whose velocity at the identity results

$$\mathbf{H}'(0) = \begin{bmatrix} \mathbf{R}'(0) & \mathbf{v}'(0) \\ \mathbf{o}^\mathsf{T} & 0 \end{bmatrix}.$$

In Section 4.3.3 it has been shown that $\mathbf{R}'(0)$ is a skew-symmetric matrix, with the subsequent identification of $\mathsf{Lie}\big(SO(3)\big)$ with $\mathfrak{so}(3)$ (cf. Definition 4.17). At the same time, since the parameterization $t \mapsto \mathbf{v}(t)$ is a curve of $\mathbb{R}^3$, the velocity $\mathbf{v}'(0)$ is a vector in $\mathbb{R}^3$.

Then, the velocity $\mathbf{H}'(0)$, which is a vector tangent to $SE(3)$ at the identity $\mathbf{I}$, is completely specified by a skew-symmetric matrix $\mathbf{W} \in \Lambda^2(\mathbb{R}^3)$ and a vector $\mathbf{x} \in \mathbb{R}^3$, so that the tangent space to $T_\mathbf{I} SE(3)$ is the space of the matrices in the form

$$\begin{bmatrix} \mathbf{W} & \mathbf{x} \\ \mathbf{o}^\mathsf{T} & 0 \end{bmatrix},$$

and, with the matrix commutator, the Lie algebra $\mathfrak{se}(3)$ is accordingly characterized.

Definition 4.18. The *special Euclidean Lie algebra* of order 3, denoted as $\mathfrak{se}(3)$, is the subalgebra of the general linear algebra $\mathfrak{gl}(4)$ such that the leading principal minor of any element is a three-dimensional skew-symmetric matrix:

$$\mathfrak{se}(3) = \left\{ \begin{bmatrix} \mathbf{W} & \mathbf{x} \\ \mathbf{o}^\mathsf{T} & 0 \end{bmatrix} \in \mathfrak{gl}(4) \;\middle|\; \mathbf{W}^\mathsf{T} = -\mathbf{W} \right\}. \tag{4.76}$$

Exploiting the usual identification of the Lie algebra of a Lie group with the tangent space at the identity, one can write

$$\mathsf{Lie}\big(SE(3)\big) \cong \mathfrak{se}(3)\,. \tag{4.77}$$

Please recall that any skew-symmetric matrix results from a vector in $\mathbb{R}^3$ by means of the isomorphism (4.41). Consequently, it is useful to consider an analogous map also for the elements of $\mathfrak{se}(3)$:

$$\begin{aligned}(\cdot)^{\widehat{}} :\ & \mathbb{R}^6 \to \mathfrak{se}(3)\\ & \mathbf{h} \mapsto \widehat{\mathbf{h}} = \begin{bmatrix} \widehat{\mathbf{a}} & \mathbf{x} \\ \mathbf{o}^\mathsf{T} & 0 \end{bmatrix},\end{aligned} \tag{4.78}$$

where $\mathbf{h}$ is the column matrix gathering $\mathbf{a} \in \mathbb{R}^3$ and $\mathbf{x} \in \mathbb{R}^3$:

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_\mathrm{r} \\ \mathbf{h}_\mathrm{t} \end{bmatrix} = \begin{bmatrix} \mathbf{a} \\ \mathbf{x} \end{bmatrix}.$$

Exponential Map and Logarithm Map of $SE(3)$

As a matrix Lie group, the exponential map of $SE(3)$ is given by the matrix exponential defined by (2.42).

With the aim to provide an explicit expression of the exponential map, let us consider the k-th power of a matrix $\widehat{\mathbf{h}} \in \mathfrak{se}(3)$, which is given by assembling $\widehat{\boldsymbol{\theta}} \in \mathfrak{so}(3)$ and $\mathbf{x} \in \mathbb{R}^3$. By induction, it is easy to verify that $\widehat{\mathbf{h}}^k$ results

$$\widehat{\mathbf{h}}^k = \begin{bmatrix} \widehat{\boldsymbol{\theta}} & \mathbf{x} \\ \mathbf{o}^\mathsf{T} & 0 \end{bmatrix}^k = \begin{bmatrix} \widehat{\boldsymbol{\theta}}^k & \widehat{\boldsymbol{\theta}}^{k-1}\mathbf{x} \\ \mathbf{o}^\mathsf{T} & 0 \end{bmatrix},$$

whence the power series defining the matrix exponential specializes to

$$e^{\widehat{\mathbf{h}}} = \sum_{k=0}^{\infty} \frac{\widehat{\mathbf{h}}^k}{k!} = \begin{bmatrix} \mathbf{I} & \mathbf{o} \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} + \begin{bmatrix} \sum_{k=1}^{\infty} \frac{\widehat{\boldsymbol{\theta}}^k}{k!} & \sum_{k=1}^{\infty} \frac{\widehat{\boldsymbol{\theta}}^{k-1}}{k!}\mathbf{x} \\ \mathbf{o}^\mathsf{T} & 0 \end{bmatrix}. \tag{4.79}$$

Please notice that principal leading minor of the matrix here above is exactly the exponential of $\widehat{\boldsymbol{\theta}}$, which gives the rotation matrix $\mathbf{R}_{\boldsymbol{\theta}}$, consistently with the definition (4.49) of the exponential map of $SO(3)$:

$$\mathbf{R}_{\boldsymbol{\theta}} = e^{\widehat{\boldsymbol{\theta}}} = \mathbf{I} + \sum_{k=1}^{\infty} \frac{\widehat{\mathbf{h}}^k}{k!}.$$

In addition, the power series multiplying the vector $\mathbf{x}$ in (4.79) can be expressed as

$$\sum_{k=0}^{\infty} \frac{\widehat{\theta}^k}{(k+1)!} = \sum_{k=0}^{\infty} (-1)^k \frac{(-\widehat{\theta})^k}{(k+1)!} \,,$$

which, comparing with (4.55), is the tangent operator relevant to $-\theta$, i.e. $\mathbf{T}_{-\theta} = \mathbf{T}_\theta^\mathsf{T}$.

In conclusion, the matrix exponential in (4.79) reads

$$e^{\widehat{\mathbf{h}}} = \mathbf{H}_\mathbf{h} = \begin{bmatrix} \mathbf{R}_\theta & \mathbf{T}_\theta^\mathsf{T}\mathbf{x} \\ \mathbf{o}^\mathsf{T} & 0 \end{bmatrix}, \tag{4.80}$$

and the exponential map of $SE(3)$ is defined as follows:

$$\begin{aligned} \exp:\ \mathfrak{se}(3) &\to SE(3) \\ \widehat{\mathbf{h}} &\mapsto \exp(\widehat{\mathbf{h}}) = \mathbf{H}_\mathbf{h} \,, \end{aligned} \tag{4.81}$$

where $\mathbf{H}_\mathbf{h}$ is the homogeneous matrix given by (4.80).

It is important to observe that, comparing (4.80) with (4.72), the matrix $\mathbf{H}_\mathbf{h}$ is exactly the homogeneous representation of the motion associated with the screw $(\theta, \mathbf{x})$.

This means that any vector $\mathbf{h} \in \mathbb{R}^6$, or equivalently any $\widehat{\mathbf{h}} \in \mathfrak{se}(3)$, is a screw representing a rigid motion of the Euclidean space, and the Lie algebra $\mathfrak{se}(3)$ can itself be defined as the space of the screws. Also, using the representation (4.75), the rotation and the translation components of $\mathbf{h}$ are $\mathbf{h}_r$ and $\mathbf{h}_t$, respectively, while the explicit representation of the screw motion associate with $\widehat{\mathbf{h}} \in \mathfrak{se}(3)$ is given by the matrix exponential in (4.80).

Let us now suppose that an homogeneous matrix $\mathbf{H} \in SE(3)$ is assigned. The map providing the screw $\widehat{\mathbf{h}} \in \mathfrak{se}(3)$ is the logarithm map of $SE(3)$, defined as:

$$\begin{aligned} \log:\ SE(3) &\to \mathfrak{se}(3) \\ \mathbf{H} &\mapsto \widehat{\mathbf{h}} = \begin{bmatrix} \widehat{\theta} & \mathbf{u} \\ \mathbf{o}^\mathsf{T} & 0 \end{bmatrix}, \end{aligned} \tag{4.82}$$

where, being $\mathbf{H} \cong (\mathbf{v}, \mathbf{R})$, the skew-symmetric matrix $\widehat{\boldsymbol{\theta}}$ results from applying the logarithm map of $SO(3)$ at $\mathbf{R}$, as specified by (4.59). Also, since the translation vector of $\mathbf{H}$ is in the form $\mathbf{v} = \mathbf{T}_{\boldsymbol{\theta}}^{\mathsf{T}}\mathbf{u}$, the vector $\mathbf{u} \in \mathbb{R}^3$ of the screw $\widehat{\mathbf{h}}$ is evaluate as

$$\mathbf{u} = \mathbf{T}_{\boldsymbol{\theta}}^{-\mathsf{T}}\mathbf{v}\,, \tag{4.83}$$

where $\mathbf{T}_{\boldsymbol{\theta}}^{-\mathsf{T}} = \mathbf{T}_{-\boldsymbol{\theta}}^{-1}$ is the inverse tangent operator of $SO(3)$, relevant to $-\boldsymbol{\theta}$, evaluated by means of (4.61).

It is worth emphasizing that, similarly to what discussed in Section 4.3.5, the exponential map $\exp : \mathfrak{se}(3) \to SE(3)$ results a diffeomorphism, and specifically a bijection, if the spectrum of the matrix $\widehat{\mathbf{h}} \in \mathfrak{se}(3)$ is characterized by $-\pi < \Im(\lambda_i) < \pi$, with $i = 1, \ldots, 4$.

Consequently, since any matrix in $\mathfrak{se}(3)$ is in the form

$$\widehat{\mathbf{h}} = \begin{bmatrix} \widehat{\boldsymbol{\theta}} & \mathbf{u} \\ \mathbf{o}^{\mathsf{T}} & 0 \end{bmatrix},$$

the eigenvalues of $\widehat{\mathbf{h}}$ are $\lambda_{1,2} = 0$ and $\lambda_{3,4} = \pm\theta i$, with $\theta = \|\boldsymbol{\theta}\|$ representing the norm of the rotation vector $\boldsymbol{\theta}$ relevant to the screw $\widehat{\mathbf{h}}$.

Then, the same condition $-\pi < \theta < \pi$ applies to both $SO(3)$ and $SE(3)$ in order to characterize the exponential map as a diffeomorphism. In addition, one should also notice that the restriction about the angle of rotation θ assures that the operator $\mathbf{T}_{\boldsymbol{\theta}}^{-\mathsf{T}}$, which is involved in the logarithm map of $SE(3)$ through (4.83), is well-defined as the inverse of the tangent operator $\mathbf{T}_{-\boldsymbol{\theta}}$ of $SO(3)$.

Adjoint Representation of $SE(3)$ and $\mathfrak{se}(3)$

The adjoint representation of the a Lie group is introduced in Definition 2.40 and for a matrix Lie group specializes to the map (2.50).

In order to find and explicit expression, let us consider a an homogeneous matrix $\mathbf{H} \cong (\mathbf{R}, \mathbf{v})$ whose adjoint map $\mathrm{Ad}_{\mathbf{H}} : \mathfrak{se}(3) \to \mathfrak{se}(3)$ applied at a screw $\widehat{\mathbf{h}}$ provides the screw $\widehat{\mathbf{s}}$:

$$\widehat{\mathbf{s}} = \mathrm{Ad}_{\mathbf{H}}(\widehat{\mathbf{h}}) = \mathbf{H}\widehat{\mathbf{h}}\mathbf{H}^{-1}\,.$$

The explicit matrix representation is

$$\begin{bmatrix} \widehat{\mathbf{s}}_{\mathsf{r}} & \mathbf{s}_{\mathsf{t}} \\ \mathbf{o}^{\mathsf{T}} & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{v} \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{h}}_{\mathsf{r}} & \mathbf{h}_{\mathsf{t}} \\ \mathbf{o}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}^{\mathsf{T}} & -\mathbf{R}^{\mathsf{T}}\mathbf{v} \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix},$$

whence

$$\begin{bmatrix} \widehat{\mathbf{s}}_{\mathsf{r}} & \widehat{\mathbf{s}}_{\mathsf{t}} \\ \mathbf{o}^{\mathsf{T}} & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{R}\widehat{\mathbf{h}}_{\mathsf{r}}\mathbf{R}^{\mathsf{T}} & -\mathbf{R}\widehat{\mathbf{h}}_{\mathsf{r}}\mathbf{R}^{\mathsf{T}}\mathbf{v} + \mathbf{R}\mathbf{h}_{\mathsf{t}} \\ \mathbf{o}^{\mathsf{T}} & 0 \end{bmatrix}.$$

It is easy to recognize that the rotation component results

$$\widehat{\mathbf{s}}_{\mathsf{r}} = \mathbf{R}\widehat{\mathbf{h}}_{\mathsf{r}}\mathbf{R}^{\mathsf{T}} = (\mathbf{R}\mathbf{h})^{\widehat{}}\,,$$

as well as the translation vector is

$$\mathbf{s}_{\mathsf{t}} = -\mathbf{R}\widehat{\mathbf{h}}_{\mathsf{r}}\mathbf{R}^{\mathsf{T}}\mathbf{v} + \mathbf{R}\mathbf{h}_{\mathsf{t}} = -(\mathbf{R}\mathbf{h}_{\mathsf{r}})^{\widehat{}}\,\mathbf{v} + \mathbf{R}\mathbf{h}_{\mathsf{t}} = \widehat{\mathbf{v}}\mathbf{R}\mathbf{h}_{\mathsf{r}} + \mathbf{R}\mathbf{h}_{\mathsf{t}}\,,$$

where the last step is justified by the following identity:

$$-(\mathbf{R}\mathbf{h}_{\mathsf{r}})^{\widehat{}}\,\mathbf{v} = -(\mathbf{R}\mathbf{h}_{\mathsf{r}}) \times \mathbf{v} = \mathbf{v} \times (\mathbf{R}\mathbf{h}_{\mathsf{r}}) = \widehat{\mathbf{v}}\mathbf{R}\mathbf{h}_{\mathsf{r}}\,.$$

Then, the two vector relations $\mathbf{s}_{\mathsf{r}} = \mathbf{R}\mathbf{h}_{\mathsf{r}}$ and $\mathbf{s}_{\mathsf{t}} = \widehat{\mathbf{v}}\mathbf{R}\mathbf{h}_{\mathsf{r}} + \mathbf{R}\mathbf{h}_{\mathsf{t}}$ can be written in matrix form as

$$\begin{bmatrix} \mathbf{s}_{\mathsf{r}} \\ \mathbf{s}_{\mathsf{t}} \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \widehat{\mathbf{v}}\mathbf{R} & \mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{h}_{\mathsf{r}} \\ \mathbf{h}_{\mathsf{t}} \end{bmatrix},$$

so that, the identification $\mathfrak{se}(3) \cong \mathbb{R}^6$ allows one to identify the adjoint representation of $\mathbf{H}$ with the linear operator $\mathrm{Ad}_{\mathbf{H}} : \mathbb{R}^6 \to \mathbb{R}^6$ having the following representation

$$\mathrm{Ad}_{\mathbf{H}} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \widehat{\mathbf{v}}\mathbf{R} & \mathbf{R} \end{bmatrix}, \tag{4.84}$$

and satisfying

$$\mathrm{Ad}_{\mathbf{H}}\,\mathbf{h} = \mathrm{Ad}_{\mathbf{H}}(\widehat{\mathbf{h}}) = \mathbf{H}\widehat{\mathbf{h}}\mathbf{H}^{-1}\,. \tag{4.85}$$

As far as the Lie algebra, the adjoint representation of is given by the Lie bracket operation (cf. Definition 2.41).

With specific reference to the matrix Lie algebra $\mathfrak{se}(3)$, given $\widehat{\mathbf{x}} \in \mathfrak{se}(3)$, the adjoint representation $\mathsf{ad}_{\widehat{\mathsf{x}}}$ is the map $\mathfrak{se}(3) \to \mathfrak{se}(3)$ such that applied at $\widehat{\mathbf{y}}$ provides the screw $\widehat{\mathbf{s}}$:

$$\widehat{\mathbf{s}} = \mathsf{ad}_{\widehat{\mathsf{x}}}(\widehat{\mathbf{y}}) = [\widehat{\mathbf{x}}, \widehat{\mathbf{y}}]\,, \tag{4.86}$$

which, by explicit computation, reads

$$\begin{bmatrix} \widehat{\mathbf{s}}_{\mathsf{r}} & \mathbf{s}_{\mathsf{t}} \\ \mathbf{o}^{\mathsf{T}} & 0 \end{bmatrix} = \begin{bmatrix} [\widehat{\mathbf{x}}_{\mathsf{r}}, \widehat{\mathbf{y}}_{\mathsf{r}}] & \widehat{\mathbf{x}}_{\mathsf{r}}\mathbf{y}_{\mathsf{t}} - \widehat{\mathbf{y}}_{\mathsf{r}}\mathbf{x}_{\mathsf{t}} \\ \mathbf{o}^{\mathsf{T}} & 0 \end{bmatrix} = \begin{bmatrix} (\widehat{\mathbf{x}}_{\mathsf{r}}\mathbf{y}_{\mathsf{t}})^{\wedge} & \widehat{\mathbf{x}}_{\mathsf{r}}\mathbf{y}_{\mathsf{t}} + \widehat{\mathbf{x}}_{\mathsf{t}}\mathbf{y}_{\mathsf{r}} \\ \mathbf{o}^{\mathsf{T}} & 0 \end{bmatrix},$$

where, since $\widehat{\mathbf{x}}_{\mathsf{r}}$ and $\widehat{\mathbf{y}}_{\mathsf{r}}$ are matrices of $\mathfrak{so}(3)$, the identity $[\widehat{\mathbf{x}}_{\mathsf{r}}, \widehat{\mathbf{y}}_{\mathsf{r}}] = (\widehat{\mathbf{x}}_{\mathsf{r}}\mathbf{y}_{\mathsf{t}})^{\wedge}$ follows from (4.52), as well as $\widehat{\mathbf{y}}_{\mathsf{r}}\mathbf{x}_{\mathsf{t}} = -\widehat{\mathbf{x}}_{\mathsf{t}}\mathbf{y}_{\mathsf{r}}$ is derived from $\widehat{\mathbf{y}}_{\mathsf{r}}\mathbf{x}_{\mathsf{t}} = \mathbf{y}_{\mathsf{r}} \times \mathbf{x}_{\mathsf{t}}$ coinciding with $-\mathbf{x}_{\mathsf{t}} \times \mathbf{y}_{\mathsf{r}} = -\widehat{\mathbf{x}}_{\mathsf{t}}\mathbf{y}_{\mathsf{r}}$.

In conclusion, the rotation vector of the screw $\widehat{\mathbf{s}}$ is $\mathbf{s}_{\mathsf{r}} = \widehat{\mathbf{x}}_{\mathsf{r}}\mathbf{y}_{\mathsf{t}}$ and the translation component is $\mathbf{s}_{\mathsf{t}} = \widehat{\mathbf{x}}_{\mathsf{r}}\mathbf{y}_{\mathsf{t}} + \widehat{\mathbf{x}}_{\mathsf{t}}\mathbf{y}_{\mathsf{r}}$, so that the evaluation of (4.86) can be equivalently represented as

$$\begin{bmatrix} \mathbf{s}_{\mathsf{r}} \\ \mathbf{s}_{\mathsf{t}} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{x}}_{\mathsf{r}} & \mathbf{0} \\ \widehat{\mathbf{x}}_{\mathsf{t}} & \widehat{\mathbf{x}}_{\mathsf{r}} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{\mathsf{r}} \\ \mathbf{y}_{\mathsf{t}} \end{bmatrix},$$

that is

$$\mathsf{ad}_{\mathsf{x}}\,\mathbf{y} \cong \mathsf{ad}_{\widehat{\mathsf{x}}}(\widehat{\mathbf{y}}) = [\widehat{\mathbf{x}}, \widehat{\mathbf{y}}]\,, \tag{4.87}$$

where the linear operator $\mathsf{ad}_{\mathsf{x}} \colon \mathbb{R}^6 \to \mathbb{R}^6$ explicitly results

$$\mathsf{ad}_{\mathsf{x}} = \begin{bmatrix} \widehat{\mathbf{x}}_{\mathsf{r}} & \mathbf{0} \\ \widehat{\mathbf{x}}_{\mathsf{t}} & \widehat{\mathbf{x}}_{\mathsf{r}} \end{bmatrix}. \tag{4.88}$$

Differential of the Exponential Map of $SE(3)$

Similarly to the differential of the exponential map of $SO(3)$, given by (4.57), the differential dexp of $SE(3)$ at a vector $\mathbf{x}$, relevant to the screw $\widehat{\mathbf{x}} \in \mathfrak{se}(3)$, is defined as follows

$$\begin{aligned} \operatorname{dexp}\big|_{\mathbf{x}} : \mathbb{R}^6 &\to T_{\exp(\widehat{\mathbf{x}})} SE(3) \\ \mathbf{a} &\mapsto \operatorname{dexp}\big|_{\mathbf{x}}(\mathbf{y}) = \exp(\widehat{\mathbf{x}})(\mathbf{T}_{\mathbf{x}}\mathbf{y})^{\widehat{}} . \end{aligned} \tag{4.89}$$

where $\mathbf{T}_{\mathbf{x}} : \mathbb{R}^6 \to \mathbb{R}^6$ is the tangent operator of $SE(3)$. Its action on $\mathbb{R}^6$ is equivalent to the one of the operator $\mathbf{D}_{\widehat{\mathbf{x}}}$ on $\mathfrak{se}(3)$:

$$(\mathbf{T}_{\mathbf{x}}\mathbf{y})^{\widehat{}} = \mathbf{D}_{\widehat{\mathbf{x}}}(\widehat{\mathbf{y}}) = \sum_{k=0}^{\infty} (-1)^k \frac{\operatorname{ad}_{\widehat{\mathbf{x}}}^k(\widehat{\mathbf{y}})}{(k+1)!} , \tag{4.90}$$

with $\mathbf{D}_{\widehat{\mathbf{x}}}$ introduced in Proposition 2.46.

Expressions of the tangent operator $\mathbf{T}_{\mathbf{x}}$ of $SE(3)$ and the inverse $\mathbf{T}_{\mathbf{x}}^{-1}$ are provided, among others, by Sonneville et al. (2014). However, a compact and useful representation, based upon the algebra of dual numbers, will be provided in Section 4.5.2.

4.5 Rigid Motions via Dual Numbers

In Section 4.4 it has been shown how any rigid motion is associated with a screw by means of (4.72).

From an algebraic point of view, the relation between a rigid motion matrix $\mathbf{H}$ and a screw $(\boldsymbol{\theta}, \mathbf{u})$ is the same as the relation between a rotation matrix $\mathbf{R}$ and a rotation vector $\boldsymbol{\theta}$. However, using screws to manage rigid motions is quite complex if compared with the pure rotation counterpart.

Manipulating screw coordinate transformations can be significantly simplified if the formalism of dual numbers is adopted. Actually, such an approach consists in using the same expressions applied for the operators of $SO(3)$, to be referred to the ring $\mathbb{D}$ of dual numbers rather than to the field $\mathbb{R}$ of real ones.

4.5.1 Dual Numbers

Dual numbers have been originally introduced by Clifford (1871) and have been widely applied in kinematic analysis of rigid body systems (see, e.g., Angeles (1998), Bottema and Roth (1990), Dimentberg (1968), Fischer (1999), McCarthy (1990), Pennestrì and Valentini (2009), Yang and Freudenstein (1964)).

The space $\mathbb{D}$ of *dual numbers* is a commutative ring containing zero divisors (cf. Appendix A.2.3). It also represents a two-dimensional commutative algebra over the real field $\mathbb{R}$, with the sum and the product defined as follows:

- $\tilde{a} + \tilde{b} = (a, a^{\mathsf{d}}) + (b, b^{\mathsf{d}}) = (a + b, a^{\mathsf{d}} + b^{\mathsf{d}})\,, \ \ \forall\,(a, a^{\mathsf{d}}), (b, b^{\mathsf{d}}) \in \mathbb{R}^2$;
- $\tilde{a}\tilde{b} = (a, a^{\mathsf{d}}) \cdot (b, b^{\mathsf{d}}) = (ab, ab^{\mathsf{d}} + bc^{\mathsf{d}})\,, \ \ \forall\,(a, a^{\mathsf{d}}), (b, b^{\mathsf{d}}) \in \mathbb{R}^2$.

A dual number can be more concisely defined by means of a non-real unit ϵ as follows.

Definition 4.19. Let $\epsilon \neq 0$ be a nonzero nilpotent unity, named the *dual unity*, satisfying $\epsilon^2 = 0$. A *dual number* $\tilde{a}$ is the number given by the sum of a *real* part a and a *dual* part a^{d}:

$$\tilde{a} = a + \epsilon a^{\mathsf{d}}\,,$$

with $a, a^{\mathsf{d}} \in \mathbb{R}$.

Since the sum is associative and the multiplication is a bilinear operation, it is trivial to verify that, for any pair of dual numbers $\tilde{a}, \tilde{b} \in \mathbb{D}$, the following relations do apply:

$$\tilde{a} + \tilde{b} = (a + b) + \epsilon(a^{\mathsf{d}} + b^{\mathsf{d}})\,,$$

as well as

$$\tilde{a}\tilde{b} = ab + \epsilon(ab^{\mathsf{d}} + ba^{\mathsf{d}})\,.$$

A direct consequence is that, exploiting the Taylor series expansion, any analytical function $f\colon U \subseteq \mathbb{R} \to \mathbb{R}$ can be easily evaluated at a dual number

$\tilde{x} = x + \epsilon x^{\mathsf{d}}$ as follows:

$$f(\tilde{x}) = f(x + \epsilon x^{\mathsf{d}}) = f(x) + \sum_{i=1}^{\infty} \frac{f^{(i)}(x)}{i!} (\epsilon x^{\mathsf{d}})^i ,$$

whence, using the property $\epsilon^2 = \epsilon^3 = \ldots = 0$, one finds

$$f(\tilde{x}) = f(x + \epsilon x^{\mathsf{d}}) = f(x) + \epsilon f'(x) x^{\mathsf{d}} . \tag{4.91}$$

The relation here above also shows that a function f can be applied at a dual variable $\tilde{x}$ only if the real part x is in the domain of f. As an example, the inverse of $\tilde{x}$ results

$$\tilde{x}^{-1} = \frac{1}{x}\left(1 - \epsilon \frac{x^{\mathsf{d}}}{x}\right) , \tag{4.92}$$

which is well-defined for any $x \neq 0$.

Hence, the inverse of a dual number in the form $\tilde{x} = \epsilon x^{\mathsf{d}}$ is not defined, even though ϵx^{d} is non-null.

By using (4.92), the division between two dual numbers reads

$$\frac{\tilde{a}}{\tilde{b}} = \frac{a + \epsilon a^{\mathsf{d}}}{b + \epsilon b^{\mathsf{d}}} = \frac{a}{b} + \epsilon \left(\frac{a^{\mathsf{d}}}{b} - \frac{a b^{\mathsf{d}}}{b^2} \right) , \tag{4.93}$$

which applies only when $b \neq 0$.

Actually, pure dual numbers, i.e. dual numbers with vanishing real part, are zero divisors in $\mathbb{D}$, that is $(\epsilon a^{\mathsf{d}})(\epsilon b^{\mathsf{d}}) = 0$ whatever a^{d} and b^{d} are. It is exactly the existence of zero divisors that makes $\mathbb{D}$ a ring rather than a field.

Even if $\mathbb{D}$ is not a field, it is possible to construct arrays whose entries are dual numbers. Then, a *dual vector* is expressed in the form

$$\tilde{\mathbf{v}} = \mathbf{v} + \epsilon \mathbf{v}^{\mathsf{d}} ,$$

where $\mathbf{v}$ and $\mathbf{v}^{\mathsf{d}}$ are the real and dual parts, respectively, of the dual vector.

Moreover, acting on the entries of the arrays, the same operations of standard vectors extend to dual vectors:

$$\begin{aligned}\tilde{c}\tilde{\mathbf{v}} &= c\mathbf{v} + \epsilon(c\mathbf{v}^{\mathsf{d}} + c^{\mathsf{d}}\mathbf{v})\,,\\ \tilde{\mathbf{u}}\cdot\tilde{\mathbf{v}} &= \mathbf{u}\cdot\mathbf{v} + \epsilon(\mathbf{u}\cdot\mathbf{v}^{\mathsf{d}} + \mathbf{u}^{\mathsf{d}}\cdot\mathbf{v})\,,\\ \tilde{\mathbf{u}}\times\tilde{\mathbf{v}} &= \mathbf{u}\times\mathbf{v} + \epsilon(\mathbf{u}\times\mathbf{v}^{\mathsf{d}} + \mathbf{u}^{\mathsf{d}}\times\mathbf{v})\,,\end{aligned} \tag{4.94}$$

where the last one specifically applies to three-dimensional dual vectors.

In addition, by (4.1), and applying the property (4.91) to the square root function, the Euclidean norm of a dual vector reads

$$\|\tilde{\mathbf{v}}\| = \|\mathbf{v}\| + \epsilon\frac{\mathbf{v}\cdot\mathbf{v}^{\mathsf{d}}}{\|\mathbf{v}\|}\,. \tag{4.95}$$

Similarly, a *dual matrix* is a matrix $\tilde{\mathbf{A}}$ whose entries are dual numbers so that, gathering the real and dual components in $\mathbf{A}$ and $\mathbf{A}^{\mathsf{d}}$, respectively, it can be expressed in the form

$$\tilde{\mathbf{A}} = \mathbf{A} + \epsilon\mathbf{A}^{\mathsf{d}}\,.$$

Given the dual matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$, whose dimensions are compatible, the matrix multiplication reads

$$\tilde{\mathbf{A}}\tilde{\mathbf{B}} = (\mathbf{A} + \epsilon\mathbf{A}^{\mathsf{d}})(\mathbf{B} + \epsilon\mathbf{B}^{\mathsf{d}}) = \mathbf{A}\mathbf{B} + \epsilon(\mathbf{A}\mathbf{B}^{\mathsf{d}} + \mathbf{A}^{\mathsf{d}}\mathbf{B})\,, \tag{4.96}$$

whence it is clear that the pure real matrix $\mathbf{I}$ is the identity matrix also in the context of the dual numbers, meaning that it is the only matrix satisfying the property $\tilde{\mathbf{A}}\mathbf{I} = \mathbf{I}\tilde{\mathbf{A}} = \tilde{\mathbf{A}}$ for any dual matrix $\tilde{\mathbf{A}}$.

The transpose of $\tilde{\mathbf{A}}$ is given by transposing the real and the dual part, that is

$$\tilde{\mathbf{A}}^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} + \epsilon{\mathbf{A}^{\mathsf{d}}}^{\mathsf{T}}\,, \tag{4.97}$$

and, when $\tilde{\mathbf{A}}$ is a square matrix, the dual inverse is

$$\tilde{\mathbf{A}}^{-1} = \mathbf{A}^{-1} - \epsilon\mathbf{A}^{-1}\mathbf{A}^{\mathsf{d}}\mathbf{A}^{-1}\,. \tag{4.98}$$

The previous relation, that is well-defined only when the real part $\mathbf{A}$ is not singular, can be derived by considering the condition $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^{-1} = \mathbf{I}$.

Finally, the determinant of a dual matrix $\tilde{\mathbf{A}}$ is a dual number to be evaluated as (see, e.g., Angeles (1998))

$$\det(\tilde{\mathbf{A}}) = \det(\mathbf{A})\big(1 + \epsilon \,\mathrm{tr}(\mathbf{A}^{\mathsf{d}}\mathbf{A}^{-1})\big)\,. \tag{4.99}$$

4.5.2 Representing Motions as Dual Rotations

The formalism of dual numbers allows one to represent a rigid motion considering only the algebra of rotations, to be intended in a dual representation sense.

Specifically, a dual orthogonal matrix $\tilde{\mathbf{R}}$ satisfies the conditions $\tilde{\mathbf{R}}^{\mathsf{T}}\tilde{\mathbf{R}} = \mathbf{I}$ and $\det(\tilde{\mathbf{R}}) = 1$.

The first requirement explicitly reads

$$(\mathbf{R} + \epsilon\mathbf{R}^{\mathsf{d}})(\mathbf{R}^{\mathsf{T}} + \epsilon{\mathbf{R}^{\mathsf{d}}}^{\mathsf{T}}) = \mathbf{R}\mathbf{R}^{\mathsf{T}} + \epsilon(\mathbf{R}{\mathbf{R}^{\mathsf{d}}}^{\mathsf{T}} + \mathbf{R}^{\mathsf{d}}\mathbf{R}^{\mathsf{T}}) = \mathbf{I}\,,$$

whence the real part satisfies the condition $\mathbf{R}\mathbf{R}^{\mathsf{T}} = \mathbf{I}$, while the vanishing of the dual part implies the skew-symmetry of $\mathbf{R}^{\mathsf{d}}\mathbf{R}^{\mathsf{T}}$. Consequently, $\mathbf{R}^{\mathsf{d}}$ can be decomposed in the form

$$\mathbf{R}^{\mathsf{d}} = \mathbf{D}\mathbf{R}\,, \tag{4.100}$$

where $\mathbf{D} = \mathbf{R}^{\mathsf{d}}\mathbf{R}^{\mathsf{T}}$ is a skew-symmetric matrix.

Moreover, by (4.99), the dual determinant results

$$\det(\tilde{\mathbf{R}}) = \det(\mathbf{R})\big(1 + \epsilon\,\mathrm{tr}(\mathbf{R}^{\mathsf{d}}\mathbf{R}^{-1})\big) = \det(\mathbf{R}) = 1\,,$$

where the property $\mathrm{tr}(\mathbf{R}^{\mathsf{d}}\mathbf{R}^{-1}) = \mathrm{tr}(\mathbf{R}^{\mathsf{d}}\mathbf{R}^{\mathsf{T}}) = 0$, following from the skew-symmetry of $\mathbf{R}^{\mathsf{d}}\mathbf{R}^{\mathsf{T}}$, has been used.

In conclusion, a dual orthogonal matrix $\tilde{\mathbf{R}}$ is such that the real part $\mathbf{R}$ is orthogonal and the dual part $\mathbf{R}^{\mathsf{d}}$ can be decomposed as the product of a skew-symmetric matrix $\mathbf{D}$ and the real orthogonal matrix $\mathbf{R}$:

$$\tilde{\mathbf{R}} = \mathbf{R} + \epsilon\mathbf{D}\mathbf{R} = (\mathbf{I} + \epsilon\mathbf{D})\mathbf{R}\,. \tag{4.101}$$

The space of the matrices $\tilde{\mathbf{R}}$ is called the *dual orthogonal group* and is denoted as $SO(3, \mathbb{D})$. It is possible to prove that there exists a Lie group homomorphism between $SO(3,\mathbb{D})$ and $SE(3)$, as shown by Daher (2013).

Here, it is just observed that the relation between a dual orthogonal matrix $\tilde{\mathbf{R}}$ and a homogeneous matrix $\mathbf{H}$ naturally prompts by comparing (4.101) with the adjoint representation given by (4.84):

$$\mathbf{H} = \begin{bmatrix} \mathbf{R} & \mathbf{v} \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix} \cong \mathrm{Ad}_{\mathbf{H}} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \hat{\mathbf{v}}\mathbf{R} & \mathbf{R} \end{bmatrix} \cong \tilde{\mathbf{R}} = \mathbf{R} + \epsilon \hat{\mathbf{v}}\mathbf{R}\,. \tag{4.102}$$

The identification $SE(3) \cong SO(3,\mathbb{D})$ summarizes the *Principle of transference*, claiming that any valid proposition about the Lie group $SO(3,\mathbb{R})$ and its Lie algebra becomes on dualisation a valid statement about $SO(3,\mathbb{D})$ and its Lie algebra and hence about $SE(3)$ and its Lie algebra (Daher, 2013). A historical review of the principle, along with a discussion about its algebraic fundamentals and validity, can be found, e.g., in Chevallier (1996), Martínez and Duffy (1993), Selig (2005).

The benefit of dual number formalism is that the operators characterizing the Lie group $SE(3)$ can be obtained from the relevant operator of $SO(3)$ applied in a dual sense.

Specifically, a screw $(\boldsymbol{\theta}, \mathbf{u})$ can be represented by means of a dual rotation vector whose real part is $\boldsymbol{\theta}$ and the dual part is the component $\mathbf{u}$:

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + \epsilon \mathbf{u}\,. \tag{4.103}$$

By applying (4.95), the dual norm of $\tilde{\boldsymbol{\theta}}$ results:

$$\|\tilde{\boldsymbol{\theta}}\| = \tilde{\theta} = \theta + \epsilon d\,, \tag{4.104}$$

where $\theta = \|\boldsymbol{\theta}\|$ is the angle of rotation and the dual part $d = \mathbf{k} \cdot \mathbf{u}$ is the contribution of the translation along the axis $\mathbf{k} = \boldsymbol{\theta}/\theta$, just as introduced by (4.73).

Moreover, using (4.93) for the component-wise division of $\tilde{\boldsymbol{\theta}}$ by $\|\tilde{\boldsymbol{\theta}}\|$, the dual unit vector results

$$\tilde{\mathbf{k}} = \frac{\tilde{\boldsymbol{\theta}}}{\tilde{\theta}} = \frac{\boldsymbol{\theta}}{\theta} + \epsilon\left(\frac{\mathbf{u}}{\theta} - \frac{d\boldsymbol{\theta}}{\theta^2}\right) = \mathbf{k} + \epsilon\frac{\mathbf{u} - d\mathbf{k}}{\theta}\,,$$

and, comparing with (4.74), the dual part is the moment $\mathbf{p} \times \mathbf{k}$ of the unit vector $\mathbf{k}$ identifying the screw axis:

$$\tilde{\mathbf{k}} = \frac{\tilde{\boldsymbol{\theta}}}{\tilde{\theta}} = \mathbf{k} + \epsilon \mathbf{p} \times \mathbf{k} \,. \tag{4.105}$$

Hence, the dual rotation vector can be decomposed as

$$\tilde{\boldsymbol{\theta}} = \tilde{\theta} \tilde{\mathbf{k}} \,, \tag{4.106}$$

in such a way the dual unit vector $\tilde{\mathbf{k}}$ completely identifies the screw, both in terms of the direction of the axis, by means of $\mathbf{k}$, and its position, through the moment $\mathbf{p} \times \mathbf{k}$; at the same time, the dual angle $\tilde{\theta}$ gives the magnitude of the screw motion, defining the angle of rotation θ and the translation d along the screw axis, considered in its actual position.

Exploiting the decomposition (4.106), the Rodrigues' rotation formula (4.45) can be applied in its dual version to provide a dual orthogonal matrix:

$$\tilde{\mathbf{R}}_{\tilde{\boldsymbol{\theta}}} = \mathbf{I} + \sin \tilde{\theta} \, \widehat{\tilde{\mathbf{k}}} + (1 - \cos \tilde{\theta}) \widehat{\tilde{\mathbf{k}}}^2 \,. \tag{4.107}$$

where $\widehat{\tilde{\mathbf{k}}}$ is the dual skew-symmetric matrix relevant to the dual unit vector $\tilde{\mathbf{k}}$ and the dual trigonometric functions $\sin \tilde{\theta}$ and $\cos \tilde{\theta}$ can be evaluated recalling the property (4.91):

$$\sin \tilde{\theta} = \sin \theta + \epsilon d \cos \theta \,, \tag{4.108}$$

as well as

$$\cos \tilde{\theta} = \cos \theta - \epsilon d \sin \theta \,. \tag{4.109}$$

The same formalism can be exploited to evaluate the tangent operator of $SE(3)$ and its inverse as the dual extension of (4.58) and (4.61):

$$\tilde{\mathbf{T}}_{\tilde{\boldsymbol{\theta}}} = \mathbf{I} - \frac{1 - \cos \tilde{\theta}}{\tilde{\theta}} \widehat{\tilde{\mathbf{k}}} + \left(1 - \frac{\sin \tilde{\theta}}{\tilde{\theta}}\right) \widehat{\tilde{\mathbf{k}}}^2 \,, \tag{4.110}$$

and

$$\tilde{\mathbf{T}}_{\tilde{\boldsymbol{\theta}}}^{-1} = \mathbf{I} + \frac{1}{2}\widehat{\tilde{\mathbf{k}}} + \left(1 - \frac{\tilde{\theta}}{2}\cot\frac{\tilde{\theta}}{2}\right)\widehat{\tilde{\mathbf{k}}}^2 , \tag{4.111}$$

where, using again (4.91), the dual evaluation of the cotangent function reads

$$\cot\frac{\tilde{\theta}}{2} = \cot\frac{\theta}{2} + \epsilon\frac{d}{4}\left(1 + \cot^2\frac{\theta}{2}\right). \tag{4.112}$$

4.6 Some Identities for Operators of $SO(3)$ and $SE(3)$

Given a rotation vector $\boldsymbol{\theta} \in \mathbb{R}^3$, let $\widehat{\boldsymbol{\theta}} \in \mathfrak{so}(3)$ be the relevant skew-symmetric matrix. In addition, consider the rotation matrix $\mathbf{R}_{\boldsymbol{\theta}}$, provided by (4.50), and the tangent operator $\mathbf{T}_{\boldsymbol{\theta}}$ expressed by (4.58).

Please, first observe that $\mathbf{R}_{-\boldsymbol{\theta}} = \mathbf{R}_{\boldsymbol{\theta}}^{-1} = \mathbf{R}_{\boldsymbol{\theta}}^{\mathsf{T}}$, as well as $\mathbf{T}_{-\boldsymbol{\theta}} = \mathbf{T}_{\boldsymbol{\theta}}^{\mathsf{T}}$, and notice that $\widehat{\boldsymbol{\theta}}$ commutes with both $\mathbf{R}_{\boldsymbol{\theta}}$ and $\mathbf{T}_{\boldsymbol{\theta}}$.

Then, setting $\theta = \|\boldsymbol{\theta}\|$, $\mathbf{k} = \boldsymbol{\theta}/\theta$ and multiplying $\mathbf{T}_{\boldsymbol{\theta}}$ with $\widehat{\boldsymbol{\theta}}$, one has:

$$\begin{aligned}
\mathbf{T}_{\boldsymbol{\theta}}\widehat{\boldsymbol{\theta}} &= \theta\widehat{\mathbf{k}} - (1-\cos\theta)\widehat{\mathbf{k}}^2 + \theta\left(1 - \frac{\sin\theta}{\theta}\right)\widehat{\mathbf{k}}^3 \\
&= \theta\widehat{\mathbf{k}} - (1-\cos\theta)\widehat{\mathbf{k}}^2 - \theta\widehat{\mathbf{k}} + \sin\theta\,\widehat{\mathbf{k}} \\
&= \mathbf{I} - \big(\mathbf{I} - \sin\theta\,\widehat{\mathbf{k}} + (1-\cos\theta)\widehat{\mathbf{k}}^2\big)
\end{aligned}$$

that is

$$\mathbf{T}_{\boldsymbol{\theta}}\widehat{\boldsymbol{\theta}} = \mathbf{I} - \mathbf{R}_{\boldsymbol{\theta}}^{\mathsf{T}} , \qquad \mathbf{R}_{\boldsymbol{\theta}} = \mathbf{I} + \mathbf{T}_{\boldsymbol{\theta}}^{\mathsf{T}}\widehat{\boldsymbol{\theta}} . \tag{4.113}$$

By using the series expansions (4.51) for $\mathbf{R}_{\boldsymbol{\theta}}$ and (4.54) for $\mathbf{T}_{\boldsymbol{\theta}}$, the following identities can be verified:

$$\mathbf{T}_{\boldsymbol{\theta}}\mathbf{R}_{\boldsymbol{\theta}} = \mathbf{R}_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}} = \mathbf{T}_{\boldsymbol{\theta}}^{\mathsf{T}} , \qquad \mathbf{T}_{\boldsymbol{\theta}}^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{\mathsf{T}} = \mathbf{R}_{\boldsymbol{\theta}}^{\mathsf{T}}\mathbf{T}_{\boldsymbol{\theta}}^{-1} = \mathbf{T}_{\boldsymbol{\theta}}^{-\mathsf{T}} . \tag{4.114}$$

Using the dual rotation vector $\tilde{\boldsymbol{\theta}}$, defined by (4.103), to represent a screw motion, the identities here above also apply to the same operators in their dual extension, given by (4.107) and (4.110-4.111).

Hence, recalling the isomorphism (4.102), the identities (4.114) likewise hold for the relevant operators of $SE(3)$ in the adjoint representation form:

$$\begin{aligned} &\mathbf{T}_{\mathbf{h}}\,\mathrm{Ad}_{\exp(\widehat{\mathbf{h}})} = \mathrm{Ad}_{\exp(\widehat{\mathbf{h}})}\,\mathbf{T}_{\mathbf{h}} = \mathbf{T}_{-\mathbf{h}}\,, \\ &\mathbf{T}_{\mathbf{h}}^{-1}\,\mathrm{Ad}_{\exp(-\widehat{\mathbf{h}})} = \mathrm{Ad}_{\exp(-\widehat{\mathbf{h}})}\,\mathbf{T}_{\mathbf{h}}^{-1} = \mathbf{T}_{-\mathbf{h}}^{-1}\,, \end{aligned} \tag{4.115}$$

where $\mathbf{T}_{\mathbf{h}}$ is the tangent operator of $SE(3)$ defined by (4.90) and $\mathrm{Ad}_{\exp(\widehat{\mathbf{h}})}$ is the adjoint representation of $\exp(\widehat{\mathbf{h}})$ given by (4.84).

Chapter 5

A Geometrically Exact Beam Model

The intuitive notion of a *beam* is related to a solid body such that a characteristic direction, coincident with the direction of the largest expansion of the solid, can be detected.

Aiming to derive a beam model as a specialization of the kinematics of a three-dimensional continuous body, by considering specific hypotheses justified by its peculiar shape, some general notion about continuum mechanics are briefly recalled.

5.1 Body, Space and Motion

Definition 5.1. A *body* $\mathscr{B}$ is a smooth manifold with boundary. A point $\mathscr{p}$ of $\mathscr{B}$ is called a *material point* of the body.

For the purposes of this work it is assumed that the body is a three-dimensional manifold, so that any material point is identified by the coordinates $\psi(\mathscr{p}) = \bar{\mathbf{x}} \in \mathbb{R}^3$. Moreover, the map $\psi\colon \mathscr{p} \mapsto \bar{\mathbf{x}}$ is the *body chart* and $\psi(\mathscr{p})$ are the *material coordinates.*

A further notion to be introduced along with a body is the physical space.

Definition 5.2. The *physical space* is a Euclidean space $\mathscr{E}$ and a point P of the physical space is called a *spatial point* of $\mathscr{E}$.

In the context of the present work, the physical space $\mathscr{E}$ is specifically intended as a 3-dimensional Euclidean affine space.

Moreover, since a Euclidean space is endowed with the Euclidean distance, the physical space $\mathscr{E}$ is also a metric space and, as such, a topological space. Also, intending the coordinates of any spatial point P as smooth functions, the physical space is also a smooth manifold.

With these premises, we say that a point $\mathscr{p}$ of the body $\mathscr{B}$ occupies a position, identified by a spatial point P of the physical space $\mathscr{E}$, if there exists a smooth mapping in the form $\mathscr{p} \mapsto \mathsf{P}$. By exploiting such an association, the notion of the configuration of a body is introduced.

Definition 5.3. Given an interval $\mathscr{J} \subset \mathbb{R}$ and a body $\mathscr{B}$, a *configuration* is a smooth map $\mathscr{C} : \mathscr{J} \times \mathscr{B} \to \mathscr{E}$ associating a material point $\mathscr{p}$ of the body $\mathscr{B}$, at a time t, with a spatial point P of the physical space $\mathscr{E}$:

$$\begin{aligned} \mathscr{C} :\ & \mathscr{J} \times \mathscr{B} \to \mathscr{E} \\ & (t, \mathscr{p}) \mapsto \mathsf{P}\,. \end{aligned} \tag{5.1}$$

With a slight abuse of terminology, the spatial point P, as image of the material point $\mathscr{p}$ through the map $\mathscr{C}$, is itself referred to as the *configuration* of $\mathscr{p}$ at the time t. Similarly, the subset $\mathscr{B}_t = \mathscr{C}(t, \mathscr{B}) \subset \mathscr{E}$ is also called the configuration of the body $\mathscr{B}$ at t.

The restriction of $\mathscr{C}$ at a specific time is the map $\mathscr{C}_t : \mathscr{B} \to \mathscr{E}$ defined by $\mathscr{C}_t(\mathscr{p}) = \mathscr{C}(t, \mathscr{p})$.

It is required that, for any $t \in \mathscr{J}$, the induced map $\mathscr{C}_t : \mathscr{B} \to \mathscr{E}$ is an embedding between the body manifold and the physical space. Actually, the injectivity of $\mathscr{C}_t$, as well as of its differential, is consistent with the two classical principles of *permanence of matter* and *impenetrability*.

Since $\mathscr{B}_t$, as a subset of $\mathscr{E}$, is itself a smooth manifold, any spatial point $\mathsf{P} \in \mathscr{B}_t$ is identified by the coordinates $\bar{\mathbf{p}} = \varphi_t(\mathsf{P}) \in \mathbb{R}^3$, which are the affine coordinates with respect to some affine frame $\mathscr{F}$ (see Figure 5.1). Hence, the configuration of a material point $\mathscr{p}$ is represented in coordinates through

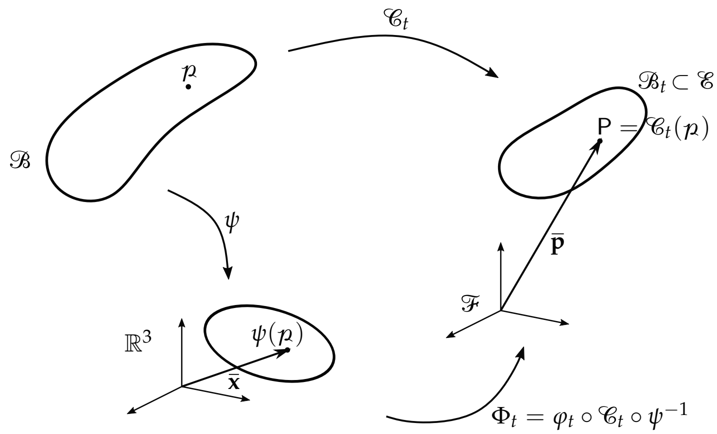


Figure 5.1: Material and spatial representation of the body $\mathscr{B}$.

the real-valued function $\Phi_t = \varphi_t \circ \mathscr{C}_t \circ \psi^{-1}$:

$$\begin{aligned} \Phi_t :\ \psi(\mathscr{B}) &\to \varphi_t(\mathscr{B}_t) \\ \bar{\mathbf{x}} &\mapsto \bar{\mathbf{p}}\,, \end{aligned} \tag{5.2}$$

where $\psi(\mathscr{B})$ and $\varphi_t(\mathscr{B}_t)$ are both subsets of $\mathbb{R}^3$.

Remark. The characterization of the physical space as an affine space is consistent with the intuitive idea of 'space' as it is commonly experienced. In short, one can say that the physical space $\mathscr{E}$ in Definition 5.2 is a mathematical model of the real environment.

On the other hand, for what concerns the idea of 'body', the notion of a manifold $\mathscr{B}$ introduced by Definition 5.1 seems to be far from the common experience of an object occupying a region of the environment.

Actually, defining a body as a manifold allows one to characterize a material point $\mathscr{p}$ independently of its position in the space. As a consequence, once the body chart ψ has been established, a material point of the body is uniquely identified by a set of material coordinates $\bar{\mathbf{x}} \in \mathbb{R}^3$, whatever is its configuration in the physical space.

At the same time, one of the drawbacks of such a theoretical notion concerns the geometric characterization of a body, which is necessarily involved when one thinks of an object in the real world. In fact, unless the manifold is endowed with some additional metric structure, the usual concepts of length and angles do not apply to $\mathscr{B}$. Consequently, if one just refers to the manifold $\mathscr{B}$, the very idea of 'shape' of the body is not defined.

As a matter of fact, the information about the geometry of a body are derived from its configuration in the physical space. In practice, the configuration map provides a connection between the mathematical definition of a body, i.e. a finite-dimensional manifold, and the idea of 'body' in the common thinking.

Moreover, the requirement for the configuration map $\mathscr{C}_t$ to be an embedding at any t, assures that the relation between the body $\mathscr{B}$ and the subset $\mathscr{B}_t \subset \mathscr{E}$ is one-to-one. Then, in specifying any geometric feature of the body, we actually refer to some configuration $\mathscr{B}_t$, that inherits the geometric properties of the Euclidean space $\mathscr{E}$.

5.2 Beam Geometric Characterization

Following Antman (2005), we consider a *beam* as a slender body such that a one characteristic direction can be detected.

It is again emphasized that the geometric characterization of a body only makes sense when one refers to a configuration $\mathscr{B}_t$ in the physical space. Then, the slenderness of $\mathscr{B}$ should actually be intended as a geometric feature of any configuration $\mathscr{B}_t = \mathscr{C}_t(\mathscr{B})$.

5.2.1 Beam Material Coordinates

The hypothesis about the shape of $\mathscr{B}_t$ translates into the possibility to introduce a scalar μ such that, for any fixed $t \in J$, distinct spatial points P are considered along the characteristic direction of $\mathscr{B}_t$ as μ varies in a closed interval $[a, b] \subset \mathbb{R}$.

At the same time, the coordinate representation of any point $\mathsf{P} \in \mathscr{B}_t$ is the image of a triple $\bar{\mathbf{x}}$ through the real-valued function Φ_t given by (5.2), where $\bar{\mathbf{x}} = (x^1, x^2, x^3)$ is the *material representation*, or the local representation, of a point $\mathscr{p}$.

Then, in defining the body chart $\psi\colon \mathscr{B} \to \mathbb{R}^3$, it is advisable to consider one of the material coordinates, say $x^3 \in \mathbb{R}$, to be coincident with the scalar parameter $\mu \in [a, b]$.

Formally, the scalar μ, which is called the *material abscissa* of the beam, is given by a map $\psi_a\colon \mathscr{B} \to \mathbb{R}$ resulting from the composition $\psi_a = \pi_3 \circ \psi$, where $\pi_3\colon \mathbb{R}^3 \to \mathbb{R}$ is the projection map defined by $\pi_3(x^1, x^2, x^3) = x^3$.

Explicitly, ψ_a reads

$$\begin{aligned} \psi_a : \mathscr{B} &\to \mathbb{R} \\ p &\mapsto \mu\,, \end{aligned}$$

where the image $\psi_a(\mathscr{B})$ is a closed real interval $[a, b] \subset \mathbb{R}$.

Moreover, for the arbitrariness of the material representation of $\mathscr{B}$, it is convenient to set $a = 0$ and $b = 1$, so that μ actually results a variable ranging in $[0, 1] \subset \mathbb{R}$.

By exploiting the material abscissa map, it is also defined the *cross-section* $\mathcal{S}_\mu$ of the beam at μ as the preimage of $\mu \in [0, 1]$ under ψ_a:

$$\mathcal{S}_\mu = \psi_a^{-1}(\mu) = \{\, p \in \mathscr{B} \mid \psi_a(p) = \mu \,\}\,.$$

It is trivial to see that the material coordinates of any point p of the cross-section $\mathcal{S}_\mu$ are in the form $\psi(p) = (x^1, x^2, \mu)$. Then, for each fixed value of the abscissa μ, the coordinates x^1 and x^2 can themselves be gathered in the pair $(x^1, x^2) \in \mathbb{R}^2$, inducing a map $\psi_{\mathcal{S}}|_\mu\colon \mathscr{B} \to \mathbb{R}^2$ defined by $\psi_{\mathcal{S}}|_\mu(p) = (x^1, x^2)$.

Similarly to the abscissa map, the material representation map $\psi_{\mathcal{S}}|_\mu$ of the cross-section at μ results from the composition $\pi_{12} \circ \psi$, where π_{12} is the projection map from $\mathbb{R}^3$ to $\mathbb{R}^2$ defined as $\pi_{12}(x^1, x^2, x^3) = (x^1, x^2)$.

It is worth noting that, by construction, both $\psi_{\mathcal{S}}|_\mu$ and ψ_a are smooth functions. In fact, being $\psi_{\mathcal{S}}|_\mu = \pi_{12} \circ \psi$, as well as $\psi_a = \pi_3 \circ \psi$, the smoothness of the body chart ψ is directly extended to the cross-section map $\psi_{\mathcal{S}}|_\mu$ and the abscissa map ψ_a.

Since each cross-section $\mathcal{S}_\mu$ is a subset of points which have the same material abscissa μ, the beam $\mathscr{B}$ can be thought as the union of all the

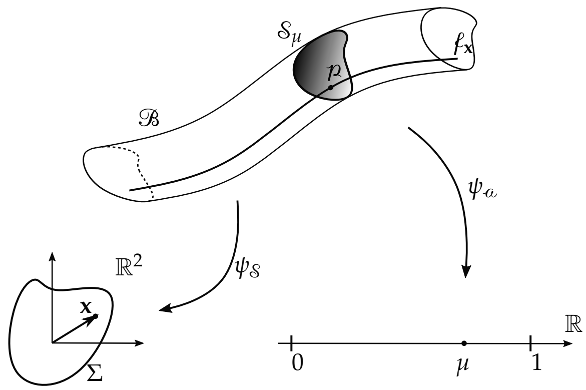


Figure 5.2: Material representation of the beam $\mathcal{B}$.

cross-sections as μ varies in $[0,1]$:

$$\mathcal{B} = \bigcup_{\mu \in [0,1]} \mathcal{S}_\mu \,. \tag{5.3}$$

At the same time, the image of the cross-section $\mathcal{S}_\mu$ is a two-dimensional connected subset of $\mathbb{R}^2$, denoted as $\Sigma_\mu = \psi_\mathcal{S}|_\mu(\mathcal{S}_\mu) \subset \mathbb{R}^2$. Then, the material representation of the beam $\mathcal{B}$ can be expressed as

$$\psi(\mathcal{B}) = \bigcup_{\mu \in [0,1]} \Sigma_\mu \times \{\mu\} \subset \mathbb{R}^3 \,.$$

If the representation map of the cross-section does not depend on the material abscissa μ, it is simply denoted as $\psi_\mathcal{S}$. In such a case, all the cross-sections $\mathcal{S}_\mu$ do have the same material representation $\Sigma = \psi_\mathcal{S}(\mathcal{S}_\mu)$ and the beam is said to have a *uniform* cross-section.

Consequently, as shown in Figure 5.2, the material representation of the beam $\mathcal{B}$ can be expressed as the Cartesian product of the cross-section Σ and the domain $[0,1]$ of the material abscissa:

$$\psi(\mathcal{B}) = \Sigma \times [0,1] \subset \mathbb{R}^3 \,.$$

Definition 5.4. Let $\mathcal{B}$ be a beam having a uniform cross-section, whose material representation is $\Sigma \subset \mathbb{R}^2$. For any $\mathbf{x} \in \Sigma$, the *(material) fiber* of the beam over $\mathbf{x}$, denoted as $f_{\mathbf{x}}$, is the preimage of $\mathbf{x}$ through the representation map of the cross-section:

$$f_{\mathbf{x}} = \psi_{\mathcal{S}}^{-1}(\mathbf{x}) = \{\, p \in \mathcal{B} \mid \psi_{\mathcal{S}}(p) = \mathbf{x} \,\}\,.$$

The set of all the material fibers of the beam $\mathcal{B}$ is denoted as

$$\mathcal{F}_{\Sigma} = \{\, f_{\mathbf{x}} = \psi_{\mathcal{S}}^{-1}(\mathbf{x}) \mid \mathbf{x} \in \Sigma \,\}\,, \tag{5.4}$$

and, since any material fiber $f_{\mathbf{x}}$ is uniquely identified by $\mathbf{x} \in \Sigma$, the beam $\mathcal{B}$ can be intended as the union of all its fibers:

$$\mathcal{B} = \bigcup_{\mathbf{x}\in\Sigma} f_{\mathbf{x}}\,. \tag{5.5}$$

Moreover, the fiber over the null element $\mathbf{o} = (0,0)$ is the *axis* a of the beam:

$$a = f_{\mathbf{o}} = \{\, p \in \mathcal{B} \mid \psi_{\mathcal{S}}(p) = (0,0) \,\}\,. \tag{5.6}$$

Please notice that, according to Definition 5.4, the axis of the beam does exist only if the null element $\mathbf{o} \in \mathbb{R}^2$ is in the material representation Σ of the cross-section. To this aim, it is useful to choose the local representation map ψ in such a way that Σ contains the null element of $\mathbb{R}^2$.

In any case, it will be shown that, with some additional assumptions that are common in beam theory, any $\mathbf{x} \in \mathbb{R}^2$ can be mapped to some spatial point of $\mathcal{E}$, even if the material fiber of $\mathbf{x}$ is not defined.

It is also remarked that, being $\psi_{\mathcal{S}} = \pi_{12} \circ \psi$, the material coordinates of any point p of the fiber over $\mathbf{x} = (x^1, x^2)$ are in the form

$$\psi(p) = (x^1, x^2, \mu)\,, \ \ \forall\, p \in f_{\mathbf{x}}\,.$$

Conversely, the bijectivity of the map ψ assures that $p = \psi^{-1}(x^1, x^2, \mu)$ is a point of the fiber $f_{\mathbf{x}}$.

Then, by fixing x^1 and x^2, the map ψ^{-1} induces a one-variable map $\gamma_{\mathbf{x}}$ between the interval $[0,1]$ of the material abscissa μ and the fiber $f_{\mathbf{x}}$ of the beam:

$$\begin{aligned} \gamma_{\mathbf{x}} :\ & [0,1] \to \mathscr{f}_{\mathbf{x}} \subset \mathscr{B} \\ & \mu \mapsto \mathscr{p} = \psi^{-1}(x^1, x^2, \mu)\,, \end{aligned} \tag{5.7}$$

where the smoothness of the coordinate map ψ implies that $\gamma_{\mathbf{x}}$ is a smooth curve on the manifold $\mathscr{B}$.

5.2.2 Beam Configuration in the Physical Space

The relation reported in (5.5) shows how the beam $\mathscr{B}$ can be thought as the union of its material fibers. Then, the configuration of the beam at t results

$$\mathscr{B}_t = \mathscr{C}_t(\mathscr{B}) = \bigcup_{\mathbf{x}\in\Sigma} \mathscr{C}_t(\mathscr{f}_{\mathbf{x}})\,, \tag{5.8}$$

where $\mathscr{C}_t(\mathscr{f}_{\mathbf{x}}) \subset \mathscr{E}$ is the configuration of the material fiber $\mathscr{f}_{\mathbf{x}}$ in the physical space $\mathscr{E}$.

Since each material point $\mathscr{p}$ of the fiber $\mathscr{f}_{\mathbf{x}}$ is the image of a scalar μ through the map $\gamma_{\mathbf{x}}$ defined by (5.7), the composition $\Gamma_{\mathbf{x}} = \mathscr{C}_t \circ \gamma_{\mathbf{x}}$ is a curve of the physical space:

$$\begin{aligned} \Gamma_{\mathbf{x}} :\ & [0,1] \to \mathscr{C}_t(\mathscr{f}_{\mathbf{x}}) \\ & \mu \mapsto \mathsf{P} = \Gamma_{\mathbf{x}}(\mu)\,. \end{aligned} \tag{5.9}$$

Using the terminology of Section 2, the tangent vector $\mathrm{d}/\mathrm{d}\mu|_{\mu}$ is mapped by the differential $\mathrm{d}\Gamma_{\mathbf{x}}|_{\mu}$ to $\Gamma'_{\mathbf{x}}(\mu)$, which is the tangent vector to the curve at the point $\mathsf{P} = \Gamma_{\mathbf{x}}(\mu)$.

Moreover, by Property 2 in Proposition 2.10, the differential of the curve $\Gamma_{\mathbf{x}}$ at μ is

$$\mathrm{d}\Gamma_{\mathbf{x}}|_{\mu} = \mathrm{d}(\mathscr{C}_t \circ \gamma_{\mathbf{x}})_{\mu} = \mathrm{d}\mathscr{C}_t|_{\mathscr{p}} \circ \gamma_{\mathbf{x}}|_{\mu}\,,$$

whence, recalling that $\gamma_{\mathbf{x}}$ is a diffeomorphism between $[0,1]$ and $\mathscr{f}_{\mathbf{x}}$, and that $\mathscr{C}_t$ is an embedding, $\mathrm{d}\Gamma_{\mathbf{x}}|_{\mu}$ is injective.

Consequently, the tangent vector $\Gamma'_{\mathbf{x}}(\mu)$, as image of the differential $\mathrm{d}\Gamma_{\mathbf{x}}|_{\mu}$, is non-vanishing for any μ in the interval $[0,1]$, and the curve $\Gamma_{\mathbf{x}}$

is *regular* (see, e.g., do Carmo (2016)).

Please observe that, by virtue of the decomposition (5.8), the configuration $\mathscr{B}_t$ of the beam is fully defined once the curves $\Gamma_{\mathbf{x}}\colon [0,1] \to \mathscr{C}_t(\mathscr{f}_{\mathbf{x}})$ have been characterized as $\mathbf{x}$ varies in Σ.

Hence, in order to formulate a one-dimensional model of the beam, it is desirable that one of the curves $\Gamma_{\mathbf{x}}$ could represent the beam as a whole, provided that the configuration of any other fiber is accordingly derived.

To this purpose, the beam axis $\mathscr{a}$ is assumed to be the representative fiber of the beam (see Figure 5.3), and its configuration in the physical space is defined by the curve $\Gamma_{\mathbf{o}}\colon [0,1] \to \mathscr{C}_t(\mathscr{a})$.

Referring again to do Carmo (2016), the *arc length* of the curve $\Gamma_{\mathbf{o}}$ is defined as the map $s\colon [0,1] \to \mathbb{R}$ given by

$$s(\mu) = \int_0^\mu \|\Gamma'_{\mathbf{o}}(\mu)\| \mathrm{d}\mu \,, \tag{5.10}$$

where $\|\Gamma'_{\mathbf{o}}(\mu)\|$ is the length of the tangent vector $\Gamma'_{\mathbf{o}}(\mu)$.

Since the curve $\Gamma_{\mathbf{o}}$ is regular, the arc length s is a differentiable function and its derivative at μ, i.e. $\mathrm{d}s/\mathrm{d}\mu|_\mu = \|\Gamma'_{\mathbf{o}}(\mu)\|$, induces the following real-valued function:

$$\begin{aligned} \lambda :\ & [0,1] \to \mathbb{R} \\ & \mu \mapsto \lambda(\mu) = \left.\frac{\mathrm{d}s}{\mathrm{d}\mu}\right|_\mu = \|\Gamma'_{\mathbf{o}}(\mu)\| \,, \end{aligned} \tag{5.11}$$

whose meaning is to provide the arc length $\mathrm{d}s = \lambda(\mu)\mathrm{d}\mu$ of the infinitesimal element of $\Gamma_{\mathbf{o}}$ between μ and $\mu + \mathrm{d}\mu$.

From (5.10), the map s clearly vanishes at $\mu = 0$, while its value at $\mu = 1$ is the *length* l of the beam, so that the arc length is properly defined as $s\colon [0,1] \to [0,l]$.

Please notice that, with a slight abuse of notation, the symbol s is typically used to denote both the map from $[0,1]$ to $[0,l]$ and the value it assumes at some μ.

Moreover, since $\Gamma_{\mathbf{o}}$ is aimed to represent the whole beam configuration in the physical space, the value $s = s(\mu)$ of the arc length of $\Gamma_{\mathbf{o}}$ at μ is also

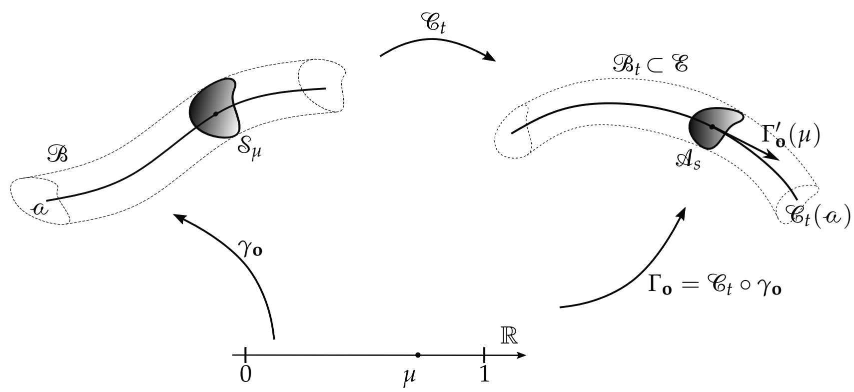


Figure 5.3: Beam material axis a and cross-section $\mathscr{S}_\mu$ and their configuration in the physical space.

called the *spatial abscissa* of the beam, relevant to the configuration $\mathscr{B}_t$.

Remark. One should observe that the composition $\Gamma_{\mathbf{o}} = \mathscr{C}_t \circ \gamma_{\mathbf{o}}$ applies to each $t \in J$. Hence, $\Gamma_{\mathbf{o}}$ is just one within a family of curves of the physical space, each of which represents a separate configuration of the beam axis a. Consequently, the spatial abscissa of the beam should be intended as a function $s = s(t, \mu)$, and the length of the beam is itself the value of the function $t \mapsto l = s(t, 1)$.

However, in order to simplify the notation, unless explicitly specified, time t is assumed to be fixed and the spatial abscissa is considered as $s = s(\mu)$, with μ as the only variable.

Internal Constraints

In addition to the configuration of the beam axis a in the physical space, a further way to look at the configuration $\mathscr{B}_t$ of the beam follows from the decomposition (5.3). Actually, by applying the configuration map at t, one has

$$\mathscr{B}_t = \mathscr{C}_t(\mathscr{B}) = \bigcup_{\mu \in [0,1]} \mathscr{C}_t(\mathscr{S}_\mu) = \bigcup_{s \in [0,l]} \mathscr{A}_s\,, \tag{5.12}$$

where $\mathscr{A}_s = \mathscr{C}_t(\mathcal{S}_\mu) \in \mathscr{E}$ is the cross-section of the beam at s. More precisely, $s = s(\mu)$ is the spatial abscissa of the configuration $\mathscr{B}_t$ relevant to the material abscissa μ, and $\mathscr{A}_s$ is the configuration of the material cross-section $\mathcal{S}_\mu$ in the physical space by $\mathscr{C}_t$.

Consistently with the approach commonly followed in beam theory (Eugster, 2015), it is now introduced the assumption of the beam cross-section to be plane and rigid.

Such hypotheses about the cross-section are usually referred to as *internal constraints*, since they provide a kinematic relation between points within the same cross-section, no matter which is the configuration of the beam in the physical space $\mathscr{E}$.

It is worth emphasizing that, since the manifold $\mathscr{B}$ is not endowed with a metric structure, any geometric assumption about the beam actually refers to its configuration in the physical space. Consequently, the beam cross-section is *plane* if its configuration $\mathscr{A}_s$, for any value of the time t, is fully contained in a plane π_s of the physical space $\mathscr{E}$:

$$\mathscr{A}_s = \mathscr{C}_t(\mathcal{S}_\mu) \subset \pi_s\,, \quad \forall\, \mu \in [0,1]\,,\ t \in J\,. \tag{5.13}$$

Similarly, in order to specify the rigidity of the cross-section, one should exploit the distance function d defined on the physical space $\mathscr{E}$ (cf. Definition 4.11). Then, the beam cross-section is *rigid* if the distance between any pair of points, in their spatial configuration, is constant, whatever the material abscissa μ and the time t are:

$$d\big(\mathscr{C}_t(p), \mathscr{C}_t(q)\big) = \text{const}\,, \quad \text{with } p, q \in \mathcal{S}_\mu\,,\ \forall\, \mu \in [0,1]\,,\ t \in J\,.$$

Moreover, if the points p and q of the cross-section $\mathcal{S}_\mu$ belong to the material fibers f_{x} and f_{y}, respectively, their configuration in $\mathscr{B}_t$ is expressed as $\Gamma_{\mathsf{x}}(\mu)$ and $\Gamma_{\mathsf{y}}(\mu)$, for some $\mu \in [0,1]$. Hence, the rigidity of the cross-section reads

$$d\big(\Gamma_{\mathsf{x}}(\mu), \Gamma_{\mathsf{y}}(\mu)\big) = \text{const}\,,\ \forall\, \mu \in [0,1]\,, \tag{5.14}$$

meaning that the distance between the fibers f_{x} and f_{y}, in any spatial configuration of the beam, is constant. Also, since such a property applies to any pair of material fibers, the relative position of the relevant spatial configurations is itself fixed.

Consequently, the kinematic constraint (5.14) justifies the assumption that the axis $\mathscr{a}$ is representative of the whole beam. Specifically, the axial curve $\Gamma_{\mathbf{o}}$ characterizes the configuration $\mathscr{B}_t$ of the beam at t in the sense that the curve $\Gamma_{\mathbf{x}}$ relevant to any other fiber $\mathscr{f}_{\mathbf{x}}$ can be deduced through the relative position of $\mathscr{C}_t(\mathscr{f}_{\mathbf{x}})$ with respect to $\mathscr{C}_t(\mathscr{a})$, which is independent of the configuration $\mathscr{C}_t$.

For this reason, it is convenient to consider a representation of the material fibers in the physical space that does not depend on t and μ, with the role to provide only the relative position of the beam fibers in their spatial configuration.

To this end, the following map is introduced

$$\begin{aligned} \mathscr{C}_* :\ & \mathcal{F}_\Sigma \to \mathscr{E} \\ & \mathscr{f}_{\mathbf{x}} \mapsto \mathsf{P}_* = \mathscr{C}_*(\mathscr{f}_{\mathbf{x}})\,, \end{aligned} \tag{5.15}$$

where $\mathcal{F}_\Sigma$, defined by (5.4), is the set of all the material fibers of $\mathscr{B}$ and the image $\mathscr{C}_*(\mathcal{F}_\Sigma)$ is a subset $\mathscr{A}_*$ of the physical space $\mathscr{E}$ such that any pair of points $\mathsf{P}_* = \mathscr{C}_*(\mathscr{f}_{\mathbf{x}})$ and $\mathsf{Q}_* = \mathscr{C}_*(\mathscr{f}_{\mathbf{y}})$ satisfies the following condition

$$d(\mathsf{P}_*, \mathsf{Q}_*) = d\big(\Gamma_{\mathbf{x}}(\mu), \Gamma_{\mathbf{y}}(\mu)\big)\,,\ \ \forall\, \mu \in [0,1]\,, \tag{5.16}$$

whatever the configuration $\mathscr{B}_t$ of the beam is.

In summary, $\mathscr{A}_*$ is a spatial representation of all material sections $\mathcal{S}_\mu$ of the beam $\mathscr{B}$, such that each configuration $\mathscr{A}_s = \mathscr{C}_t(\mathcal{S}_\mu)$ can be obtained through an affine transformation $\mathcal{H}_s : \mathscr{A}_* \to \mathscr{A}_s$ mapping a point P_*, representative of a material fiber $\mathscr{f}_{\mathbf{x}}$ in the physical space, to the point P of the cross-section $\mathscr{A}_s$ belonging to the configuration at t of the same material fiber:

$$\begin{aligned} \mathcal{H}_s :\ & \mathscr{A}_* \to \mathscr{A}_s \\ & \mathsf{P}_* \mapsto \mathsf{P}\,, \end{aligned} \tag{5.17}$$

where $\mathsf{P}_* = \mathscr{C}_*(\ell_{\mathbf{x}})$ and $\mathsf{P} = \Gamma_{\mathbf{x}}(\mu)$, whence the rigidity condition (5.16) reads

$$d(\mathsf{P}_*, \mathsf{Q}_*) = d\big(\mathcal{H}_s(\mathsf{P}_*), \mathcal{H}_s(\mathsf{Q}_*)\big)\,, \quad \forall\, \mathsf{P}_*, \mathsf{Q}_* \in \mathscr{A}_*\,. \tag{5.18}$$

Please notice that, by Definition 5.4, any fiber $\ell_{\mathbf{x}}$ is uniquely associated with the material coordinates $\mathbf{x} \in \Sigma$ through $\psi_{\mathcal{S}}^{-1}$. Then, the composition $\varphi_* = \mathscr{C}_* \circ \psi_{\mathcal{S}}^{-1}$ is well-defined:

$$\begin{aligned} \mathscr{C}_* \circ \psi_{\mathcal{S}}^{-1} :\ & \Sigma \to \mathscr{A}_* \\ & \mathbf{x} \mapsto \mathsf{P}_*\,. \end{aligned} \tag{5.19}$$

Moreover, since it is assumed that the configuration $\mathscr{A}_s$ of the cross-section is plane, so is $\mathscr{A}_*$ and the plane of the physical space containing $\mathscr{A}_*$ is denoted as π_*. With this specification, the map $\mathscr{C}_* \circ \psi_{\mathcal{S}}^{-1}$ can be intended as the restriction to Σ of a map φ_* defined between $\mathbb{R}^2$ and π_*:

$$\begin{aligned} \varphi_* :\ & \mathbb{R}^2 \to \pi_* \\ & \mathbf{x} \mapsto \mathsf{P}_*\,, \end{aligned} \tag{5.20}$$

and then the affine transformation $\mathcal{H}_s$ can be defined between the plane π_* and π_s:

$$\begin{aligned} \mathcal{H}_s :\ & \pi_* \to \pi_s \\ & \mathsf{P}_* \mapsto \mathsf{P}\,. \end{aligned} \tag{5.21}$$

It is worth observing that φ_* defined by (5.20) is a map between the material representation of the beam cross-section, in a wide sense, and its configuration in the physical space $\mathscr{E}$. As a matter of fact, φ_* is a bijective map between the vector space $\mathbb{R}^2$ and the two-dimensional affine space π_* and, as such, can be applied to any $\mathbf{x} \in \mathbb{R}^2$, even when it does not detect any material fiber of the beam $\mathscr{B}$.

Notably, as shown in Figure 5.4, the point $\mathsf{O}_* = \varphi_*(0,0) \in \pi_*$ is considered as the representation of the beam axis in the physical space, even if the material fiber $a = \psi_{\mathcal{S}}^{-1}(0,0)$ is not defined, and the axis configuration at the time t is represented by the curve $\Gamma_{\mathbf{o}}$, defined by $\Gamma_{\mathbf{o}}(\mu) = \mathcal{H}_s(\mathsf{O}_*)$.

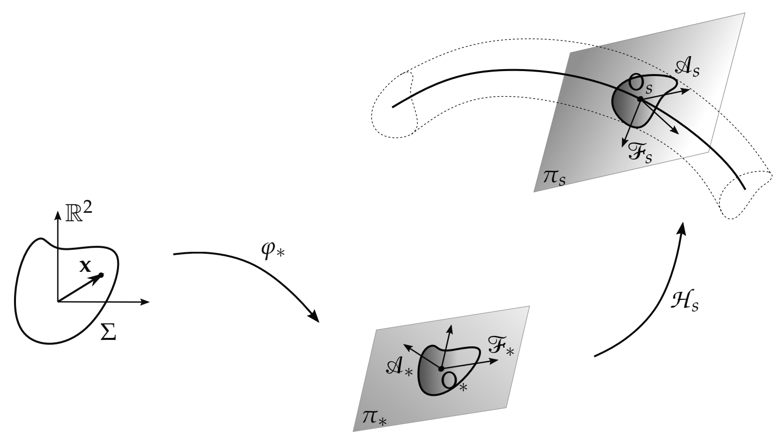


Figure 5.4: Cross-section configuration resulting from a rigid motion in $\mathscr{E}$.

Moreover, it is convenient to consider O_* as the origin of a reference frame $\mathscr{F}_*$ in such a way that the coordinates of any $\mathsf{P}_* \in \pi_*$ do coincide with $\mathbf{x} \in \mathbb{R}^2$. With this assumption, the material representation Σ of any beam cross-section $\mathcal{S}_\mu$ does coincide with the configuration $\mathscr{A}_*$ represented in the reference system $\mathscr{F}_*$.

Then, the point $\Gamma_{\mathbf{o}}(\mu) = \mathcal{H}_s(\mathsf{O}_*) = \mathsf{O}_s$ can be considered as the origin of a reference frame $\mathscr{F}_s$ attached at the plane π_s containing $\mathscr{A}_s$, and the motion $\mathcal{H}_s$ can considered as the change of frame from $\mathscr{F}_*$ to $\mathscr{F}_s$.

In addition, the hypothesis for the cross-section to be rigid induces to consider $\mathscr{F}_*$, as well as any $\mathscr{F}_s$, as an orthonormal frame (cf. Definition 4.12). Consequently, $\mathcal{H}_s$ results a proper Euclidean motion of $\mathscr{E}$ whose matrix representation, given by (4.27), takes the following form:

$$\mathcal{H}_s \cong \mathbf{H}_s = \begin{bmatrix} \mathbf{R}_s & \mathbf{v}_s \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix}. \tag{5.22}$$

Please observe that the subscript 's' in (5.22), representing the spatial abscissa relevant to $\mathscr{A}_s = \mathscr{C}_t(\mathcal{S}_\mu)$, should be intended as $s = s(t, \mu)$, so that the homogeneous matrix $\mathbf{H}_s$ is actually a function of t and μ, and the same applies to the entries $\mathbf{R}_s$ and $\mathbf{v}_s$:

$$\mathbf{H}_s = \mathbf{H}(t, \mu) = \begin{bmatrix} \mathbf{R}(t, \mu) & \mathbf{v}(t, \mu) \\ \mathbf{o}^\mathsf{T} & 1 \end{bmatrix} . \tag{5.23}$$

It is important to emphasize that, with the internal constraints on $\mathscr{A}_s$, the configuration map introduced by Definition 5.3 specializes to

$$\begin{aligned} \mathbb{Q} : \; & J \times [0, 1] \to \mathscr{E} \\ & (t, \mu) \mapsto \mathbb{Q}(t, \mu) = \mathscr{A}_s = \mathscr{C}_t(\mathscr{S}_\mu) \, . \end{aligned} \tag{5.24}$$

Moreover, by the introduction of the reference cross-section configuration $\mathscr{A}_*$ as image of the map (5.19), the same cross-section $\mathscr{A}_s$ also results from applying the Euclidean motion $\mathscr{H}_s$ defined by (5.21), that is

$$\mathscr{A}_s = \mathbb{Q}(t, \mu) = \mathscr{H}_s(\mathscr{A}_*) \, ,$$

and, recalling from (5.10) that $s = s(t, \mu)$, the map $(t, \mu) \mapsto \mathscr{H}_s$ is isomorphic with $\mathbb{Q}$.

Consequently, the *configuration space* of the beam, that is the space of the configuration maps $\mathbb{Q}$ as in (5.24), coincides with the space of the rigid motions $\mathbf{H}_s$, i.e. the special Euclidean group $SE(\mathscr{E})$ specified by Definition 4.16.

Further, the isomorphism $SE(\mathscr{E}) \cong SE(3)$ makes it possible to consider the group $SE(3)$ as the configuration space for the one-dimensional beam theory.

5.3 Beam Kinematics

It has been shown how the configuration $\mathscr{B}_t$ of the beam is defined by the configuration of any cross-section $\mathscr{A}_s$, relevant to the time $t \in J$ and the material abscissa μ. Moreover, under the hypotheses specified in Section 5.2.2, the configuration $\mathscr{A}_s$ is completely detected by a Euclidean proper motion $\mathscr{H}_s$, which is identified with the homogeneous matrix $\mathbf{H}_s$ in (5.22).

Hence, representing the configuration space through $SE(3)$, a comprehensive description of the beam kinematics comes from the characterization of $SE(3)$ as matrix Lie group.

Specifically, since a parameterization in the form $s \mapsto \mathbf{H}_s$ is a curve on the Lie group $SE(3)$, any information about the configuration of the beam can be derived considering the differential structure of $SE(3)$ as a smooth manifold.

However, in order to highlight the mechanical meaning of the algebraic and differential structure of $SE(3)$, it seems convenient to first derive the compatibility conditions of the beam kinematics from the perspective of its configuration in $\mathscr{E}$. Then, the validity of such conditions, within the Lie group structure of $SE(3)$, is discussed.

5.3.1 Beam Compatibility Equations

Let $\mathbf{H}_s$ be the homogeneous matrix defining the configuration of the cross-section $\mathscr{A}_s$. Supposing $\mathbf{H}_s$ to be fixed, let us consider the homogeneous matrix $\mathbf{H}_{s+\Delta s}$ associated with the configuration of the cross-section $\mathscr{A}_{s+\Delta s}$ at $s + \Delta s$.

Recalling that the spatial abscissa s is a function of μ by means of (5.10), the change of configuration from $\mathscr{A}_s$ to $\mathscr{A}_{s+\Delta s}$ actually corresponds to a variation of the material abscissa from μ to $\mu + h$.

Hence, making reference to a fixed configuration $\mathscr{C}_t$ and omitting the explicit dependence on t, the values of the spatial abscissa of the cross-sections $\mathscr{A}_s$ and $\mathscr{A}_{s+\Delta s}$ are

$$s = s(\mu)\,, \qquad s + \Delta s = s(\mu + h)\,,$$

while the relevant configuration matrices turn out to be

$$\mathbf{H}_s = \mathbf{H}(\mu)\,, \qquad \mathbf{H}_{s+\Delta s} = \mathbf{H}(\mu + h)\,.$$

Since $\mathbf{H}_{s+\Delta s}$ represents the configuration of the reference frame $\mathscr{F}_{s+\Delta s}$ attached to $\mathscr{A}_{s+\Delta s}$ with respect to the global frame $\mathscr{F}_*$, it can be thought as resulting from two subsequent transformations (see Figure 5.5).

Specifically, the configuration of the frame $\mathscr{F}_{s+\Delta s}$ is first represented with

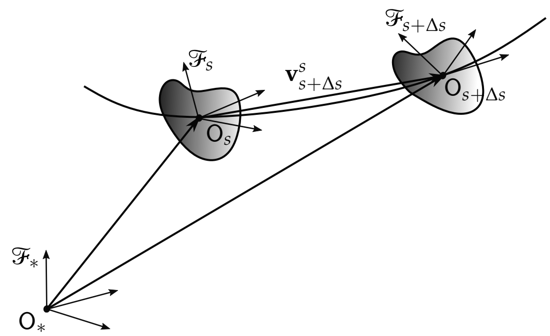


Figure 5.5: Relative configuration of beam cross-section frames $\mathscr{F}_s$ and $\mathscr{F}_{s+\Delta s}$.

respect to $\mathscr{F}_s$ by the matrix $\mathbf{H}^s_{s+\Delta s}$, and then $\mathscr{F}_s$ is referred to the global frame $\mathscr{F}_*$ through $\mathbf{H}_s$:

$$\mathbf{H}_{s+\Delta s} = \mathbf{H}_s \mathbf{H}^s_{s+\Delta s} \,,$$

or equivalently, emphasizing the dependence on the material abscissa,

$$\mathbf{H}(\mu + h) = \mathbf{H}(\mu)\mathbf{Q}(h) \,,$$

where the relative configuration matrix $\mathbf{H}^s_{s+\Delta s} = \mathbf{Q}(h)$ actually depends only on the increment h of the material abscissa, since $\mathscr{A}_s$ is taken as a reference and $\mathbf{H}_s$ is considered fixed.

Consequently, when one moves from $\mathscr{A}_s$ to $\mathscr{A}_{s+\Delta s}$, the variation of the configuration matrix results

$$\Delta\mathbf{H} = \mathbf{H}(\mu + h) - \mathbf{H}(\mu) = \mathbf{H}(\mu)\big(\mathbf{Q}(h) - \mathbf{I}\big) \,, \tag{5.25}$$

and the rate of variation of $\mathbf{H}$ is given by

$$\frac{\Delta\mathbf{H}}{\Delta s} = \frac{\mathbf{H}(\mu)\big(\mathbf{Q}(h) - \mathbf{I}\big)}{s(\mu + h) - s(\mu)} \,,$$

since the arc length between $\mathscr{A}_s$ and $\mathscr{A}_{s+\Delta s}$ is $\Delta s = s(\mu + h) - s(\mu)$.

With the aim to characterize the rate of variation of $\mathbf{H}$ per unitary length, let us consider the limit, as Δs approaches 0, of the above ratio, also corresponding to $h \to 0$:

$$\frac{\mathrm{d}\mathbf{H}}{\mathrm{d}s} = \lim_{\Delta s \to 0} \frac{\Delta \mathbf{H}}{\Delta s} = \lim_{h \to 0} \frac{\mathbf{H}(\mu + h) - \mathbf{H}(\mu)}{s(\mu + h) - s(\mu)}$$
$$= \lim_{h \to 0} \frac{\mathbf{H}(\mu + h) - \mathbf{H}(\mu)}{h} \lim_{h \to 0} \frac{h}{s(\mu + h) - s(\mu)} \,,$$

whence, exploiting the map λ defined by (5.11) and denoting as $\mathbf{H}'(\mu)$ the derivative of $\mathbf{H}$ with respect to μ, one finds

$$\left.\frac{\mathrm{d}\mathbf{H}}{\mathrm{d}s}\right|_{s(\mu)} = \frac{\mathbf{H}'(\mu)}{\lambda(\mu)} \,. \tag{5.26}$$

It is worth recalling that $\lambda(\mu)$ is the norm of the vector $\Gamma'_{\mathbf{o}}(\mu)$ tangent to the beam axis in its spatial configuration. Hence, due the regularity of such a curve of $\mathscr{E}$, the condition $\lambda(\mu) \neq 0$ is satisfied, and the derivative here above is well-defined.

In addition, by means of (5.25), the derivative of $\mathbf{H}$ with respect to the material abscissa reads:

$$\mathbf{H}'(\mu) = \lim_{h \to 0} \frac{\mathbf{H}(\mu + h) - \mathbf{H}(\mu)}{h} = \mathbf{H}(\mu) \lim_{h \to 0} \frac{\mathbf{Q}(h) - \mathbf{I}}{h} \,. \tag{5.27}$$

Please recall that $\mathbf{Q}(h) = \mathbf{H}^{s}_{s+\Delta s}$ is the configuration matrix of $\mathscr{F}_{s+\Delta s}$ with respect to $\mathscr{F}_s$, so that it can be expressed as

$$\mathbf{Q}(h) = \begin{bmatrix} \mathbf{R}(h) & \mathbf{v}(h) \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix} , \tag{5.28}$$

where $\mathbf{R}(h) = \mathbf{R}^{s}_{s+\Delta s}$ represents the rotation of $\mathscr{F}_{s+\Delta s}$ with respect to $\mathscr{F}_s$ and $\mathbf{v}(h) = \mathbf{v}^{s}_{s+\Delta s}$ is the position vector of the origin $\mathsf{O}_{s+\Delta s}$ of $\mathscr{F}_{s+\Delta s}$ with respect to the origin O_s of $\mathscr{F}_s$ (see Figure 5.5).

Moreover, having assumed that the configuration of the cross-section changes with continuity along the beam axis, it is expected that $\mathscr{F}_{s+\Delta s}$ overlaps $\mathscr{F}_s$ when $\Delta s = 0$, that is when $h = 0$. Hence, the condition $\mathbf{Q}(0) = \mathbf{I}$ applies, which is the same as $\mathbf{R}(0) = \mathbf{I}$ and $\mathbf{v}(0) = \mathbf{o}$.

Furthermore, since $\mathbf{Q}(h)$ represents a proper Euclidean motion, it is associated with a screw $\big(\boldsymbol{\omega}(h), \mathbf{x}(h)\big) \cong \mathbf{q}(h) \in \mathbb{R}^6$ (cf. Section 4.4). Hence, using (4.72), it can be written as

$$\mathbf{Q}(h) = \begin{bmatrix} \mathbf{R}_{\boldsymbol{\omega}(h)} & \mathbf{T}^{\mathsf{T}}_{\boldsymbol{\omega}(h)}\mathbf{x}(h) \\ \mathbf{o}^{\mathsf{T}} & 1 \end{bmatrix},$$

where, setting $\omega = \|\boldsymbol{\omega}\|$, the rotation matrix $\mathbf{R}_{\boldsymbol{\omega}(h)}$ is expressed by the Rodrigues' formula in (4.50) as

$$\mathbf{R}_{\boldsymbol{\omega}(h)} = \mathbf{I} + \sin\omega\frac{\widehat{\boldsymbol{\omega}}}{\omega} + (1-\cos\omega)\frac{\widehat{\boldsymbol{\omega}}^2}{\omega^2}, \tag{5.29}$$

and $\mathbf{T}^{\mathsf{T}}_{\boldsymbol{\omega}(h)} = \mathbf{T}_{-\boldsymbol{\omega}(h)}$ is the tangent operator of $SO(3)$, whose expression, by means of (4.58), is given by

$$\mathbf{T}^{\mathsf{T}}_{\boldsymbol{\omega}(h)} = \mathbf{I} + \frac{1-\cos\omega}{\omega}\frac{\widehat{\boldsymbol{\omega}}}{\omega} + \left(1 - \frac{\sin\omega}{\omega}\right)\frac{\widehat{\boldsymbol{\omega}}^2}{\omega^2}. \tag{5.30}$$

Please also observe that the condition $\mathbf{Q}(0) = \mathbf{I}$ is reflected in the vanishing of the associated screw, that is $\mathbf{q}(0) = \mathbf{o}$, or equivalently $\boldsymbol{\omega}(0) = \mathbf{o}$ and $\mathbf{x}(0) = \mathbf{o}$.

With these specifications, the limit in (5.27) reads

$$\lim_{h\to 0}\frac{\mathbf{Q}(h)-\mathbf{I}}{h} = \begin{bmatrix} \lim_{h\to 0}\dfrac{\mathbf{R}_{\boldsymbol{\omega}(h)}-\mathbf{I}}{h} & \lim_{h\to 0}\dfrac{\mathbf{T}^{\mathsf{T}}_{\boldsymbol{\omega}(h)}\mathbf{x}(h)}{h} \\ \mathbf{o}^{\mathsf{T}} & 0 \end{bmatrix}. \tag{5.31}$$

In order to evaluate the entries of the matrix here above, let us first report the following limit evaluations:

$$\lim_{h\to 0}\frac{\sin\omega}{\omega} = 1, \qquad \lim_{h\to 0}\frac{1-\cos\omega}{\omega^2} = \frac{1}{2}, \qquad \lim_{h\to 0}\frac{\omega-\sin\omega}{\omega^3} = \frac{1}{6}, \tag{5.32}$$

which, considering $\omega(0) = \|\boldsymbol{\omega}(0)\| = 0$, can be easily verified by applying L'Hospital's rule.

Accordingly, using the expression (5.29), the first limit in (5.31) is evaluated as follows:

$$\begin{aligned}\lim_{h\to 0}\frac{\mathbf{R}_{\boldsymbol{\omega}(h)}-\mathbf{I}}{h} &= \lim_{h\to 0}\frac{\sin\omega}{\omega}\frac{\widehat{\boldsymbol{\omega}}}{h}+\lim_{h\to 0}\frac{1-\cos\omega}{\omega^2}\frac{\widehat{\boldsymbol{\omega}}^2}{h} \\ &= \lim_{h\to 0}\frac{\widehat{\boldsymbol{\omega}}(h)}{h}+\frac{1}{2}\lim_{h\to 0}\frac{\widehat{\boldsymbol{\omega}}(h)}{h}\widehat{\boldsymbol{\omega}}(h)\,,\end{aligned} \tag{5.33}$$

where the first addend is the derivative of the skew-symmetric matrix $\widehat{\boldsymbol{\omega}}(h)$, associated with the vector $\boldsymbol{\omega}(h)$. The existence of such a derivative will be discussed in Section 5.3.2. Here, it is enough to say that $\boldsymbol{\omega}'(0)$, as well as $\widehat{\boldsymbol{\omega}}'(0)$, does exist and its value is a vector $\mathbf{d}_{\mathsf{r}}$:

$$\lim_{h\to 0}\frac{\boldsymbol{\omega}(h)}{h}=\boldsymbol{\omega}'(0)=\mathbf{d}_{\mathsf{r}}\,; \tag{5.34}$$

this result, being $\boldsymbol{\omega}(0)=\mathbf{o}$, also implies the vanishing of the second addend in (5.33), and the limit on the left-hand side is evaluated as

$$\lim_{h\to 0}\frac{\mathbf{R}_{\boldsymbol{\omega}(h)}-\mathbf{I}}{h}=\widehat{\mathbf{d}}_{\mathsf{r}}\,. \tag{5.35}$$

Regarding the limit of the translation vector in (5.28), please notice that the operator $\mathbf{T}^{\mathsf{T}}_{\boldsymbol{\omega}(h)}$ tends to $\mathbf{I}$ as $h\to 0$. Actually, recalling the expression (5.30) of $\mathbf{T}^{\mathsf{T}}_{\boldsymbol{\omega}(h)}$ and using the evaluations (5.32), one infers

$$\begin{aligned}\lim_{h\to 0}\mathbf{T}^{\mathsf{T}}_{\boldsymbol{\omega}(h)} &= \mathbf{I}+\lim_{h\to 0}\frac{1-\cos\omega}{\omega^2}\widehat{\boldsymbol{\omega}}+\lim_{h\to 0}\frac{\omega-\sin\omega}{\omega^3}\widehat{\boldsymbol{\omega}}^2 = \\ &= \mathbf{I}+\frac{1}{2}\lim_{h\to 0}\widehat{\boldsymbol{\omega}}(h)+\frac{1}{6}\lim_{h\to 0}\widehat{\boldsymbol{\omega}}^2(h)=\mathbf{I}\,.\end{aligned}$$

In addition, let us denote as $\mathbf{d}_{\mathsf{t}}$ the derivative of $\mathbf{x}(h)$ with respect to h:

$$\lim_{h\to 0}\frac{\mathbf{x}(h)}{h}=\mathbf{x}'(0)=\mathbf{d}_{\mathsf{t}}\,, \tag{5.36}$$

whose existence, similarly to $\mathbf{d}_{\mathsf{r}}=\boldsymbol{\omega}'(0)$, will be discussed in Section 5.3.2. Accordingly, the translation vector in (5.28) becomes

$$\lim_{h\to 0}\frac{\mathbf{v}(h)}{h}=\lim_{h\to 0}\frac{\mathbf{T}^{\mathsf{T}}_{\boldsymbol{\omega}(h)}\mathbf{x}(h)}{h}=\lim_{h\to 0}\frac{\mathbf{x}(h)}{h}=\mathbf{d}_{\mathsf{t}}\,. \tag{5.37}$$

In conclusion, using the evaluations (5.35) and (5.37) in (5.31), one finds

$$\lim_{h\to 0}\frac{\mathbf{Q}(h)-\mathbf{I}}{h}=\begin{bmatrix}\widehat{\mathbf{d}}_{\mathsf{r}} & \mathbf{d}_{\mathsf{t}}\\ \mathbf{o}^{\mathsf{T}} & 0\end{bmatrix}=\begin{bmatrix}\mathbf{d}_{\mathsf{r}}\\ \mathbf{d}_{\mathsf{t}}\end{bmatrix}^{\widehat{}}=\widehat{\mathbf{d}}\,, \tag{5.38}$$

where the map (4.78) ha been used, since $\widehat{\mathbf{d}}_{\mathsf{r}}$ is the skew-symmetric matrix relevant to the vector $\mathbf{d}_{\mathsf{r}}\in\mathbb{R}^3$ and $\mathbf{d}_{\mathsf{t}}$ is a vector in $\mathbb{R}^3$.

Therefore, using the result (5.38), in (5.27) the derivative of the configuration matrix $\mathbf{H}$ with respect to the beam material abscissa is

$$\mathbf{H}'(\mu)=\mathbf{H}(\mu)\widehat{\mathbf{d}}(\mu)\,, \tag{5.39}$$

where the explicit dependence of $\widehat{\mathbf{d}}$ on μ is introduced to highlight that, consistently with the described procedure, it is actually associated with the cross-section $\mathscr{A}_s$ relevant the specific value μ of the material abscissa.

Moreover, from (5.26), the variation of $\mathbf{H}$ per unitary length becomes

$$\left.\frac{\mathrm{d}\mathbf{H}}{\mathrm{d}s}\right|_{s(\mu)}=\frac{\mathbf{H}'(\mu)}{\lambda(\mu)}=\mathbf{H}(\mu)\frac{\widehat{\mathbf{d}}(\mu)}{\lambda(\mu)}\,. \tag{5.40}$$

Please recall from (5.11) that $\lambda(\mu)$ is the length of the tangent vector $\Gamma'_{\mathbf{o}}(\mu)$ to the configuration of the beam axis. Moreover, in coordinates, such a vector is given by the translation component of $\mathbf{H}'(\mu)$, resulting

$$\mathbf{v}'(\mu)=\mathbf{R}(\mu)\mathbf{d}_{\mathsf{t}}(\mu)\,; \tag{5.41}$$

in turn, being $\mathbf{R}(\mu)$ orthogonal, this implies

$$\lambda(\mu)=\|\mathbf{v}'(\mu)\|=\|\mathbf{d}_{\mathsf{t}}(\mu)\|\,. \tag{5.42}$$

5.3.2 Algebra of Beam Deformations

In Section 5.3.1 the kinematics of the beam has been characterized in terms of configuration of each cross-section in the physical space, and has been represented by means of an homogeneous matrix $\mathbf{H}$ varying smoothly along the beam axis.

Considering the Lie group structure of the space $SE(3)$, the parameterization $\mu \mapsto \mathbf{H}(\mu)$ is a smooth curve, whose velocity at μ is a vector $\mathbf{H}'(\mu)$ of the tangent space $T_{\mathbf{H}(\mu)}SE(3)$.

Moreover, for each $\mu \in [0,1]$, the point $\mathbf{H}(\mu) = \mathbf{H}_\mu$ of $SE(3)$ can be considered as the starting point of a smooth curve defined as follows:

$$\begin{aligned} \sigma :\ & \mathbb{R} \to SE(3) \\ & h \mapsto \sigma(h) = \mathbf{H}(\mu + h)\,, \end{aligned}$$

that can be easily seen to satisfy the initial conditions $\sigma(0) = \mathbf{H}_\mu$ and $\sigma'(0) = \mathbf{H}'(\mu)$.

At same time, by left translation, the curve σ induces a further curve $\gamma = L_{\mathbf{H}_\mu^{-1}} \circ \sigma$, defined by $\gamma(h) = (L_{\mathbf{H}_\mu^{-1}} \circ \sigma)(h) = \mathbf{H}_\mu^{-1}\mathbf{H}(\mu + h)$.

It is easy to verify that the starting point of γ is $\gamma(0) = \mathbf{I}$, and that the relevant velocity is a vector of the tangent space $T_{\mathbf{I}}SE(3)$ evaluated as follows (cf. Proposition 2.15):

$$\gamma'(0) = (L_{\mathbf{H}_\mu^{-1}} \circ \sigma)'(0) = \mathrm{d}L_{\mathbf{H}_\mu^{-1}}\Big|_{\sigma(0)} \sigma'(0) = \mathbf{H}_\mu^{-1}\mathbf{H}'(\mu)\,, \tag{5.43}$$

where the initial condition $\sigma'(0) = \mathbf{H}'(\mu)$ has been used, along with the property $\mathrm{d}L_{\mathbf{H}_\mu^{-1}}\Big|_{\sigma(0)} = \mathbf{H}_\mu^{-1}$, as given by (2.48) for matrix Lie groups.

Let us also notice that the point of γ at h is given by

$$\gamma(h) = \mathbf{Q}(h)\,,$$

where $\mathbf{Q}(h)$ is the homogeneous matrix, represented by (5.28), that provides the relative configuration of the beam cross-section $\mathcal{A}_{s+\Delta s}$ with respect to $\mathcal{A}_s$ (cf. Section 5.3.1).

Furthermore, the screw associated with the motion $\mathbf{Q}(h)$, denoted as $\widehat{\mathbf{q}}(h) \in \mathfrak{se}(3)$, results from applying the logarithm map of $SE(3)$ at $\mathbf{Q}(h)$, accordingly with the definition (4.82).

Hence, the parameterization $h \mapsto \widehat{\mathbf{q}}(h)$, given by $\widehat{\mathbf{q}}(h) = (\log \circ \gamma)(h)$, is a smooth curve of $\mathfrak{se}(3)$, as a consequence of the smoothness of both γ and log.

Specifically, the starting point of $\log \circ \gamma$ is the null vector of $\mathfrak{se}(3)$, i.e.

$$\widehat{\mathbf{q}}(0) = \log(\mathbf{I}) = \mathbf{0}\,,$$

and the initial velocity is evaluated as follows (cf. Proposition 2.15):

$$\widehat{\mathbf{q}}'(0) = (\log \circ \gamma)'(0) = \left.\frac{\mathrm{d}}{\mathrm{d}h}\right|_0 (\log \circ \gamma) = \mathrm{dlog}|_{\gamma(0)}\gamma'(0) = \mathrm{dexp}|_{\mathbf{0}}^{-1}\gamma'(0)\,,$$

where $\gamma'(0) = \mathbf{H}_\mu^{-1}\mathbf{H}'(\mu)$ is given by (5.43).

Accordingly, recalling from (2.39) that the differential of the exponential at $\mathbf{0}$ is the identity map, and so is its inverse $\mathrm{dexp}|_{\mathbf{0}}^{-1}$, one finally finds

$$\widehat{\mathbf{q}}'(0) = \mathbf{H}_\mu^{-1}\mathbf{H}'(\mu) = \widehat{\mathbf{d}}(\mu) \in \mathfrak{se}(3)\,, \tag{5.44}$$

or, using the vector space isomorphism $\mathfrak{se}(3) \cong \mathbb{R}^6$,

$$\mathbf{q}'(0) = \mathbf{d}(\mu) \in \mathbb{R}^6\,. \tag{5.45}$$

Please observe that, specifying the rotation $\boldsymbol{\omega}(h)$ and translation component $\mathbf{x}(h)$ of the screw vector $\mathbf{q}(h)$, the relation (5.45) also reads

$$\boldsymbol{\omega}'(0) = \mathbf{d}_\mathsf{r}\,, \qquad \mathbf{x}'(0) = \mathbf{d}_\mathsf{t}\,,$$

so that the limits in (5.34) and (5.36) are actually well-defined.

Moreover, the evaluation (5.44) for the velocity of the curve $h \mapsto \widehat{\mathbf{q}}(h)$ also implies the compatibility condition of the beam kinematics expressed by (5.39):

$$\mathbf{H}'(\mu) = \mathbf{H}(\mu)\widehat{\mathbf{d}}(\mu)\,, \tag{5.46}$$

that, with the usual identification $\mathfrak{se}(3) \cong T_\mathbf{I}SE(3)$, is formally given by the following map:

$$\begin{aligned} \mathrm{d}L_{\mathbf{H}(\mu)}\big|_\mathbf{I} :\ & T_\mathbf{I}SE(3) \to T_{\mathbf{H}(\mu)}SE(3) \\ & \widehat{\mathbf{d}}(\mu) \mapsto \mathbf{H}'(\mu) = \mathrm{d}L_{\mathbf{H}(\mu)}\big|_\mathbf{I}\widehat{\mathbf{d}}(\mu)\,. \end{aligned} \tag{5.47}$$

Hence, in view of the smooth manifold structure of $SE(3)$, the derivative

$\mathbf{H}'(\mu)$ of the configuration matrix $\mathbf{H}(\mu)$, associated with the beam cross-section $\mathscr{A}_s$, is actually a vector within the tangent space $T_{\mathbf{H}(\mu)}SE(3)$ resulting from the left translation of the screw $\widehat{\mathbf{d}}(\mu) \in \mathfrak{se}(3) \cong T_{\mathbf{I}}SE(3)$.

In other words, the parameterization $\mu \mapsto \mathbf{H}(\mu)$ represents the integral curve of a vector field $\boldsymbol{V}$ on $SE(3)$, that satisfies $\boldsymbol{V}_{\mathbf{H}(\mu)} = \mathbf{H}'(\mu)$ at each $\mathbf{H}(\mu)$ (cf. Section 2.3.3).

Furthermore, with reference to the beam spatial abscissa, the real-valued function $\mu \mapsto s(\mu)$, defined by (5.10), is a diffeomorphism between $[0, 1]$ and $[0, l]$, whose derivative is the function λ given by (5.11). Consequently, the inverse map $\tau \colon s \mapsto \mu$ is itself a real-valued function and the derivative is $\tau' \colon s \mapsto 1/\lambda\big(\mu(s)\big)$.

Finally, recalling Proposition 2.25, the map $s \mapsto \mathbf{H}(s)$ is the integral curve of a vector field $\tau'\boldsymbol{V}$ defined by the condition

$$\left.\frac{\mathrm{d}\mathbf{H}}{\mathrm{d}s}\right|_s = \tau'(s)\boldsymbol{V}_{\mathbf{H}(s)} = \frac{\boldsymbol{V}_{\mathbf{H}(\mu(s))}}{\lambda\big(\mu(s)\big)} = \frac{\mathbf{H}'\big(\mu(s)\big)}{\lambda\big(\mu(s)\big)}\,, \tag{5.48}$$

that is the same as (5.40) considering s as the independent variable.

5.4 Force Representation and Equilibrium Configurations

The internal constraints introduced in Section 5.2.2, that characterize the beam cross-section as plane and rigid in any spatial configuration, has led to consider the Lie group $SE(3)$ as the configuration space of the beam.

Consequently, in considering the virtual displacements of the beam, with the classical meaning of arbitrary infinitesimal changes of a configuration (see, e.g., Arnold (1989), Goldstein et al. (2001)), the smooth manifold structure of $SE(3)$ is also involved.

Specifically, when the configuration space of a mechanical system is represented in terms of a smooth manifold, the virtual displacements should be intended as elements of the tangent space (see, e.g., Epstein (2010), Epstein and Segev (1980), Eugster (2015)). Moreover, since a specific tangent space $T_{\mathbf{H}}SE(3)$ is associated with any point $\mathbf{H} \in SE(3)$, it is appropriate to define the *virtual displacement field* as the tangent bundle $TSE(3)$ of $SE(3)$ (cf. Section 2.2.4).

At this point, following the same approach proposed by Epstein (2010), the *space of forces* is introduced as the cotangent bundle $T^*SE(3)$ of $SE(3)$, intended as the dual space of the tangent bundle (cf. Section 2.2.4).

Such a notion of a force follows from the assumption the configuration space is $SE(3)$ and, consequently, the virtual displacements are tangent vectors on $SE(3)$. Accordingly, for any point $\mathbf{H} \in SE(3)$, a virtual displacement is a vector $\delta\mathbf{H}$ of the tangent space $T_{\mathbf{H}}SE(3)$, and a force is an element $\mathbf{f}^*_{\mathbf{H}}$ of the cotangent space $T^*_{\mathbf{H}}SE(3)$.

As a consequence, a force $\mathbf{f}^*_{\mathbf{H}}$, considered as a covector within the dual space $T^*_{\mathbf{H}}SE(3)$ of the tangent space at $\mathbf{H}$, plays the role of a functional acting on any $\delta\mathbf{H} \in T_{\mathbf{H}}SE(3)$ to give a scalar.

More properly, the real number resulting from applying $\mathbf{f}^*_{\mathbf{H}}$ at $\delta\mathbf{H}$ is the *virtual work* of the force $\mathbf{f}^*_{\mathbf{H}}$ by the virtual displacement $\delta\mathbf{H}$:

$$\delta W = \mathbf{f}^*_{\mathbf{H}}(\delta\mathbf{H})\,. \tag{5.49}$$

5.4.1 Virtual Displacement Screws and Wrenches

Let us first observe that any virtual displacement $\delta\mathbf{H} \in T_{\mathbf{H}}SE(3)$ can be thought as resulting from the left translation of a screw $\delta\widehat{\mathbf{h}} \in \mathfrak{se}(3)$. Hence, recalling that (2.48) applies for matrix Lie groups under the usual identification $T_{\mathbf{I}}SE(3) \cong \mathfrak{se}(3)$, one can write

$$\delta\mathbf{H} = \mathbf{H}\delta\widehat{\mathbf{h}} \in T_{\mathbf{H}}SE(3)\,, \tag{5.50}$$

where $\mathbf{H}$ plays the role of a linear map $\mathbf{H}\colon \mathfrak{se}(3) \to T_{\mathbf{H}}SE(3)$.

As a consequence, the virtual displacement field, represented by the tangent bundle $TSE(3)$, can actually be identified with $SE(3) \times \mathfrak{se}(3)$, meaning that any $(\mathbf{H}, \delta\widehat{\mathbf{h}})$, or simply $\delta\widehat{\mathbf{h}}$ if the point $\mathbf{H}$ appears from the context, gives the virtual displacement $\delta\mathbf{H} = \mathbf{H}\delta\widehat{\mathbf{h}} \in T_{\mathbf{H}}SE(3)$.

Similarly, considering the dual map $\mathbf{H}^{\mathsf{T}} : T^*_{\mathbf{H}}SE(3) \to \mathfrak{se}^*(3)$ of $\mathbf{H}$, any force $\mathbf{f}^*_{\mathbf{H}} \in T^*_{\mathbf{H}}SE(3)$ can be represented by its pullback $\widehat{\mathbf{f}}^* \in \mathfrak{se}^*(3)$:

$$\widehat{\mathbf{f}}^* = \mathbf{H}^{\mathsf{T}}\mathbf{f}^*_{\mathbf{H}} \in \mathfrak{se}^*(3)\,, \tag{5.51}$$

a quantity that, recalling Definition B.26, satisfies

$$\delta W = \mathbf{f}^*_{\mathbf{H}}(\delta\mathbf{H}) = \widehat{\mathbf{f}}^*(\delta\widehat{\mathbf{h}})\,. \tag{5.52}$$

Any vector $\widehat{\mathbf{f}}^* \in \mathfrak{se}^*(3)$, acting as a functional for the screws in $\mathfrak{se}(3)$, is called a *coscrew*.

In addition, since the dual space $\mathfrak{se}^*(3)$ is itself a six-dimensional vector space, hence isomorphic to $\mathbb{R}^6$; consequently, any coscrew $\widehat{\mathbf{f}}^*$ can be represented by a vector $\mathbf{f}^* \in \mathbb{R}^6$ such that, exploiting the Euclidean structure of $\mathbb{R}^6$, the virtual work $\delta W = \widehat{\mathbf{f}}^*(\delta\widehat{\mathbf{h}})$ can be evaluated by the dot product as in (4.5):

$$\delta W = \widehat{\mathbf{f}}^*(\delta\widehat{\mathbf{h}}) = \mathbf{f}^* \cdot \delta\mathbf{h}\,, \tag{5.53}$$

where $\delta\mathbf{h} \in \mathbb{R}^6$ is the column vector associated with the screw $\delta\widehat{\mathbf{h}}$ by means of the map (4.78).

Expressing the virtual work δW in the form (5.53) provides a physical meaning to the coscrew $\mathbf{f}^*$. Actually, the column vector $\mathbf{f}^* \in \mathbb{R}^6$ can be thought as made of two components as follows:

$$\mathbf{f}^* = \begin{bmatrix} \mathbf{f}_{\mathsf{m}} \\ \mathbf{f}_{\mathsf{f}} \end{bmatrix}, \tag{5.54}$$

where $\mathbf{f}_{\mathsf{f}} \in \mathbb{R}^3$ is a vector representing a force applied at a point of the physical space $\mathscr{E}$, and $\mathbf{f}_{\mathsf{m}} \in \mathbb{R}^3$ represents its moment with respect to some point of $\mathscr{E}$.

Consequently, denoting as $\delta\mathbf{h}_{\mathsf{r}}$ and $\delta\mathbf{h}_{\mathsf{t}}$ the rotation and the translation components of the screw $\delta\mathbf{h}$, the virtual work δW in (5.53) becomes

$$\delta W = \mathbf{f}^* \cdot \delta\mathbf{h} = \mathbf{f}_{\mathsf{f}} \cdot \delta\mathbf{h}_{\mathsf{t}} + \mathbf{f}_{\mathsf{m}} \cdot \delta\mathbf{h}_{\mathsf{r}}\,. \tag{5.55}$$

It is clear from (5.54) that the entries of $\mathbf{f}^* \in \mathbb{R}^6$ provide the matrix representation $\mathbf{f} \in \mathbb{R}^6$ of a screw $\widehat{\mathbf{f}} \in \mathfrak{se}(3)$ that, in turn, can be identified with the pair $(\mathbf{f}_\mathsf{f}, \mathbf{f}_\mathsf{m})$ by means of (4.75), that is

$$(\mathbf{f}_\mathsf{f}, \mathbf{f}_\mathsf{m}) \cong \mathbf{f} = \begin{bmatrix} \mathbf{f}_\mathsf{f} \\ \mathbf{f}_\mathsf{m} \end{bmatrix};$$

this quantity, called a *wrench*, has direction defined by the force component $\mathbf{f}_\mathsf{f}$ and position associated with the moment component $\mathbf{f}_\mathsf{m}$. At the same time, the energetic coupling of a wrench $(\mathbf{f}_\mathsf{f}, \mathbf{f}_\mathsf{m})$ with a kinematic screw takes the form of the dot product in $\mathbb{R}^6$ if the coscrew representation $\mathbf{f}^*$ is considered.

It is important to remark that a screw $(\delta\mathbf{h}_\mathsf{r}, \delta\mathbf{h}_\mathsf{t})$, being related with the virtual displacement $\delta\mathbf{H} \in T_\mathbf{H}SE(3)$ by means of (5.50), has the kinematic meaning of the infinitesimal variation of a reference system $\mathscr{F}$ of the physical space $\mathscr{E}$, represented by the configuration matrix $\mathbf{H} \in SE(3)$.

In parallel, a vector $\delta\mathbf{h} \in \mathbb{R}^6$ actually takes the meaning of a screw representing a virtual displacement only if the reference system $\mathscr{F}$ is specified, or equivalently the configuration $\mathbf{H}$ at which such a variation is applied.

In conclusion, the virtual displacement of a framed point in $\mathscr{E}$ should be intended as a pair $(\mathbf{H}, \delta\mathbf{h}) \cong (\mathbf{H}, \delta\mathbf{H}) \in TSE(3)$.

This is also reflected on the representation of forces within the physical space $\mathscr{E}$. In fact, since a wrench $(\mathbf{f}_\mathsf{f}, \mathbf{f}_\mathsf{m})$ is necessarily associated with a screw $(\delta\mathbf{h}_\mathsf{r}, \delta\mathbf{h}_\mathsf{t})$ by means of (5.55), it is well-defined only by concurrently specifying the reference system $\mathscr{F}$, seen as a framed point of $\mathscr{E}$, at which it is applied.

Consequently, a wrench should be in turn understood as a pair $(\mathbf{H}, \mathbf{f}) \cong (\mathbf{H}, \mathbf{f}^*_\mathbf{H}) \in T^*SE(3)$, where the identification with the cotangent vector $\mathbf{f}^*_\mathbf{H} \in T^*_\mathbf{H}SE(3)$ follows from the definition of the virtual work as an invariant scalar quantity:

$$\delta W = \mathbf{f}^* \cdot \delta\mathbf{h} = \mathbf{f} \cdot \delta\mathbf{h}^* = \mathbf{f}^*_\mathbf{H}(\delta\mathbf{H}) \,. \tag{5.56}$$

Following again the approach outlined, among others, by Epstein (2010), Epstein and Segev (1980), Eugster (2015), a configuration of a system is said to be an *equilibrium configuration* if the virtual work δW does vanish for all virtual displacements.

In the context of this work, a configuration $\mathbf{H}$ of a framed point of the physical space $\mathscr{E}$, consisting of an affine frame $\mathscr{F}$, is an equilibrium configuration if the virtual work δW is null for any virtual displacement in the tangent space of $SE(3)$ at $\mathbf{H}$:

$$\delta W = \mathbf{f}^*_{\mathbf{H}}(\delta\mathbf{H}) = 0\,, \quad \forall\,\delta\mathbf{H} \in T_{\mathbf{H}}SE(3)\,. \tag{5.57}$$

It is worth noting that the assumption stated by (5.57), that specializes the *Principle of Virtual Work* to a framed point of the physical space, is actually full consistent with the notion of equilibrium in classical mechanics (see, e.g., Arnold (1989), Goldstein et al. (2001)).

In fact, recalling Part 2 of Proposition B.25, the arbitrariness of $\delta\mathbf{H}$ in (5.57) implies that $\mathbf{f}^*_{\mathbf{H}}$ is null. At the same time, since any cotangent vector $\mathbf{f}^*_{\mathbf{H}} \in T^*_{\mathbf{H}}SE(3)$ has a relevant physical representation by a wrench $(\mathbf{f}_{\mathsf{f}}, \mathbf{f}_{\mathsf{m}})$, the condition $\mathbf{f}^*_{\mathbf{H}} = \mathbf{o}^*$ is reflected on the vanishing of the resultant force $\mathbf{f}_{\mathsf{f}}$ and of the resultant moment $\mathbf{f}_{\mathsf{m}}$ acting on $\mathbf{H}$.

Hence, the configuration of a point P of the physical space, along with an attached frame $\mathscr{F}$, is an equilibrium configuration if the resultant force and moment acting at P do vanish.

5.4.2 Virtual Displacement Field of a Beam

With specific reference to the beam kinematics, the representation in terms of a curve $\mu \mapsto \mathbf{H}(\mu)$ on $SE(3)$ implies that the virtual displacement field is actually a subspace of $TSE(3)$, and consists of all tangent vectors $\delta\mathbf{H}(\mu)$ satisfying the condition $\delta\mathbf{H}(\mu) \in T_{\mathbf{H}(\mu)}SE(3)$, for each $\mu \in [0,1]$.

Moreover, recalling that any $\mathbf{H} \in SE(3)$ concurrently applies as a linear operator $\mathbf{H}\colon T_{\mathbf{I}}SE(3) \to T_{\mathbf{H}}SE(3)$, a compatible virtual displacement $\delta\mathbf{H}(\mu)$ can be obtained by left translation as

$$\delta\mathbf{H}(\mu) = \mathbf{H}(\mu)\delta\widehat{\mathbf{h}}(\mu)\,, \tag{5.58}$$

where $\delta\widehat{\mathbf{h}}(\mu) \in T_{\mathbf{I}}SE(3) \cong \mathfrak{se}(3)$ is a screw variation compatible with the virtual displacement field of the beam.

Furthermore, let us consider the relative configuration screw $\widehat{\mathbf{d}}(\mu) \in \mathfrak{se}(3)$, associated with the derivative $\mathbf{H}'(\mu)$ by the compatibility condition (5.39). Intending any quantity evaluated at an arbitrary $\mu \in [0,1]$, the infinitesimal variation of $\widehat{\mathbf{d}}$ reads

$$\delta\widehat{\mathbf{d}} = \delta(\mathbf{H}^{-1}\mathbf{H}') = \delta\mathbf{H}^{-1}\mathbf{H}' + \mathbf{H}^{-1}\delta(\mathbf{H}') = -\delta\widehat{\mathbf{h}}\,\mathbf{H}^{-1}\mathbf{H}' + \mathbf{H}^{-1}(\delta\mathbf{H})'\,,$$

where the property $\delta\mathbf{H}^{-1} = -\delta\widehat{\mathbf{h}}\,\mathbf{H}^{-1}$ can be deduced from $\delta(\mathbf{H}\mathbf{H}^{-1}) = \mathbf{0}$ and, noting that $\mathbf{H}'$ and $\delta\mathbf{H}$ are in the same linear space $T_{\mathbf{H}}SE(3)$, the operators $(\cdot)'$ and $\delta(\cdot)$ do commute.

Consequently, the above equation further simplifies as follows:

$$\begin{aligned}\delta\widehat{\mathbf{d}} &= -\delta\widehat{\mathbf{h}}\,\widehat{\mathbf{d}} + \mathbf{H}^{-1}(\mathbf{H}\delta\widehat{\mathbf{h}})' = -\delta\widehat{\mathbf{h}}\,\widehat{\mathbf{d}} + \mathbf{H}^{-1}\mathbf{H}'\delta\widehat{\mathbf{h}} + \mathbf{H}^{-1}\mathbf{H}\,\delta\widehat{\mathbf{h}}' \\ &= -\delta\widehat{\mathbf{h}}\,\widehat{\mathbf{d}} + \mathbf{H}^{-1}\mathbf{H}\widehat{\mathbf{d}}\,\delta\widehat{\mathbf{h}} + \delta\widehat{\mathbf{h}}' = \delta\widehat{\mathbf{h}}' + \widehat{\mathbf{d}}\,\delta\widehat{\mathbf{h}} - \delta\widehat{\mathbf{h}}\,\widehat{\mathbf{d}}\,,\end{aligned}$$

whence, recalling that the matrix commutator is the Lie bracket for a matrix Lie algebra (cf. Definition B.56), one infers

$$\delta\widehat{\mathbf{d}} = \delta\widehat{\mathbf{h}}' + [\widehat{\mathbf{d}}, \delta\widehat{\mathbf{h}}]\,; \tag{5.59}$$

using the equivalence (4.87), the previous relation in vector form reads

$$\delta\mathbf{d} = \delta\mathbf{h}' + \mathrm{ad}_{\mathbf{d}}\,\delta\mathbf{h}\,, \tag{5.60}$$

and represents the variation $\delta\mathbf{d}(\mu)$ associated with the field $\delta\mathbf{h}(\mu)$ of the screw variation compatible with the virtual displacement field $\delta\mathbf{H}(\mu) \in T_{\mathbf{H}(\mu)}SE(3)$.

5.4.3 Static Equilibrium

The principle of virtual work expressed by (5.57) for a single framed point of $\mathscr{E}$, can be extended to the whole beam.

Actually, intending the configuration of the beam in the physical space as a framed curve, the virtual work of $\mathscr{B}_t$ can be expressed as

$$\delta W = \int_{\mathscr{B}_t} \mathrm{d}(\delta W)\,,$$

where $\mathsf{d}(\delta W) = \delta W(\mu + \mathsf{d}\mu) - \delta W(\mu)$ represents the virtual work relevant to the infinitesimal element of the beam, between the cross-sections at μ and at $\mu + \mathsf{d}\mu$.

Moreover, as usual in one-dimensional beam theory, the virtual work is split in the external contribution and the internal one. Actually, consistently with the representation of forces as covectors of virtual displacement screws, that holds true by virtue of the identification (5.56), the external virtual work is associated with the external forces acting on the beam, and the same applies for the internal contribution.

Specifically, the *external load* is a screw vector field $\mu \mapsto \mathbf{g}(\mu)$ acting on a compatible virtual displacement screw field $\mu \mapsto \delta\mathbf{h}(\mu)$:

$$\mathsf{d}(\delta W^{\mathsf{ext}}) = \mathbf{g}^*(\mu) \cdot \delta\mathbf{h}(\mu)\mathsf{d}\mu\,, \tag{5.61}$$

and, considering also the *external forces* $\mathbf{f}_{\mathsf{A}}$ and $\mathbf{f}_{\mathsf{B}}$, applied as wrenches at the framed points $\mathscr{F}_{\mathsf{A}}$ and $\mathscr{F}_{\mathsf{B}}$ relevant to the extremities of the beam, the *external virtual work* of the beam reads

$$\delta W^{\mathsf{ext}} = \mathbf{f}^*_{\mathsf{A}} \cdot \delta\mathbf{h}_{\mathsf{A}} + \mathbf{f}^*_{\mathsf{B}} \cdot \delta\mathbf{h}_{\mathsf{B}} + \int_0^1 \mathbf{g}^* \cdot \delta\mathbf{h}\,\mathsf{d}\mu\,, \tag{5.62}$$

where the dependence of the screw in the integral is intended upon the material abscissa μ.

In parallel, with the aim to represent the static interaction between two subsequent beam cross-sections, the *internal forces* are introduced as the screw field $\mu \mapsto \mathbf{l}(\mu)$ acting on the virtual variation of the deformation screw field $\mu \mapsto \delta\mathbf{d}(\mu)$, compatible via (5.60) with the virtual displacement screw:

$$\mathsf{d}(\delta W^{\mathsf{int}}) = \mathbf{l}^*(\mu) \cdot \delta\mathbf{d}(\mu)\mathsf{d}\mu\,; \tag{5.63}$$

thus, omitting the explicit dependence on μ, the *internal virtual work* of the beam turns out to be

$$\delta W^{\mathsf{int}} = \int_0^1 \mathbf{l}^* \cdot \delta\mathbf{d}\,\mathsf{d}\mu\,, \tag{5.64}$$

and the virtual work of the beam is

$$\delta W = \delta W^{\mathsf{ext}} + \delta W^{\mathsf{int}}\,. \tag{5.65}$$

In accordance with the principle of virtual work, the configuration $\mathscr{B}_t$ of the beam is an equilibrium configuration if the virtual work vanishes for all compatible virtual displacement $\delta\mathbf{h}$:

$$\delta W(\delta\mathbf{h}) = \delta W^{\mathsf{ext}}(\delta\mathbf{h}) + \delta W^{\mathsf{int}}(\delta\mathbf{h}) = 0\,,\ \ \forall\,\delta\mathbf{h} = \delta\mathbf{h}(\mu)\,. \tag{5.66}$$

Let us point out that the required compatibility of $\delta\mathbf{h}(\mu)$ consists in assuring that the relevant screw field $\delta\widehat{\mathbf{h}}(\mu)$ satisfies (5.58). In addition, it is required that the virtual screw of A and B in (5.62) do satisfy $\delta\mathbf{h}_{\mathsf{A}} = \delta\mathbf{h}(0)$ and $\delta\mathbf{h}_{\mathsf{B}} = \delta\mathbf{h}(1)$, as well as the virtual increment of the deformation vector $\delta\mathbf{d}$ appearing in (5.64) fulfills the compatibility condition (5.60).

Specifically, under such compatibility condition, the internal virtual work becomes

$$\begin{aligned}\delta W^{\mathsf{int}} &= \int_0^1 \mathbf{l}^* \cdot \delta\mathbf{h}'\,\mathrm{d}\mu + \int_0^1 \mathbf{l}^* \cdot \mathrm{ad}_{\mathbf{d}}\,\delta\mathbf{h}\,\mathrm{d}\mu \\ &= [\mathbf{l}^* \cdot \delta\mathbf{h}]_0^1 - \int_0^1 \mathbf{l}^{*\prime} \cdot \delta\mathbf{h}\,\mathrm{d}\mu + \int_0^1 \mathrm{ad}_{\mathbf{d}}^{\mathsf{T}}\,\mathbf{l}^* \cdot \delta\mathbf{h}\,\mathrm{d}\mu \\ &= \mathbf{l}^*(1)\cdot\delta\mathbf{h}_{\mathsf{A}} - \mathbf{l}^*(0)\cdot\delta\mathbf{h}_{\mathsf{B}} + \int_0^1 (\mathrm{ad}_{\mathbf{d}}^{\mathsf{T}}\,\mathbf{l}^* - \mathbf{l}^{*\prime})\cdot\delta\mathbf{h}\,\mathrm{d}\mu\,.\end{aligned}$$

so that the principle of virtual work specializes to

$$\begin{aligned}\delta W &= \big(\mathbf{f}_{\mathsf{A}}^* + \mathbf{l}^*(1)\big)\cdot\delta\mathbf{h}_{\mathsf{A}} + \big(\mathbf{f}_{\mathsf{B}}^* - \mathbf{l}^*(0)\big)\cdot\delta\mathbf{h}_{\mathsf{B}} \\ &\quad + \int_0^1 (\mathbf{g}^* + \mathrm{ad}_{\mathbf{d}}^{\mathsf{T}}\,\mathbf{l}^* - \mathbf{l}^{*\prime})\cdot\delta\mathbf{h}\,\mathrm{d}\mu = 0\,,\ \ \forall\,\delta\mathbf{h} = \delta\mathbf{h}(\mu)\,.\end{aligned} \tag{5.67}$$

The arbitrariness of the screw vector field $\delta\mathbf{h}(\mu)$, and using the screw representation of the forces in place of the coscrew one, implies the differential equilibrium equation of the beam, i.e.

$$\mathbf{g}(\mu) + \mathrm{ad}_{\mathbf{d}(\mu)}\,\mathbf{l}(\mu) - \mathbf{l}'(\mu) = \mathbf{o}\,, \tag{5.68}$$

along with the following boundary conditions:

$$\mathbf{l}(1) = -\mathbf{f}_{\mathsf{A}}\,, \quad \mathbf{l}(0) = \mathbf{f}_{\mathsf{B}}\,. \tag{5.69}$$

5.4.4 Internal Forces

The internal force wrench $\mathbf{l}(\mu)$, relevant to the beam cross-section at μ, has been introduced in (5.63) in terms of the virtual work $\mathrm{d}(\delta W^{\mathsf{int}})$ provided by its action on the screw variation $\delta\mathbf{d}(\mu)\mathrm{d}\mu$.

However, a complete characterization of the beam internal forces requires to specify some explicit relation with the strain parameters associated with the cross-section.

At this stage, it is useful to recall that, within the framework of an induced beam theory, beam kinematics has been derived by specializing a three-dimensional solid body through the progressive introduction of specific geometric assumptions and constraints about the position field. Hence, it would be desirable to do the same for the statics by establishing an equivalence between the stress distribution relevant to the beam cross-section and the internal forces of the one-dimensional model.

In this respect, satisfactory characterizations of a beam cross-section are available from linear beam theories so that one asks if it is possible, and to what extent it is legitimate, to exploit the results derived from the cross-section analysis based on linear models, in order to describe the elastic behavior of a beam undergoing large displacements.

Actually, a first issue concerns the definition of suitable strain measures for the beam cross-section that can somehow be compared to the relevant strain measures usually adopted in linear analysis.

As a matter of fact, in linear models, strain measures relevant to a current configuration, have the mechanical meaning of variation of the beam deformation status, with respect to a reference configuration, per unitary length.

However, while the hypothesis of small displacements and displacement gradients, characterizing any linear beam model, allows one to recognize strain parameters unambiguously, the extension of the same reasoning to a beam undergoing large displacements is not straightforward.

With the aim to introduce a proper set of strain parameters, let us recall from (5.38) that the screw vector $\mathbf{d}(\mu)\mathrm{d}\mu$ represents the relative configuration of the cross-section at $\mu + \mathrm{d}\mu$ with respect to the one at μ. Accordingly, the quantity $\delta\mathbf{d}(\mu)\mathrm{d}\mu$ has the meaning of infinitesimal variation of the relative cross-section configuration, when an infinitesimal displacement field is applied at the beam configuration $\mathscr{B}_t$. At this point, the cross-section strains resulting from $\delta\mathbf{d}(\mu)\mathrm{d}\mu$ can be obtained by referring such a variation to the length of the beam incremental element.

However, differently from linear beam theories, the length $\mathrm{d}s$ of such an element itself depends upon the configuration $\mathscr{B}_t$. Specifically, the length of the infinitesimal beam element is given by the differential of the spatial abscissa $s(\mu)$ with respect to μ, and can be expressed as $\mathrm{d}s = \lambda(\mu)\mathrm{d}\mu$, with λ defined by (5.11).

Accordingly, the ratio between the variation $\delta\mathbf{d}(t,\mu)\mathrm{d}\mu$ and the associated arc length $\mathrm{d}s = \lambda(t,\mu)\mathrm{d}\mu$, with the explicit dependence on time t, is assumed to represent the infinitesimal variation $\delta\boldsymbol{\chi}$ of the cross-section strain screw:

$$\delta\boldsymbol{\chi}(t,\mu) = \frac{\delta\mathbf{d}(t,\mu)}{\lambda(t,\mu)} \,. \tag{5.70}$$

In practice, the screw $\delta\boldsymbol{\chi}(t,\mu)$ has the function to represent the linearized strain measures of the cross-section when the beam configuration $\mathscr{B}_t$ is deformed slightly enough to remain close to $\mathscr{B}_t$ itself. As such, the incremental screw $\delta\boldsymbol{\chi}$ plays the same role with respect to the configuration $\mathscr{B}_t$ at time t, as the strain measures of linear beam theory with respect to the reference configuration.

With these specifications, if one refers to the infinitesimal variation $\delta\boldsymbol{\chi}(t,\mu)$, the results emerging from the analysis of classical linear beam models appear appropriate to be extended to the context of a beam subject to large displacements.

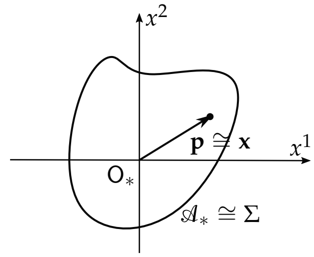


Figure 5.6: Beam cross-section representation.

Cross-Section Stiffness Matrix

It has been shown in Section 5.2.1 how the material representation of any beam-cross section $\mathscr{S}_\mu$ is a domain $\Sigma \subset \mathbb{R}^2$. Moreover, by virtue of the internal constraints introduced in Section 5.2.2, any spatial configuration $\mathscr{A}_s$ of the cross-section $\mathscr{S}_\mu$ is obtained by a proper rigid motion applied at a plane subset $\mathscr{A}_*$ of the physical space $\mathscr{E}$, the relation between $\Sigma \subset \mathbb{R}^2$ and $\mathscr{A}_* \subset \pi_*$ being provided by the map (5.19), or equivalently by its extension $\varphi_* \colon \mathbb{R}^2 \to \pi_*$ defined by (5.20).

Let us assume that the reference system $\mathscr{F}_*$ attached to $\mathscr{A}_*$, made of the origin O_* and the orthonormal basis $\mathscr{B}_* = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, is such that the plane π_* containing $\mathscr{A}_*$ is spanned by $\mathbf{e}_1$ and $\mathbf{e}_2$.

Concurrently, since the material representation Σ of the beam cross-section is arbitrary, it is convenient to choose a local chart such that the pair (x^1, x^2) defining $\mathbf{x} \in \mathbb{R}^2$ also provides the coordinates of the relevant point $\mathsf{P}_* \in \pi_*$ with respect to $\mathscr{F}_*$.

Explicitly, as shown in Figure 5.6, given $\mathbf{x} = (x^1, x^2)$ and considering $\mathsf{P}_* = \varphi_*(\mathbf{x})$, the position vector of P_* in the frame $\mathscr{F}_*$ is $\mathbf{p} = x^1\mathbf{e}_1 + x^2\mathbf{e}_2$.

Moreover, since the strain variation $\delta\boldsymbol{\chi}$, arising from the deformation screw vector $\mathbf{d}$ through (5.70), is referred to the cross-section configuration $\mathscr{A}_s$, the representation of the relevant rotation and translation components, with respect to the local frame $\mathscr{F}_s$, turn out to be

$$\delta\boldsymbol{\chi}_{\mathsf{r}} = \begin{bmatrix} \delta\boldsymbol{\chi}_{\mathsf{ben}} \\ \delta\chi_{\mathsf{tor}} \end{bmatrix}, \quad \delta\boldsymbol{\chi}_{\mathsf{t}} = \begin{bmatrix} \delta\boldsymbol{\chi}_{\mathsf{sh}} \\ \delta\chi_{\mathsf{ax}} \end{bmatrix}, \tag{5.71}$$

where the subscripts 'ben' and 'tor' for the entries of the rotation component $\delta\boldsymbol{\chi}_{\mathsf{r}}$ refer to the cross-section strains associated with bending and torsion, respectively, as well as the subscripts 'sh' and 'ax' for the components of the translation vector $\delta\boldsymbol{\chi}_{\mathsf{t}}$ identify the strains induced by the shear and axial effects, respectively.

A comprehensive description of the effects associated with the above-mentioned load conditions in linear elastic beams can be found in Paradiso et al. (2019), as well as the resulting beam cross-section characterization is reported in Paradiso et al. (2021, 2020).

Thus, denoting by $\mathbf{K}_\Sigma$ the *cross-section stiffness matrix*, consistently with the meaning given to the strain screw variation $\delta\boldsymbol{\chi}$, the infinitesimal variation of the cross-section internal forces is expressed as

$$\delta\mathbf{l} = \mathbf{K}_\Sigma \delta\boldsymbol{\chi}\,, \tag{5.72}$$

whence the internal force vector $\mathbf{l}(t,\mu)$, relevant to the cross-section at μ for the beam configuration $\mathscr{B}_t$, becomes

$$\mathbf{l}(t,\mu) = \int_{t_0}^{t} \mathbf{K}_\Sigma \delta\boldsymbol{\chi}(\bar{t},\mu)\,. \tag{5.73}$$

where $\bar{t} \in [t_0, t]$ identifies a configuration $\mathscr{B}_{\bar{t}}$ of the beam between the *reference configuration* $\mathscr{B}_0 = \mathscr{C}(t_0, \mathscr{B})$ and the current one $\mathscr{B}_t$.

Please notice that the reference configuration represents a condition where the internal forces do vanish:

$$\mathbf{l}(t_0,\mu) = \mathbf{o}\,,\ \ \forall\, \mu \in [0,1]\,.$$

Hence, $\mathscr{B}_0$ defines a configuration where no external loads are applied, or, more generally, a configuration where the equilibrium is fulfilled, providing the starting point for any further variation.

Moreover, in accordance with its derivation from the linear beam theory, the stiffness matrix $\mathbf{K}_\Sigma$ is assumed to not depend on the specific beam configuration $\mathscr{B}_t$, and the wrench vector in (5.73) becomes

$$\mathbf{l}(t,\mu) = \mathbf{K}_\Sigma\, \boldsymbol{\chi}(t,\mu)\,, \tag{5.74}$$

where the meaning of the screw vector $\boldsymbol{\chi}(t, \mu)$ is to represent the cumulative strain of the cross-section at μ between the reference configuration $\mathscr{B}_0$ and the current one $\mathscr{B}_t$.

Taking also account of (5.70), the cumulative strain vector becomes

$$\boldsymbol{\chi}(t, \mu) = \int_{t_0}^{t} \delta\boldsymbol{\chi}(\bar{t}, \mu) = \int_{t_0}^{t} \frac{\delta\mathbf{d}(\bar{t}, \mu)}{\lambda(\bar{t}, \mu)}\,, \tag{5.75}$$

an expression that will be exploited in numerical simulations.

Chapter 6

Finite Element Formulation

A Finite Element approach to the solution of the beam mechanical problem is presented.

In agreement with such a method, the kinematics of a beam element is approximated by means of shape functions interpolating the configurations $\mathbf{H}_\mathsf{A}$ and $\mathbf{H}_\mathsf{B}$ of the element's nodes, say A and B. In particular, consistently with the geometrically exact formulation, the interpolating functions are defined within the context of the Lie group $SE(3)$, so that they are introduced as curves on $SE(3)$.

Then, the equilibrium of the beam element is formulated in weak form by considering the virtual displacement field of the finite element, that specializes the field introduced in Section 5.4.2.

6.1 Shape Functions

The compatibility condition governing the beam kinematics is given by (5.39) and has been derived in Section 5.3.1 with the geometric meaning of establishing the relation between the local variation $\mathbf{H}'(\mu)$ of the configuration matrix $\mathbf{H}(\mu)$ and the deformation screw $\widehat{\mathbf{d}}(\mu)$.

Moreover, in Section 5.3.2 it has shown how the same equation actually defines a vector field $\boldsymbol{V}$ on the Lie group $SE(3)$ whose value at $\mathbf{H}(\mu)$ is $\boldsymbol{V}_{\mathbf{H}(\mu)} = \mathbf{H}'(\mu)$ and specializes to $\boldsymbol{V}_\mathbf{I} = \widehat{\mathbf{d}}(\mu)$ at the identity.

In order to define the shape functions of the beam element, it is useful to recall that $\boldsymbol{V}$, assumed to be a left-invariant vector field of $SE(3)$, is completely identified by its value at the identity (cf. Section 2.4.1). In such a case the value $\boldsymbol{V}_{\mathbf{I}}$ is uniquely defined, that is the screw $\widehat{\mathbf{d}} \in \mathfrak{se}(3)$ does not depend on μ, and, by Proposition 2.35, the integral curve of $\boldsymbol{V}$ starting at the identity with velocity $\boldsymbol{V}_{\mathbf{I}} = \widehat{\mathbf{d}}$ is the one-parameter subgroup generated by $\boldsymbol{V}$.

In addition, recalling also Proposition 2.37, the integral curve of $\boldsymbol{V}$ starting at the identity is provided by the exponential map in the form

$$\gamma_{\mathbf{I}}(\mu) = \exp(\mu\widehat{\mathbf{d}})\,,$$

where the identification $\boldsymbol{V} \cong \boldsymbol{V}_{\mathbf{I}} = \widehat{\mathbf{d}}$ has been exploited.

The benefit of assuming $\boldsymbol{V}$ as a left-invariant vector field is that any integral curve can be obtained from $\gamma_{\mathbf{I}}$ by left-translation. Consequently, the integral curve of $\boldsymbol{V}$ starting at $\mathbf{H}_{\mathsf{A}}$, which satisfies the condition $\mathbf{H}(0) = \mathbf{H}_{\mathsf{A}}$, provides the shape functions for the beam element in the form

$$\mathbf{H}(\mu) = \mathbf{H}_{\mathsf{A}} \exp(\mu\widehat{\mathbf{d}})\,, \tag{6.1}$$

where the exponential map for $SE(3)$ is explicitly provided by (4.80).

Moreover, since it is also required $\mathbf{H}(1) = \mathbf{H}_{\mathsf{B}}$, one can promptly verify that $\exp(\widehat{\mathbf{d}}) = \mathbf{H}_{\mathsf{A}}^{-1}\mathbf{H}_{\mathsf{B}}$, that is

$$\widehat{\mathbf{d}} = \log(\mathbf{H}_{\mathsf{A}}^{-1}\mathbf{H}_{\mathsf{B}})\,, \tag{6.2}$$

with the logarithm map of $SE(3)$ defined by (4.82).

Please recall that the deformation screw $\widehat{\mathbf{d}}(\mu)$ has been introduced in Section 5.3.1 with kinematic meaning of the local change of configuration between two subsequent beam cross-sections. Hence, the assumption of $\boldsymbol{V}$ to be a left-invariant vector field on $SE(3)$, implies that such a deformation is uniform along the beam axis.

Moreover, invoking (6.2), the deformation screw $\widehat{\mathbf{d}}$ defines the motion between the beam cross-sections $\mathscr{A}_{\mathsf{A}}$ to $\mathscr{A}_{\mathsf{B}}$, identified with the configuration matrices $\mathbf{H}_{\mathsf{A}}$ and $\mathbf{H}_{\mathsf{B}}$, respectively. This also means that the beam deformation is provided by the average of the initial and final cross-section configurations, in a sense consistent with the group structure of the configuration space $SE(3)$.

In order to provide a weak form for the equilibrium of the beam element, consistently with the standard approach of the finite element method (Zienkiewicz et al., 2013), let us consider a virtual displacement of the configuration matrices $\mathbf{H}_\mathsf{A}$ and $\mathbf{H}_\mathsf{B}$:

$$\delta\mathbf{H}_\mathsf{A} = \mathbf{H}_\mathsf{A}\delta\widehat{\mathbf{h}}_\mathsf{A}\,, \qquad \delta\mathbf{H}_\mathsf{B} = \mathbf{H}_\mathsf{B}\delta\widehat{\mathbf{h}}_\mathsf{B}\,, \tag{6.3}$$

where $\delta\widehat{\mathbf{h}}_\mathsf{A}$ and $\delta\widehat{\mathbf{h}}_\mathsf{B}$ are the variational screws of $\mathfrak{se}(3)$ associated with the six-dimensional vectors $\delta\mathbf{h}_\mathsf{A}$ and $\delta\mathbf{h}_\mathsf{B}$ by the map (4.78); for convenience, they are assembled in a unique variational vector $\delta\mathbf{h}_\mathsf{AB} \in \mathbb{R}^{12}$:

$$\delta\mathbf{h}_\mathsf{AB} = \begin{bmatrix} \delta\mathbf{h}_\mathsf{A} \\ \delta\mathbf{h}_\mathsf{B} \end{bmatrix}. \tag{6.4}$$

Hence, with the aim to provide a weak form of the equilibrium equation (5.68), let us express the vector variation $\delta\mathbf{d}$, associated with the variation $\delta\widehat{\mathbf{d}}$ of the deformation screw given by (6.2), as a function of $\delta\mathbf{h}_\mathsf{AB}$:

$$\delta\mathbf{d} = \mathbf{B}_\mathbf{d}\delta\mathbf{h}_\mathsf{AB}\,, \tag{6.5}$$

where $\mathbf{B}_\mathbf{d}$ is the *element deformation operator*.

Similarly, the infinitesimal variation $\delta\mathbf{h}(\mu)$ relevant to the beam virtual displacement screw $\delta\widehat{\mathbf{h}}(\mu) = \mathbf{H}(\mu)^{-1}\delta\mathbf{H}(\mu)$, can be expressed by means of the *element displacement operator*, denoted as $\mathbf{P}_\mathbf{d}(\mu)$, as

$$\delta\mathbf{h}(\mu) = \mathbf{P}_\mathbf{d}(\mu)\delta\mathbf{h}_\mathsf{AB}\,. \tag{6.6}$$

The explicit expression of both the deformation and the displacement operators is derived in the next section.

6.2 Element Operators of Deformation and Displacement

Let us preliminarily observe that, given $\delta\widehat{\mathbf{h}} \in \mathfrak{se}(3)$ such that $\delta\mathbf{H} = \mathbf{H}\delta\widehat{\mathbf{h}} \in T_{\mathbf{H}}SE(3)$, the infinitesimal variation of the inverse element $\mathbf{H}^{-1} \in SE(3)$ satisfies

$$\begin{aligned}\delta(\mathbf{H}^{-1}\mathbf{H}) &= \delta\mathbf{H}^{-1}\mathbf{H} + \mathbf{H}^{-1}\delta\mathbf{H} = \delta\mathbf{H}^{-1}\mathbf{H} + \mathbf{H}^{-1}\mathbf{H}\delta\widehat{\mathbf{h}} \\ &= \delta\mathbf{H}^{-1}\mathbf{H} + \delta\widehat{\mathbf{h}} = \mathbf{0}\,,\end{aligned}$$

whence

$$\delta\mathbf{H}^{-1} = -\delta\widehat{\mathbf{h}}\mathbf{H}^{-1}\,. \tag{6.7}$$

Now, consider the exponential of the screw $\widehat{\mathbf{d}}$, given by (6.2), and apply an infinitesimal variation:

$$\begin{aligned}\delta\big(\exp(\widehat{\mathbf{d}})\big) &= \delta(\mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B}) = \delta\mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B} + \mathbf{H}_\mathsf{A}^{-1}\delta\mathbf{H}_\mathsf{B} \\ &= -\delta\widehat{\mathbf{h}}_\mathsf{A}\mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B} + \mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B}\delta\widehat{\mathbf{h}}_\mathsf{B} \\ &= \mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B}\big(-(\mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B})^{-1}\delta\widehat{\mathbf{h}}_\mathsf{A}(\mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B}) + \delta\widehat{\mathbf{h}}_\mathsf{B}\big)\,,\end{aligned}$$

where, accordingly with (2.50), $(\mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B})^{-1}\delta\widehat{\mathbf{h}}_\mathsf{A}(\mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B})$ is the adjoint representation of $(\mathbf{H}_\mathsf{A}^{-1}\mathbf{H}_\mathsf{B})^{-1}$ applied at $\delta\widehat{\mathbf{h}}_\mathsf{A}$.

Since within $SE(3)$ the isomorphism $\mathfrak{se}(3) \cong \mathbb{R}^6$ allows one to write the adjoint map in the form (4.84), the above relation can be expressed as

$$\delta\big(\exp(\widehat{\mathbf{d}})\big) = \exp(\widehat{\mathbf{d}})\big(-\operatorname{Ad}_{\exp(-\widehat{\mathbf{d}})}\delta\mathbf{h}_\mathsf{A} + \delta\mathbf{h}_\mathsf{B}\big)^{\widehat{}}\,.$$

At the same time, the variation of $\exp(\widehat{\mathbf{d}})$ can be expressed by means of the differential dexp of $SE(3)$ given by (4.89):

$$\delta\big(\exp(\widehat{\mathbf{d}})\big) = \operatorname{dexp}|_{\mathbf{d}}(\delta\mathbf{d}) = \exp(\widehat{\mathbf{d}})(\mathbf{T}_\mathbf{d}\delta\mathbf{d})^{\widehat{}}\,, \tag{6.8}$$

so that, by comparison, one finds

$$\mathbf{T}_\mathbf{d}\delta\mathbf{d} = -\operatorname{Ad}_{\exp(-\widehat{\mathbf{d}})}\delta\mathbf{h}_\mathsf{A} + \delta\mathbf{h}_\mathsf{B}\,;$$

upon assembling $\delta\mathbf{h}_{\mathsf{A}}$ and $\delta\mathbf{h}_{\mathsf{B}}$ in $\delta\mathbf{h}_{\mathsf{AB}}$, the previous relation reads

$$\delta\mathbf{d} = \begin{bmatrix} -\mathbf{T}_{\mathbf{d}}^{-1}\,\mathrm{Ad}_{\exp(-\widehat{\mathbf{d}})} & \mathbf{T}_{\mathbf{d}}^{-1} \end{bmatrix} \delta\mathbf{h}_{\mathsf{AB}}\,.$$

Finally, using the second identity in (4.115), one gets

$$\delta\mathbf{d} = \mathbf{B}_{\mathbf{d}}\delta\mathbf{h}_{\mathsf{AB}} = \begin{bmatrix} -\mathbf{T}_{-\mathbf{d}}^{-1} & \mathbf{T}_{\mathbf{d}}^{-1} \end{bmatrix} \delta\mathbf{h}_{\mathsf{AB}}\,, \tag{6.9}$$

where the element deformation operator explicitly reads

$$\mathbf{B}_{\mathbf{d}} = \begin{bmatrix} -\mathbf{T}_{-\mathbf{d}}^{-1} & \mathbf{T}_{\mathbf{d}}^{-1} \end{bmatrix}. \tag{6.10}$$

A similar procedure can be applied to derive the displacement operator. Specifically let us consider the virtual variation of $\mathbf{H}(\mu)$ expressed by (6.1):

$$\begin{aligned}
\delta\mathbf{H}(\mu) &= \delta\big(\mathbf{H}_{\mathsf{A}}\exp(\mu\widehat{\mathbf{d}})\big) = \delta\mathbf{H}_{\mathsf{A}}\exp(\mu\widehat{\mathbf{d}}) + \mathbf{H}_{\mathsf{A}}\delta\big(\exp(\mu\widehat{\mathbf{d}})\big)\\
&= \mathbf{H}_{\mathsf{A}}\delta\widehat{\mathbf{h}}_{\mathsf{A}}\exp(\mu\widehat{\mathbf{d}}) + \mathbf{H}_{\mathsf{A}}\,\mathrm{dexp}|_{\mu\mathbf{d}}(\mu\delta\mathbf{d}) =\\
&= \mathbf{H}_{\mathsf{A}}\exp(\mu\widehat{\mathbf{d}})\big(\exp(-\mu\widehat{\mathbf{d}})\delta\widehat{\mathbf{h}}_{\mathsf{A}}\exp(\mu\widehat{\mathbf{d}}) + (\mu\mathbf{T}_{\mu\mathbf{d}}\delta\mathbf{d})^{\widehat{}}\big)\,,
\end{aligned}$$

where the variation $\delta\big(\exp(\mu\widehat{\mathbf{d}})\big)$ has been evaluated as in (6.8) considering the screw $\mu\widehat{\mathbf{d}}$.

Moreover, recognizing the adjoint representation of $\exp(-\mu\widehat{\mathbf{d}})$ applied at $\delta\widehat{\mathbf{h}}_{\mathsf{A}}$, one also has

$$\delta\mathbf{H}(\mu) = \mathbf{H}(\mu)\big(\mathrm{Ad}_{\exp(-\mu\widehat{\mathbf{d}})}\,\delta\mathbf{h}_{\mathsf{A}} + \mu\mathbf{T}_{\mu\mathbf{d}}\delta\mathbf{d}\big)^{\widehat{}}\,,$$

and, since $\delta\mathbf{H}(\mu) = \mathbf{H}(\mu)\delta\widehat{\mathbf{h}}(\mu)$, by comparison, one infers

$$\delta\mathbf{h} = \mathrm{Ad}_{\exp(-\mu\widehat{\mathbf{d}})}\,\delta\mathbf{h}_{\mathsf{A}} + \mu\mathbf{T}_{\mu\mathbf{d}}\delta\mathbf{d}\,.$$

Furthermore, exploiting the explicit form of $\delta\mathbf{d}$ given by (6.9), the following relation is finally obtained

$$\begin{aligned}
\delta\mathbf{h}(\mu) &= \mathbf{P}_{\mathbf{d}}(\mu)\delta\mathbf{h}_{\mathsf{AB}}\\
&= \begin{bmatrix} \mathrm{Ad}_{\exp(-\mu\widehat{\mathbf{d}})} - \mu\mathbf{T}_{\mu\mathbf{d}}\mathbf{T}_{-\mathbf{d}}^{-1} & \mu\mathbf{T}_{\mu\mathbf{d}}\mathbf{T}_{\mathbf{d}}^{-1} \end{bmatrix} \delta\mathbf{h}_{\mathsf{AB}}\,,
\end{aligned} \tag{6.11}$$

whence the expression of the element displacement operator is

$$\mathbf{P}_{\mathbf{d}}(\mu) = \left[\mathrm{Ad}_{\exp(-\mu\widehat{\mathbf{d}})} \; -\mu\mathbf{T}_{\mu\mathbf{d}}\mathbf{T}_{-\mathbf{d}}^{-1} \quad \mu\mathbf{T}_{\mu\mathbf{d}}\mathbf{T}_{\mathbf{d}}^{-1}\right]. \tag{6.12}$$

6.3 Element Forces

Adopting the usual procedure of the finite element method, both the external distributed load $\mathbf{g}(\mu)$ and the internal forces $\mathbf{l}(\mu)$ of the beam element are expressed in terms of equivalent nodal forces, denoted as $\mathbf{f}_{\mathsf{AB}}^{\mathsf{ext}} \in \mathbb{R}^{12}$ and $\mathbf{f}_{\mathsf{AB}}^{\mathsf{int}} \in \mathbb{R}^{12}$, respectively.

Please recall that, from the algebraic point of view, the vector $\mathbf{g}(\mu) \in \mathbb{R}^6$ represents a coscrew $\widehat{\mathbf{g}}^*(\mu) \in \mathfrak{se}^*(3)$, and the action on the virtual displacement vector $\delta\mathbf{h}(\mu) \in \mathbb{R}^6$, associated with $\delta\widehat{\mathbf{h}}(\mu) \in \mathfrak{se}(3)$, applies as in (5.61):

$$\delta W^{\mathsf{ext}} = \int_0^1 \mathbf{g}^* \cdot \delta\mathbf{h}\,\mathrm{d}\mu = \int_0^1 \mathbf{g}(\delta\mathbf{h})\,\mathrm{d}\mu\,,$$

where the explicit dependence on μ has been omitted.

In addition, using (6.11), the above relation becomes

$$\delta W^{\mathsf{ext}}(\delta\mathbf{h}_{\mathsf{AB}}) = \int_0^1 \mathbf{g}(\mathbf{P}_{\mathbf{d}}\delta\mathbf{h}_{\mathsf{AB}})\,\mathrm{d}\mu = \int_0^1 (\mathbf{P}_{\mathbf{d}}^{\mathsf{T}}\mathbf{g})(\delta\mathbf{h}_{\mathsf{AB}})\,\mathrm{d}\mu\,,$$

where, invoking Definition B.26 of dual map, the transpose of the displacement operator $\mathbf{P}_{\mathbf{d}}^{\mathsf{T}}(\mu)$ is a linear map from $\mathbb{R}^6 \cong \mathfrak{se}^*(3)$ to $\mathbb{R}^{12} \cong \mathfrak{se}^*(3) \times \mathfrak{se}^*(3)$.

Hence, considering $\delta\mathbf{h}_{\mathsf{AB}}^* \cong (\delta\widehat{\mathbf{h}}_{\mathsf{A}}^*, \delta\widehat{\mathbf{h}}_{\mathsf{B}}^*) \in \mathfrak{se}^*(3) \times \mathfrak{se}^*(3)$, that is

$$\delta\mathbf{h}_{\mathsf{AB}}^* = \begin{bmatrix} \delta\mathbf{h}_{\mathsf{A}}^* \\ \delta\mathbf{h}_{\mathsf{B}}^* \end{bmatrix}, \tag{6.13}$$

the external virtual work can be written as

$$\delta W^{\mathsf{ext}}(\delta\mathbf{h}_{\mathsf{AB}}) = \int_0^1 \mathbf{P}_{\mathbf{d}}^{\mathsf{T}}\mathbf{g}\,\mathrm{d}\mu \cdot \delta\mathbf{h}_{\mathsf{AB}}^* = \mathbf{f}_{\mathsf{AB}}^{\mathsf{ext}} \cdot \delta\mathbf{h}_{\mathsf{AB}}^*\,,$$

where the element external forces have been gathered in $\mathbf{f}_{\mathsf{AB}}^{\mathsf{ext}} \in \mathbb{R}^{12}$:

$$\mathbf{f}_{\mathsf{AB}}^{\mathsf{ext}} = \int_0^1 \mathbf{P}_{\mathbf{d}}^{\mathsf{T}}\mathbf{g}\,\mathrm{d}\mu\,, \tag{6.14}$$

whose components are

$$\mathbf{f}^{\mathsf{ext}}_{\mathsf{AB}} = \begin{bmatrix} \mathbf{f}^{\mathsf{ext}}_{\mathsf{A}} \\ \mathbf{f}^{\mathsf{ext}}_{\mathsf{B}} \end{bmatrix}.$$

It is worth remarking that $\mathbf{f}^{\mathsf{ext}}_{\mathsf{A}}$ and $\mathbf{f}^{\mathsf{ext}}_{\mathsf{B}}$ define the wrenches acting on the virtual displacement screws $\delta\mathbf{h}_{\mathsf{A}}$ and $\delta\mathbf{h}_{\mathsf{B}}$, respectively, in such a way that the virtual work $\mathbf{f}^{\mathsf{ext}}_{\mathsf{A}}(\delta\mathbf{h}_{\mathsf{A}}) + \mathbf{f}^{\mathsf{ext}}_{\mathsf{B}}(\delta\mathbf{h}_{\mathsf{B}}) = \mathbf{f}^{\mathsf{ext}}_{\mathsf{A}} \cdot \delta\mathbf{h}^*_{\mathsf{A}} + \mathbf{f}^{\mathsf{ext}}_{\mathsf{B}} \cdot \delta\mathbf{h}^*_{\mathsf{B}}$ equals the virtual work δW^{ext} of the beam element due to the external load wrench $\mathbf{g}(\mu)$.

A similar procedure can be applied to evaluate the element internal forces. In particular, the action of the internal force vector $\mathbf{l}(\mu)$ on the infinitesimal variation $\delta\mathbf{d}$ of the deformation vector is defined by (5.63) and the virtual internal work turns out to be:

$$\delta W^{\mathsf{int}} = \int_0^1 \mathbf{l}^* \cdot \delta\mathbf{d}\,\mathrm{d}\mu = \int_0^1 \mathbf{l}(\delta\mathbf{d})\,\mathrm{d}\mu\,,$$

which, by means of (6.9) becomes

$$\delta W^{\mathsf{int}}(\delta\mathbf{h}_{\mathsf{AB}}) = \int_0^1 \mathbf{l}(\mathbf{B}_{\mathbf{d}}\delta\mathbf{h}_{\mathsf{AB}})\,\mathrm{d}\mu = \int_0^1 (\mathbf{B}^{\mathsf{T}}_{\mathbf{d}}\mathbf{l})(\delta\mathbf{h}_{\mathsf{AB}})\,\mathrm{d}\mu\,,$$

where, similarly to the displacement operator, the transpose of the deformation operator is the linear map $\mathbf{B}^{\mathsf{T}}_{\mathbf{d}}$ between $\mathbb{R}^6 \cong \mathfrak{se}^*(3)$ and $\mathbb{R}^{12} \cong \mathfrak{se}^*(3) \times \mathfrak{se}^*(3)$.

Furthermore, considering also (6.13), the internal virtual work becomes

$$\delta W^{\mathsf{int}}(\delta\mathbf{h}_{\mathsf{AB}}) = \int_0^1 \mathbf{B}^{\mathsf{T}}_{\mathbf{d}}\mathbf{l}\,\mathrm{d}\mu \cdot \delta\mathbf{h}^*_{\mathsf{AB}} = \mathbf{f}^{\mathsf{int}}_{\mathsf{AB}} \cdot \delta\mathbf{h}^*_{\mathsf{AB}}\,,$$

where $\mathbf{f}^{\mathsf{int}}_{\mathsf{AB}} \in \mathbb{R}^{12}$ is the vector gathering the element internal forces:

$$\mathbf{f}^{\mathsf{int}}_{\mathsf{AB}} = \int_0^1 \mathbf{B}^{\mathsf{T}}_{\mathbf{d}}\mathbf{l}\,\mathrm{d}\mu\,, \tag{6.15}$$

that is

$$\mathbf{f}^{\mathsf{int}}_{\mathsf{AB}} = \begin{bmatrix} \mathbf{f}^{\mathsf{int}}_{\mathsf{A}} \\ \mathbf{f}^{\mathsf{int}}_{\mathsf{B}} \end{bmatrix}.$$

Please observe again that $\mathbf{f}_{\mathsf{A}}^{\mathsf{int}}$ and $\mathbf{f}_{\mathsf{B}}^{\mathsf{int}}$ represent the wrench vectors whose action on the displacement screws $\delta\mathbf{h}_{\mathsf{A}}$ and $\delta\mathbf{h}_{\mathsf{B}}$, respectively, provides the virtual work $\mathbf{f}_{\mathsf{A}}^{\mathsf{int}}(\delta\mathbf{h}_{\mathsf{A}}) + \mathbf{f}_{\mathsf{B}}^{\mathsf{int}}(\delta\mathbf{h}_{\mathsf{B}}) = \mathbf{f}_{\mathsf{A}}^{\mathsf{int}} \cdot \delta\mathbf{h}_{\mathsf{A}}^{*} + \mathbf{f}_{\mathsf{B}}^{\mathsf{int}} \cdot \delta\mathbf{h}_{\mathsf{B}}^{*}$, that is the same as the virtual work δW^{int} relevant to the internal force wrench $\mathbf{l}$.

Moreover, the internal force vector $\mathbf{l}(t)$, relevant to the current configuration $\mathscr{B}_t$, is expressed by (5.74) as linearly dependent on the cumulative strain vector $\boldsymbol{\chi}(t)$, given by (5.75). Also, since (6.2) implies a uniform deformation screw vector $\mathbf{d}$, the cumulative cross-section strains result in

$$\boldsymbol{\chi}(t) = \int_{t_0}^{t} \delta\boldsymbol{\chi}(\bar{t}) = \int_{t_0}^{t} \frac{\delta\mathbf{d}(\bar{t})}{\lambda(\bar{t})}\,, \tag{6.16}$$

whence the internal force vector $\mathbf{l} = \mathbf{K}_{\Sigma}\boldsymbol{\chi}$ does not depend on the material abscissa μ.

Consequently, (6.15) becomes

$$\mathbf{f}_{\mathsf{AB}}^{\mathsf{int}} = \mathbf{B}_{\mathbf{d}}^{\mathsf{T}}\mathbf{K}_{\Sigma}\boldsymbol{\chi} = \mathbf{B}_{\mathbf{d}}^{\mathsf{T}}\mathbf{K}_{\Sigma} \int_{t_0}^{t} \frac{\delta\mathbf{d}(\bar{t})}{\lambda(\bar{t})}\,. \tag{6.17}$$

It is worth highlighting that if the beam axis elongation is constant, that is $\lambda(\bar{t}) = \lambda_0 = \lambda(t_0)$ for any intermediate configuration $\mathscr{B}_{\bar{t}}$, the ratio $\delta\mathbf{d}(\bar{t})/\lambda_0$ becomes an exact differential. Then, the cumulative strain vector is given by the difference between the current deformation $\mathbf{d}(t)$ and the reference one $\mathbf{d}(t_0) = \mathbf{d}_0$, divided by length of the beam axis at the reference configuration $l_0 = \lambda_0 = \|\mathbf{d}_{0\mathsf{t}}\|$.

With this assumption, the internal forces can be explicitly evaluated as

$$\mathbf{f}_{\mathsf{AB}}^{\mathsf{int}} = \mathbf{B}_{\mathbf{d}}^{\mathsf{T}}\mathbf{K}_{\Sigma}\frac{\mathbf{d}(t) - \mathbf{d}_0}{l_0}\,, \tag{6.18}$$

which is the same evaluation proposed by Sonneville et al. (2014).

In more general hypotheses, the cumulative strain vector can be evaluated by a numerical computation of the integral (6.16), obtaining the relevant element internal forces $\mathbf{f}_{\mathsf{AB}}^{\mathsf{int}}$ for the current configuration by (6.17).

6.4 Element Stiffness Matrix

The element stiffness matrix $\mathbf{K}^{\mathsf{el}}$ can be numerically evaluated considering its role as the gradient of the element force vector.

Specifically, assuming that the distributed loads are null, the element force vector specializes to $\mathbf{f}^{\mathsf{el}}_{\mathsf{AB}} = \mathbf{f}^{\mathsf{int}}_{\mathsf{AB}}$, to be evaluated by means of (6.18) as a function $\boldsymbol{\psi}$ of the current configurations $\mathbf{H}_{\mathsf{A}}$ and $\mathbf{H}_{\mathsf{B}}$:

$$\mathbf{f}^{\mathsf{el}}_{\mathsf{AB}} = \boldsymbol{\psi}(\mathbf{H}_{\mathsf{A}}, \mathbf{H}_{\mathsf{B}})\,, \tag{6.19}$$

in such a way that the current deformation $\mathbf{d}$ is the vector representation of the screw $\widehat{\mathbf{d}} = \log(\mathbf{H}_{\mathsf{A}}^{-1}\mathbf{H}_{\mathsf{B}})$.

Supposing that infinitesimal screw vectors $\delta\mathbf{h}_{\mathsf{A}}$ and $\delta\mathbf{h}_{\mathsf{B}}$ are applied, the relevant variations of the configuration matrices $\mathbf{H}_{\mathsf{A}}$ and $\mathbf{H}_{\mathsf{B}}$ result

$$\delta\mathbf{H}_{\mathsf{A}} = \mathbf{H}_{\mathsf{A}}\delta\widehat{\mathbf{h}}_{\mathsf{A}}\,, \qquad \delta\mathbf{H}_{\mathsf{B}} = \mathbf{H}_{\mathsf{B}}\delta\widehat{\mathbf{h}}_{\mathsf{B}}\,,$$

and, differentiating (6.19), the infinitesimal variation of the force vector can be expressed as

$$\delta\mathbf{f}^{\mathsf{el}}_{\mathsf{AB}} = \frac{\partial\boldsymbol{\psi}}{\partial\mathbf{h}_{\mathsf{AB}}}\delta\mathbf{h}_{\mathsf{AB}}\,, \tag{6.20}$$

where $\delta\mathbf{h}_{\mathsf{AB}}$ gathers the vectors $\delta\mathbf{h}_{\mathsf{A}}$ and $\delta\mathbf{h}_{\mathsf{B}}$.

Since the tangent stiffness matrix plays the role to provide the variation $\delta\mathbf{f}^{\mathsf{el}}_{\mathsf{AB}}$ when an increment $\delta\mathbf{h}_{\mathsf{AB}}$ is applied to the configurations of the element's nodes, $\mathbf{K}^{\mathsf{el}}$ is formally the gradient of the function $\boldsymbol{\psi}$ with respect to the infinitesimal variation $\delta\mathbf{h}_{\mathsf{AB}}$:

$$\mathbf{K}^{\mathsf{el}} = \frac{\partial\boldsymbol{\psi}}{\partial\mathbf{h}_{\mathsf{AB}}}\,. \tag{6.21}$$

At this point, in order to evaluate $\mathbf{K}^{\mathsf{el}}$, automatic differentiation via dual numbers can be applied.

Dual numbers have been introduced in Section 4.5.1 to provide a useful representation of rigid motions. At the same time, when considered as variables of an analytical function, they provide the derivative of such a function, as specified by (4.91).

Then, applying the vector function $\boldsymbol{\psi}$ in (6.19) at the dual variables $\tilde{\mathbf{H}}_{\mathsf{A}} = \mathbf{H}_{\mathsf{A}} + \epsilon\delta\mathbf{H}_{\mathsf{A}}$ and $\tilde{\mathbf{H}}_{\mathsf{B}} = \mathbf{H}_{\mathsf{B}} + \epsilon\delta\mathbf{H}_{\mathsf{B}}$, the dual extension of $\mathbf{f}^{\mathsf{el}}_{\mathsf{AB}}$ is obtained:

$$\tilde{\mathbf{f}}^{\mathsf{el}}_{\mathsf{AB}} = \boldsymbol{\psi}(\tilde{\mathbf{H}}_{\mathsf{A}}, \tilde{\mathbf{H}}_{\mathsf{B}}) = \mathbf{f}^{\mathsf{el}}_{\mathsf{AB}} + \epsilon\delta\mathbf{f}^{\mathsf{el}}_{\mathsf{AB}} = \mathbf{f}^{\mathsf{el}}_{\mathsf{AB}} + \epsilon\frac{\partial\boldsymbol{\psi}}{\partial\mathbf{h}_{\mathsf{AB}}}\delta\mathbf{h}_{\mathsf{AB}}\,,$$

that is

$$\tilde{\mathbf{f}}^{\mathsf{el}}_{\mathsf{AB}} = \boldsymbol{\psi}(\tilde{\mathbf{H}}_{\mathsf{A}}, \tilde{\mathbf{H}}_{\mathsf{B}}) = \mathbf{f}^{\mathsf{el}}_{\mathsf{AB}} + \epsilon\mathbf{K}^{\mathsf{el}}\delta\mathbf{h}_{\mathsf{AB}}\,. \tag{6.22}$$

Exploiting the relation here above, the element stiffness matrix $\mathbf{K}^{\mathsf{el}}$ can be assembled observing that its i-th column is the dual part of the dual vector $\tilde{\mathbf{f}}^{\mathsf{el}}_{\mathsf{AB}}$ when $\delta\mathbf{h}_{\mathsf{AB}}$ is the i-th unitary column vector $\mathbf{e}_i = [0 \cdots 1 \cdots 0]^{\mathsf{T}}$.

6.5 Numerical Tests

The effectiveness of the described formulation is shown by considering two representative numerical tests also presented in the specialized literature.

6.5.1 Cantilever Beam

As an introductory application let us consider the test case of the cantilever beam in Figure 6.1, which is frequently considered in the specialized literature (Jelenić and Crisfield, 1999, Magisano et al., 2020, Simo and Vu-Quoc, 1986, Sonneville et al., 2014).

The beam axis has the shape of a circular arc with radius $r = 100$, subtending an angle measuring $\pi/4$, while the cross-section is a unit square. Young's modulus is $E = 1 \times 10^7$ and the Poisson ratio is $\nu = 0$.

A transverse force is applied at the free-end, and two values are considered, that is $F = 300$ and $F = 600$.

The numerical analysis has been performed by means of the Newton method, whose convergence and stability have been tested by considering several discretizations of the beam and of the load increments, for both the loading conditions.

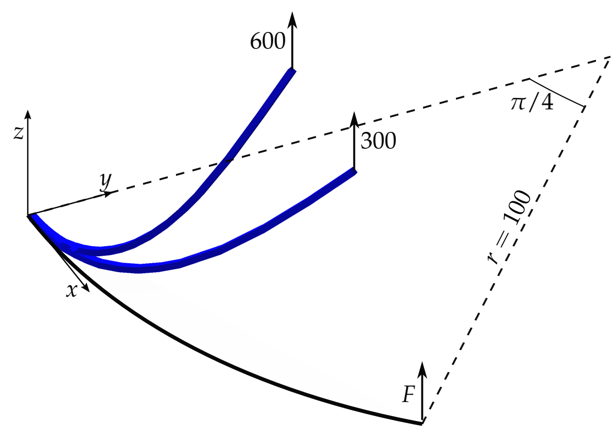


Figure 6.1: Bent cantilever beam loaded by a transverse force.

Table 6.1: Average number of iterations by Newton method and out of plane displacement of the beam in Figure 6.1 for $F = 300$.

Load steps	Number of elements				
	1	2	4	8	16
1	38	Fails	Fails	Fails	Fails
5	11.2	11	11.4	11.4	11.4
10	8	8.7	8.6	8.2	8.1
20	7	7	7	7	7
40	6.3	6.4	6	6	6.3
u_z	32.4353	38.5746	40.0100	40.3634	40.4514

Tables 6.1 and 6.2 show the average number of iterations required for convergence at each load increment, for $F = 300$ and $F = 600$, respectively. In addition, the values of the displacement in z-direction of the loaded point are also reported. For each beam mesh, these values do not depend on the number of load steps, confirming the path-independence of the formulation.

As expected the standard Newton method fails to converge when the equilibrium configuration is too far from the initial one, and a multi-step analysis is required. However, since the implemented formulation is intrinsically path-independent, the equilibrium configuration can be reached by gradually increasing the applied loads without affecting the validity of the

Table 6.2: Average number of iterations by Newton method and out of plane displacement of the beam in Figure 6.1 for $F = 600$.

Load steps	Number of elements				
	1	2	4	8	16
1	Fails	Fails	Fails	Fails	Fails
5	14.8	16	Fails	Fails	Fails
10	10.2	10.2	10.1	10.3	10.4
20	7.8	8.3	8.2	8	8.1
40	6.8	6.9	6.8	6.8	6.8
u_z	44.5472	51.2959	53.0373	53.4659	53.5725

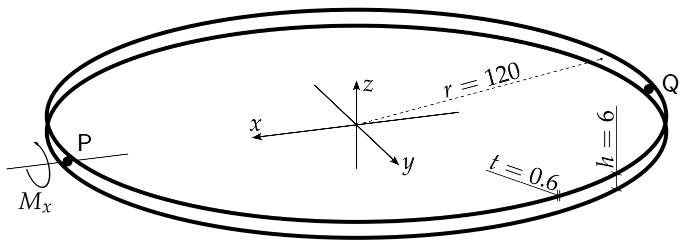


Figure 6.2: Deployable ring undergoing wrapping rotation.

solution.

6.5.2 Deployable Ring

In order to verify the capability of the finite element formulation to handle arbitrarily large rotations, the example in Figure 6.2 is considered. Such a test case, introduced by Yoshiaki et al. (1992) and also reported by Magisano et al. (2020), depicts the deformation of an elastic ring wrapping in on itself as an effect of a rotation around a diametrical axis.

The ring has a radius $r = 120$, with a thin rectangular cross-section having height $h = 6$ and thickness $t = 0.6$. The Young modulus is $E = 2 \times 10^5$ and the Poisson ration is $\nu = 0.3$.

The point Q is fully clamped, while the control point P is allowed to move only along the x-direction. The rotation ϕ_x of P around the x-axis is assumed as a control kinematic parameter, while M_x is obtained as the relevant constraint reaction moment. Hence, controlling the values of the

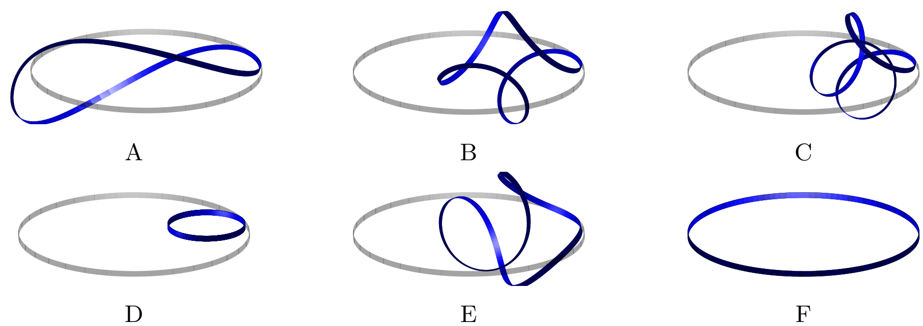


Figure 6.3: Evolving configuration of the elastic ring in Figure 6.2.

rotation ϕ_x, the analysis can be performed by the standard Newton method.

Figure 6.3 shows the evolution of the configuration under increasing values of ϕ_x. First, the ring twists about the diametrical axis. Then, it wraps around itself and fold into a smaller ring when ϕ_x equals 2π. Finally, keeping on increasing the control parameter, the ring untangles and goes back to its initial configuration at $\phi_x = 4\pi$.

In order to verify the effective path-independence of the numerical formulation and its stability under very large rotations, five complete cycles are considered, so that ϕ_x is gradually increased between 0 and 20π.

The elastic energy relevant to the configuration $\mathscr{C}_t$ at the time t can be evaluated by halving the mechanical work of the external loads. Specifically, the only non vanishing contribution is the work of the reaction M_x at the control point P:

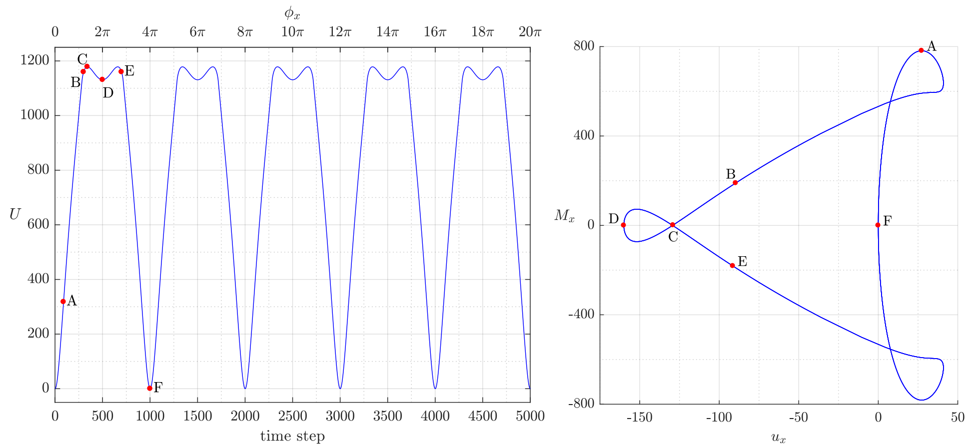
$$U(t) = \int_{t_0}^{t} \mathrm{d}U = \frac{1}{2}\int_{t_0}^{t} \mathrm{d}W^{\text{ext}} = \frac{1}{2}\int_{t_0}^{t} M_x(\bar{t})\dot{\phi}_x(\bar{t})\mathrm{d}\bar{t}\,,$$

where $\mathrm{d}\phi_x = \dot{\phi}_x(\bar{t})\mathrm{d}\bar{t}$ is the infinitesimal variation of the rotation ϕ_x.

Moreover, since very small variations of the rotation ϕ_x are applied in the numerical analysis, the elastic energy at the i-th step can be estimated as

$$U(t_i) \approxeq U_i = \frac{1}{2}\sum_{j=1}^{i} M_{xi}\Delta\phi_{x_i} = \frac{1}{2}\sum_{j=1}^{i} M_{xi}(\phi_{x_i} - \phi_{x_{i-1}})\,.$$

The elastic energy U of the deformed ring is plotted in Figure 6.4a with



(a) (b)

Figure 6.4: Deformation process for the elastic ring in Figure 6.2: (a) elastic energy; (b) configuration path. The marked steps refer to the configurations in Figure 6.3.

respect to the time steps and the rotation ϕ_x, for five consecutive cycles. The steps relevant to the configurations in Figure 6.3 are also highlighted.

Please observe that $U(t)$ has a local minimum when the rings completely folds (D), and vanishes when the ring overlaps to its initial configuration (F). Then, the graph recurs identically for the successive cycles.

The cyclic behavior is also clear from Figure 6.4b, which shows the configuration path following the moment M_x and the horizontal displacement u_x of the control point P. The five cycles perfectly overlap, proving again the path-independence and the stability of the proposed formulation.

Chapter 7

Conclusions and Outlook

This dissertation has presented a comprehensive mathematical framework for modeling beams undergoing large displacements, what arises from an appropriate characterization of the beam kinematics.

In particular, it has been shown how the reduction process from a slender body to a one-dimensional continuous system, in line with the geometrically exact approach, induces one to represent the configuration of a beam cross-section by means of a rigid-body motion within the physical space. Then, the configuration space of a beam is naturally identified with the space of the proper Euclidean motions.

For this reason, a mathematically consistent discussion on the beam mechanics cannot be separated from the algebraic characterization of such a space.

As a matter of fact, a proper Euclidean motion is a particular kind of affine transformation, i.e. the one preserving the distance between affine points and the orientation of affine frames. As such, the group of Euclidean motions, namely $SE(3)$, inherits the same algebraic structure of the affine group.

In particular, with specific reference to a rigid body motion, any decomposition involving a translation should take into account the action of the rotation component on the translation one. At the same time, this group action, necessarily involved in the subgroup decomposition of $SE(3)$, results in the coupling of rotations and translations within the physical space, that is the coupling of rotation and translation kinematic parameters of any beam cross-section.

In parallel, in accordance with the assumed continuity and differentiability of the beam kinematics, the group $SE(3)$ has also the structure of a smooth manifold, being specifically indicated as a Lie group.

However, the differential properties of a Lie group are more general than the ones of a Euclidean vector space, so that some refined notions, specific of the differential geometry, are required. In some way, one could claim that it is exactly because of the non-linear algebraic structure of the configuration space that a beam exhibits a non-linear behavior in the physical space.

The description of the beam configuration within the context of a Lie group has also effects on statics. In fact, when one refers to a non-flat manifold, the notion of a force is algebraically well-defined only if it is related to its action on virtual displacements, intended as infinitesimal variations on the Lie group. Moreover, since a virtual displacement belongs to a local linearization of the neighborhood of a point, one should specify the point of the non-linear space it is associated to, the same applying for the relevant static quantities.

On account of such algebraic characterization of the configuration space of a beam, a finite beam element has been formulated by introducing the shape functions as integral curves of a vector field on the Lie group $SE(3)$.

Hence, representing the element's nodes as points of $SE(3)$, they are interpolated in their current configuration, leading to a formulation that is intrinsically path-independent. Such a property is confirmed by the numerical results, that also validate the effectiveness of the automatic differentiation, based upon the algebra of dual numbers, in evaluating the element stiffness matrix.

Nevertheless, some issues still require further in-depth studies. In particular, the standard Newton method results ineffective to converge when the equilibrium configuration is not sufficiently close to the reference one, requiring a large number of load steps. Moreover, the Newton method requires a prescribed load history, in terms of either force or displacement parameters, and cannot be applied for general configuration paths adapting forces and displacements simultaneously.

In this respects, some refined numerical methodologies already presented in the specialized literature (Magisano et al., 2017, 2020), as well as specifically derived from Lie group methods (Iserles et al., 2000, Munthe-Kaas, 1998), could be adapted to the beam model specifically described.

Appendix A

Review of Algebraic Structures

A.1 Basic Definitions

We introduce the basics of some algebraic structures mostly following Dummit and Foote (2003) and Lang (2002), as well as other standard textbooks in abstract algebra.

First, basic notions and symbols about sets and maps are briefly recalled. Then, we will see how several algebraic structures can be constructed by introducing specific operations on sets.

A.1.1 Sets and Maps

We denote a *set* by capital letters, such as A, with the usual meaning of a collection of objects shearing a common property.

If an object a is an *element*, or a *member*, of the set A, we write $a \in A$. Otherwise, we write $a \notin A$.

If A and B are sets, we write $A \subseteq B$ to say that A is a subset of B, or also $B \supseteq A$ to say that B contains A. If A is a *proper* subset of B, i.e. $A \neq B$, we write $A \subsetneq B$, or simply $A \subset B$. Similarly, $B \supsetneq A$, as well as $B \supset A$, means that B properly contains A.

The *Cartesian product* of two sets A and B is the collection of ordered pairs of elements from A and B:

$$A \times B = \{\, (a, b) \mid a \in A\,,\ b \in B \,\}\,.$$

Repeating n times the Cartesian product of a set A with itself defines its n-ary Cartesian power:

$$A^n = \underbrace{A \times \cdots \times A}_{n \text{ times}} = \{\, (a_1, \ldots, a_n) \mid a_i \in A\,,\ i = 1, \ldots, n \,\}\,.$$

Definition A.1. A *mapping*, or simply a *map*, from the set A to the set B is a relation f which associates each element of A with an element of B:

$$\begin{aligned} f:\ & A \to B \\ & a \mapsto b\,. \end{aligned}$$

We say that A and B are the *domain* and the *codomain* of f, respectively, and that the element $a \in A$ is mapped to $b \in B$. We also say that b is the *image* of a through f and write $f(a) = b$. More generally, the image of a subset $X \subseteq A$ under f is the subset $Y \subseteq B$ collecting the images of the elements in X:

$$Y = f(X) = \{\, f(a) \mid a \in X \subseteq A \,\} \subseteq B\,.$$

If $X = A$, we also say that $f(A)$ is the *image*, or the *range*, of f:

$$\mathrm{Im}(f) = f(A) = \{\, f(a) \mid a \in A \,\} \subseteq B\,.$$

For each subset $Y \subseteq B$, its *preimage*, or its *inverse image*, is the subset of A having Y as image under f:

$$f^{-1}(Y) = \{\, a \in A \mid f(a) \in Y \,\}\,.$$

The preimage of the set $\{b\}$, consisting of a single element $b \in B$, is called the *fiber* of f over b.

Remark. Please notice that f^{-1} is not, in general, a map. Actually, depending on the properties of f and the subset Y, the preimage of Y could be an empty set, as well as there could be several elements of A mapping to the same element $b \in B$.

A.1.2 Properties of Maps

The fiber over $f(a)$, with $a \in A$, induces a binary relation on the domain:

$$a \sim_f a' \quad \Leftrightarrow \quad f(a) = f(a')\,. \tag{A.1}$$

One can easily verify that this is an equivalence relation on A and implies the following definition (see, e.g., Bergman (2011), Mac Lane and Birkhoff (1999)).

Definition A.2. Let f be a map from A to B. The equivalence relation (A.1) on the domain A is called the *equivalence kernel* of f. Explicitly, the kernel of f can be represented as a subset of the Cartesian power A^2:

$$\mathrm{Ker}(f) = \{\, (a, a') \in A^2 \mid f(a) = f(a')\,\}\,. \tag{A.2}$$

The notion of kernel here introduce has a general meaning. However, in the subsequent sections, such a definition will be specified to the features of the map f and the involved sets.

The primary map to be defined on a set A is the *identity map*. It is the unary operation $\mathbf{id}_A$ associating each element of A with itself:

$$\begin{aligned} \mathrm{id}_A :\ & A \to A \\ & a \mapsto a\,. \end{aligned} \tag{A.3}$$

Definition A.3. Let $f\colon A \to B$ and $g\colon B \to C$ be maps. The *composition* of g with f is the map $g \circ f\colon A \to C$ such that

$$(g \circ f)(a) = g\big(f(a)\big)\,, \quad \forall\, a \in A\,.$$

We also say that $g \circ f$ is the *composite map* of g and f.

The maps f and g are *composable* if their composition, in a given order, is meaningful. Actually, the composite map $g \circ f$ only makes sense if

$$f\colon A \to B\,, \quad g\colon C \subseteq B \to D\,.$$

Moreover, if the codomain of f is not fully contained within the domain of g, we usually restrict f to a subset A' in such a way that $C \subseteq f(A')$.

Remark. Even if composite map $g \circ f$ is well-defined, the composition in the reverse order, i.e. $f \circ g$, only makes sense when $A \subseteq D$. Also, in general the composite maps are not the same, that is

$$g \circ f \neq f \circ g\,.$$

On the contrary, when $g \circ f = f \circ g$ we say that f and g *commute*.

A property of mapping composition which always applies is associativity. Explicitly, given the composable maps f, g and h, we can write

$$h \circ (g \circ f) = (h \circ g) \circ f = h \circ g \circ f\,.$$

Definition A.4. Let $f\colon A \to B$ be a map between two sets A and B. The following definitions hold:

- f is *injective*, or f is an *injection*, if distinct elements of the domain A are mapped to distinct elements of the codomain B:

  $$a_1 \neq a_2 \;\;\Rightarrow\;\; f(a_1) \neq f(a_2)\,;$$

- f is *surjective*, or f is a *surjection*, if the preimage of the codomain B corresponds with the domain A:

  $$\forall\, b \in B,\;\; \exists\, a \in A : \, f(a) = b\,;$$

- f is *bijective*, or f is a *bijection*, if it is both injective and surjective:

  $$\forall\, b \in B,\;\; \exists!\, a \in A : \, f(a) = b\,;$$

- f has a *left inverse* if there exists a map $g : B \to A$, providing the identity map when is composed to the left of f:

$$g \circ f = \mathrm{id}_A \,;$$

- f has a *right inverse* if there exists a map $h\colon B \to A$ which provides the identity map by right composition with f:

$$f \circ h = \mathrm{id}_B \,.$$

Proposition A.5. *Let f be a map from A to B. The following properties hold true:*

1. *f is injective if and only if it has a left inverse;*
2. *f is surjective if and only if it has a right inverse;*
3. *f is bijective if and only if there exists a map $g\colon B \to A$ which is both a left and right inverse of f.*

For the proofs one can refers, among others, to Mac Lane and Birkhoff (1999).

The unique map g that is both a left and right inverse of a bijection f is called the *two-sided inverse*, or simply the *inverse*, of f.

We also say that f is *invertible*, meaning that the two-sided inverse does exist and is represented by f^{-1}.

The composite map of injections is always injective, and the same holds for surjective maps. As a consequence, the composition of two bijections f and g provides the bijective map $g \circ f$. Such a map is invertible and the inverse map satisfies the following property:

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1} \,.$$

Definition A.6. A bijection from a set A to itself is called a *permutation* of A. The collection of all the permutations of A is itself a set and is referred to as $\mathrm{Perm}(A)$.

A.1.3 Homomorphisms

Here we recall the fundamental notions useful for the further developments, referring to the specialized literature for the details (see, e.g., Cohn (1981)).

An *algebraic structure* is a set A endowed with a collection of operations satisfying a finite number of properties. The number of arguments taken by any operation is called its *arity*.

Definition A.7. A map between two algebraic structures is called a *homomorphism*, or a *homomorphic mapping*, if it is structure-preserving.

With reference to a single operation μ of arity k, we say that the map $f\colon A \to B$ between the sets A and B preserves μ, or is *compatible* with this operation, if

$$f\big(\mu_A(a_1, \ldots, a_k)\big) = \mu_B\big(f(a_1), \ldots, f(a_k)\big) , \ \ \forall\, a_1, \ldots, a_k \in A .$$

As an example, considering a binary operation, the map f is compatible if

$$f(x * y) = f(x) * f(y) , \ \ \forall\, x, y \in A .$$

Please notice that in the last relation there is a slight abuse of notation, since the same symbol '$*$' has been used to represent the operations μ_A and μ_B associated with two distinct algebraic structures.

An injective homomorphism is called a *monomorphism*, and a surjective one is called an *epimorphism*. We say that a homomorphism $f\colon A \to B$ is an *isomorphism* if it is bijective. An isomorphism is also invertible, i.e. there exists a map $g\colon B \to A$ such that

$$g \circ f = \mathrm{id}_A , \quad f \circ g = \mathrm{id}_B .$$

The inverse map g is itself a homomorphism.

Like any map between sets, the *kernel* of the homomorphism f can be defined as the following equivalence relation:

$$x \sim_f x' \ \ \Leftrightarrow \ \ f(x) = f(x') ,$$

or, explicitly, as the subset of ordered pairs of elements of A mapped to the same element of B:

$$\mathrm{Ker}(f) = \{\, (x, x') \in A^2 \mid f(x) = f(x') \,\} \,. \tag{A.4}$$

The kernel relation of the homomorphism f is properly a *congruence relation*, since it is compatible with the algebraic structure of A (see, e.g. Bergman (2011)). As an example, denoting by '$*$' the law of composition defined over A, the relation $\equiv$ is a congruence on A if it is an equivalence and satisfies

$$x \equiv x' \,,\; y \equiv y' \;\;\Rightarrow\;\; x * y \equiv x' * y' \,.$$

With reference to definition (A.4) of kernel of f, one can verify that

$$f(x *_A y) = f(x) *_B f(y) = f(x') *_B f(y') = f(x' *_A y') \,,$$

whence

$$x \sim_f x' \,,\; y \sim_f y' \;\;\Rightarrow\;\; x *_A y \sim_f x' *_A y' \,.$$

A homomorphism $f\colon A \to A$, such that the domain and the codomain are the same, is called an *endomorphism*. If f is also an isomorphism, we say that it is an *automorphism*.

A permutation of the set A represents an automorphism of A, intended as an algebraic structure, when it is compatible with the relevant algebraic operations. The set of all the automorphisms of A is $\mathrm{Aut}(A)$.

It is important to point out that the definition of an algebraic structure strictly depends on the operations defined on the underlying set. This means that the same set can provide several algebraic structures depending on the collection of operations considered.

A.2 Algebraic Structures

A.2.1 Monoids

Let us consider a set S and a binary operation $S \times S \to S$, mapping the pair (a, b) to the element $a \cdot b$, also denoted by ab. The algebraic structure $(S, \cdot)$ is a *monoid* if the following properties are satisfied:

- associativity:

$$a(bc) = (ab)c\,, \ \ \forall\, a, b, c \in S\,;$$

- existence of *identity*, also called *unit element*, $e \in S$:

$$ae = ea = a\,, \ \ \forall\, a \in S\,.$$

The binary operation defining the monoid structure is sometimes called *law of composition* of S into itself (see, e.g., Lang (2002)).

It is easy to verify that the identity element is unique. Actually, if e_1 and e_2 both satisfy the identity axiom, one has $e_1 = e_1 e_2 = e_2$.

The identity e can be seen as the value of a constant operator, i.e. a 0-ary operation with no arguments (Cohn, 1981). Consequently, a monoid is specified as the triple $(S, \cdot, e)$ made of the set S, the binary operation $\cdot$ and the nullary operation e.

Given n elements $a_1, \ldots, a_n$ of the monoid S, one can recursively apply the composition, obtaining the sequence $p_i = p_{i-1} a_i\,, i = 1, \ldots, n$ with initial element $p_0 = e$. Then, the p_n is the product of the sequence $(a_1, \ldots, a_n)$, also written as

$$p_n = \prod_{i=1}^{n} a_i\,.$$

The special case when $a_i = a$, the product p_n provides the n-th power of the element a, represented as a^n.

If $(M, \cdot, e_M)$ and $(N, *, e_N)$ are two monoids, we say that a map $f\colon M \to N$ is a *monoid homomorphism* if

- $f(x \cdot y) = f(x) * f(y)\,, \ \ \forall\, x, y \in M$;

- $f(e_M) = e_N$.

Let us observe that such properties fulfill the requirement for f to preserve the monoid structure, i.e. to be compatible with both the binary operations $(\cdot, *)$ and the nullary ones (e_M, e_N).

A.2.2 Groups

A group is a monoid $(G, \cdot)$ such that for every element $a \in G$ there exists an element $b \in G$ verifying $ab = ba = e$. The element b is called the *inverse* of a.

Definition A.8. A *group* is a set G with a binary operation $G \times G \to G$ satisfying the following axioms:

- associativity:

$$a(bc) = (ab)c\,, \;\; \forall\, a, b, c \in G\,;$$

- existence of the identity element $e \in G$:

$$ae = ea = a\,, \;\; \forall\, a \in G\,;$$

- existence of the inverse element:

$$ab = ba = e\,, \;\; \forall\, a \in G\,.$$

A group G inherits all the properties of the monoid structure, e.g. the uniqueness of the unity element e and the recursive definition of the n-th power a^n of an element a.

In addition, the group axioms also imply the uniqueness of the inverse for any element of G. Actually, let b be the inverse of $a \in G$. If c is also the inverse of a, one has:

$$c = ce = cab = eb = e\,.$$

The inverse of $a \in G$ is usually denoted as a^{-1}. Moreover, for any non-negative integer n, let us observe that

$$e = aa^{-1} = aea^{-1} = a(aa^{-1})a^{-1} = \underbrace{a \cdots a}_{n \text{ times}} \underbrace{a^{-1} \cdots a^{-1}}_{n \text{ times}} = a^n (a^{-1})^n ,$$

whence we find

$$(a^n)^{-1} = (a^{-1})^n = a^{-n} .$$

The existence of the inverse can be interpreted as the effect of the following unary operation:

$$\begin{aligned} \theta :\ & G \to G \\ & a \mapsto a^{-1} , \end{aligned}$$

so that the group is identified with the quadruple $(G, \cdot, e, \theta)$ made of the set G, the binary operation $\cdot$, the nullary operation e and the unary operation θ.

We recall that a map can be qualified as homomorphism if it is compatible with the operations characterizing the algebraic structure. With specific reference to a group, the following definition is introduced.

Definition A.9. Let $(G, \cdot, e_G, \theta_G)$ and $(H, *, e_H, \theta_H)$ be groups. The map $f \colon G \to H$ is a *group homomorphism*, if the following properties are satisfied:

- $f(x \cdot y) = f(x) * f(y) , \ \forall x, y \in G$;
- $f(e_G) = e_H$;
- $f(x^{-1}) = f(x)^{-1} \ \Leftrightarrow \ f\big(\theta_G(x)\big) = \theta_H\big(f(x)\big) , \ \forall x \in G$.

Examples of Groups

An important example of group is the set of permutations defined over a set A, along with the operation of map composition. Specifically, let A be a set and $\mathrm{Perm}(A)$ the set of all its permutations. We can verify that $\mathrm{Perm}(A)$, endowed with the composition rule, forms a group.

First, we recall that composition is associative and that $f \circ g$ is itself a bijection from A to itself, for all f and g in $\mathrm{Perm}(A)$.

Moreover, since the identity map $\mathbf{id}_A$ satisfies

$$f \circ \mathrm{id}_A = f\,, \quad \mathrm{id}_A \circ f = f\,, \ \ \forall f \in \mathrm{Perm}(A)\,,$$

it represents the identity element for $\mathrm{Perm}(A)$.

Finally, since any bijection is invertible, the existence of the inverse element f^{-1} is ensured for any f in $\mathrm{Perm}(A)$.

Hence, the group axioms are satisfied and we can conclude that $\mathrm{Perm}(A)$ along with map composition, is a group.

When the set A is endowed with a number of operations, it provides an algebraic structure and the set $\mathrm{Aut}(A)$ of all the automorphisms is defined. Similarly to $\mathrm{Perm}(A)$, one can easily prove that $\mathrm{Aut}(A)$, with the composition operation, has a group structure.

Translations and Conjugations

The uniqueness of the inverse element in a group G imply that, given two elements a and b, there is a unique solution in G satisfying the equation $ax = b$. One can easily verify that such a solution is $x = a^{-1}b$.

As a consequence, for any fixed $a \in G$, it is possible to define a bijection L_a as follows

$$\begin{aligned} L_a :\ & G \to G \\ & x \mapsto xa\,, \end{aligned} \tag{A.5}$$

whose inverse map is $L_a^{-1} = L_{a^{-1}}$.

The above-defined bijection is called the *left multiplication* by a, or the *left translation* by a. Similarly, it is possible to define the *right multiplication*, or the *right translation*, by a as the map

$$\begin{aligned} R_a :\ & G \to G \\ & x \mapsto xa\,, \end{aligned}$$

and it is trivial to verify that the inverse map is $R_a^{-1} = R_{a^{-1}}$. The existence of such an inverse is ensured by the fact that the equation $xa = b$ has the unique solution $x = ba^{-1} \in G$.

Please notice that left and right translations represent, in general, two distinct maps. Only if the binary operation defining the group structure is commutative, they coincide:

$$L_a(x) = ax = xa = R_a(x)\,, \ \ \forall\, a, x \in G\,.$$

When commutativity holds, we say that the G is an *abelian group*, or a *commutative group*.

It is worth noting that the translation map is not a group homomorphism, since it does not preserve the group structure. For example, one can verify that $L_a(e) = ae = a \neq e$ (the same holds for the right translation). The only exception is the translation by e, in which case both left and right multiplication maps coincide with the identity map:

$$L_e(x) = ex = R_e(x) = xe = x\,, \ \ \forall\, x \in G \quad \Leftrightarrow \quad L_e = R_e = \mathrm{id}_G\ .$$

Let us now consider an element $a \in G$ and compose the left translation L_a and the inverse of the right translation $R_{a^{-1}}$. Since the multiplication in G is associative, the resulting map does not depends on the order of composition and any element $x \in G$ transforms as follows:

$$(L_a \circ R_{a^{-1}})(x) = (R_{a^{-1}} \circ L_a)(x) = axa^{-1}\,, \ \ \forall\, a, x \in G\,.$$

The above operation is called the *conjugation* by a and is defined as

$$\begin{aligned} C_a :\ & G \to G \\ & x \mapsto axa^{-1}\,. \end{aligned} \tag{A.6}$$

Since C_a arises from the composition of bijective maps, it is a bijection itself. Moreover, differently from translation, conjugation is a group homomorphism, since it preserves the identity, the law of composition and the inversion in G:

- $C_a(e) = aea^{-1} = aa^{-1} = e\,;$
- $C_a(xy) = axya^{-1} = axeya^{-1} = (axa^{-1})(aya^{-1}) = C_a(x)C_a(y)\,,$ $\forall\, x, y \in G\,;$

- $C_a(x^{-1}) = ax^{-1}a^{-1} = a(ax)^{-1} = (axa^{-1})^{-1} = (C_a(x))^{-1}\,,\ \forall\, x \in G\,.$

In the end, a conjugation is an automorphism of G. Specifically, all the automorphisms in the form of a conjugation are said *inner* and form a subgroup of $\mathrm{Aut}(G)$ denoted by $\mathrm{Inn}(G)$.

Two elements x and y of G are said to be *conjugate* in G if there exists some $a \in G$ such that $x = C_a(y) = aya^{-1}$. The same definition applies to subsets X and Y of G such that $X = C_a(Y) = aYa^{-1}$.

One can easily verify that conjugacy is an equivalence relation and, as such, induces a partition of G. The equivalence class induced by $a \in G$ is called the *conjugacy class* of a.

Group Actions

Let G be a group with identity element e and A a set. A *group action* of G on A is a map

$$\begin{aligned} \alpha :\ & G \times A \to A \\ & (g, a) \mapsto \alpha(g, a) = g \cdot a\,, \end{aligned}$$

satisfying the following properties (Dummit and Foote, 2003):

- compatibility:

$$g_1 \cdot (g_2 \cdot a) = (g_1 g_2) \cdot a\,,\ \forall\, g_1, g_2 \in G\,,\ a \in A\,;$$

- identity:

$$e \cdot a = a\,,\ \forall\, a \in A\,.$$

When $g \in G$ is fixed, the map α is reduced to a map $\sigma_g \colon A \to A$ such that

$$\sigma_g(a) = \alpha(g, a) = g \cdot a\,,\ \forall\, a \in A\,.$$

It can bee proved that σ_g is a permutation of A (see, e.g., Dummit and Foote (2003)) whose inverse is $\sigma_g^{-1} = \sigma_{g^{-1}}$. Moreover, the map between G

and $\mathrm{Perm}(A)$ is a group homomorphism:

$$\begin{aligned} \sigma :\; & G \to \mathrm{Perm}(A) \\ & g \mapsto \sigma_g \,. \end{aligned}$$

Some authors, such as Lang (2002), introduce the definition of group action of G on A referring to the group homomorphism σ and then derive the map α with the relevant properties.

We also say that such a homomorphism is the *permutation representation* of the action, meaning that each element of G is associated with a permutation in $\mathrm{Perm}(A)$. The set A is called a G-*set*.

The *kernel* of the action is the set of elements of G that act trivially on every element of A:

$$\mathrm{Ker}(\sigma) = \{\, g \in G \mid g \cdot a = \sigma_g(a) = a\,,\ \ \forall\, a \in A \,\} \subseteq G\,.$$

In practice, the kernel of an action is precisely the same as the kernel of the associated permutation representation.

If the kernel of σ is the identity $e \in G$, we say that such an action is *faithful*. In other words, distinct elements in G induce distinct permutations of A, that is the associated permutation representation is injective.

For each element $a \in A$, we call the *stabilizer* of a in G the set G_a of elements of G that fix the element a (Dummit and Foote, 2003):

$$G_a = \{\, g \mid g \cdot a = \sigma_g(a) = a\,,\ a \in A \,\} \subseteq G \;. \tag{A.7}$$

The stabilizer of a has a group structure and is also called the *isotropy group* of a in G (see, Lang (2002)). A *fixed point* of G is an element $a \in A$ such that $g \cdot a = \sigma_g(a) = a$ for all $g \in G$, so that $G_a = G$.

Let G be a group acting on A and $a \in A$. We call the *orbit* of a under G the subset Ga of A made of all the elements resulting as $g \cdot a$, with $g \in G$ (Lang, 2002):

$$Ga = \{\, x \mid x = g \cdot a\,,\ g \in G \,\} \subseteq A\,.$$

The action of G on A is *transitive* if for any element of A its orbit under

G is not empty. Explicitly, given any two elements a and b in A, there exists $g \in G$ such that $a = g \cdot b$. If such g is unique, the action is said *simply transitive*, i.e. it is both faithful and transitive.

We want to observe that the map $G \times A \to A$ should be more precisely referred to as a *left action*, since the group element appears on the left of the set one. Similarly, a *right action* of G on A can be introduced by means of a map $A \times G \to A$.

In this respect, it is also clear that the left (right) translation over a group G defines a left (right) group action of G on itself.

Subgroups, Group Homomorphisms and Products

Let G be a group and $H \subseteq G$ a nonempty subset. We say that H is a *subgroup* of G if it is a group under the same law of composition defined in G.

In practice, H is a subgroup of G if it contains the identity element e and it is closed under products and inverses (see, e.g., Dummit and Foote (2003), Lang (2002)):

- $e \in H$;
- $x, y \in H \;\; \Rightarrow \;\; xy \in H\,, \;\; \forall\, x, y \in G$;
- $x \in H \;\; \Rightarrow \;\; x^{-1} \in H\,, \;\; \forall\, x \in G$.

Suppose H a subgroup of G and g an arbitrary element of G. We call the *left coset* of H in G the subset of G defined as

$$gH = \{\, gh \mid h \in G \,\} \subseteq G\,,$$

and any element of a coset is called a *representative* of the coset.

Please notice that the left coset gH can be thought as the left translate of H by $g \in G$ or, equivalently, as the image of the bijection $h \mapsto gh$. Also, since H is closed under products, $lH = H$ for any $l \in H$.

Let us observe that any element $g \in G$ represents the left translation of the identity e by itself, i.e. $g = ge$. Hence, if H is a subgroup of G, e is in H and then $g \in gH$.

At the same time, if x is a representative of gH, there exists $l \in H$ such that $x = gl$. Hence, $xH = glH = gH$, because $lH = H$ when l is in H. In other words, x is an element of exactly one left coset of H, that is $xH = gH$.

Any pair of representatives of a left coset gH, say x and y, are characterized by the property $xy^{-1} \in H$. Such a property is an equivalence relation, and so any left coset of a subgroup H provides an equivalence class. Then, x and y are coset representatives in the sense of class representatives.

As an equivalence class, the cosets of a subgroup H form a partition of the underlying group G.

Of course, one can also introduce an analogous definition for a *right coset* of H in G as

$$Hg = \{\, hg \mid h \in H \,\} \subseteq G \,,$$

and similar properties as for the left cosets holds.

Please notice that the left coset of H with respect to g is also the right coset of the conjugate of H by g:

$$gH = gHg^{-1}g = C_g(H)\, g \,.$$

Definition A.10. Let H and K be subgroups of a group G. The *multiplication* of H and K is the subset

$$HK = \{\, hk \mid h \in H \,,\; k \in K \,\} \subseteq G \,.$$

The multiplication of H and K is a union of the left cosets of K, as well as a union of the right cosets of K:

$$HK = \bigcup_{h \in H} hK = \bigcup_{k \in K} Hk \,.$$

Proposition A.11. *Let H and K be subgroups of G. Then, the multiplication HK is also a subgroup of G if, and only if, $HK = KH$.*

Proof. Let us verify that with $HK = KH$ the multiplication satisfies the group axioms.

- Since H and K are subgroups, they both contain the identity e. Then,

$ee = e \in HK$.

- The inverse of $hk \in HK$ is in $HK = KH$, since

$$(hk)^{-1} = k^{-1}h^{-1} \in KH = HK\,, \;\; \forall\, h \in H\,,\; k \in K\,.$$

- Let us consider $h_1, h_2 \in H$ and $k_1, k_2 \in K$, so that h_1k_1 and h_2k_2 belong to HK. By associativity, the multiplication results

$$(h_1k_1)(h_2k_2) = h_1(k_1h_2)k_2\,, \;\; \forall\, h_1, h_2 \in H\,,\; k_1, k_2 \in K\,.$$

  Because of the permutation property, there exist $h_2' \in H$ and $k_1' \in K$ such that

$$k_1h_2 = h_2'k_1'\,,$$

  whence

$$(h_1k_1)(h_2k2) = (h_1h_2')(k_1'k_2) \in HK\,.$$

  Then, HK is closed under composition of elements.

Conversely, let us assume HK is a subgroup of G and consider an arbitrary element hk in HK, with $h \in H$ and $k \in K$. By group axiom of inverse, the element $(hk)^{-1}$ is in HK, meaning that there exist $h_1 \in H$ and $k_1 \in K$ satisfying

$$(hk)^{-1} = h_1k_1 \in HK\,,$$

or, equivalently,

$$hk = k_1^{-1}h_1^{-1} \in HK\,,$$

Since k_1^{-1} is in K and h_1^{-1} is in H, for the arbitrary $hk \in HK$ we have found $k' = k_1^{-1}$ and $h' = h_1^{-1}$ such that

$$hk = k'h' \in KH\,. \qquad \square$$

When $HK = KH$, we say that the two subgroups *permute*. However, this does not imply that the elements of H commute with those of K. Actually, the permutation of H and K means that for any element $hk \in HK$, with $h \in H$ and $k \in K$, there exist $h' \in H$ and $k' \in K$ such that

$$hk = k'h' \in KH\,.$$

Definition A.12. A subgroup N of G is called *normal* if its left and right cosets coincide and we write $N \trianglelefteq G$ (Dummit and Foote, 2003):

$$gN = Ng\,,\ \ \forall\, g \in G\,.$$

The above definition implies that N equals its own conjugate by any element of G, and we say that N is *invariant* under conjugation:

$$gNg^{-1} = N\,,\ \ \forall\, g \in G\,.$$

Proposition A.13. *Let H and K be subgroups of G. Then, if either H or K is normal, the multiplication HK is a subgroup of G.*

Proof. Suppose $K \trianglelefteq G$. Recalling that the multiplication HK represents the collection of left cosets of K by the elements of H, the subgroups do permute:

$$HK = \bigcup_{h\in H} hK = \bigcup_{h\in H} Kh = KH\,,$$

whence, by Proposition A.11, HK is a subgroup of G.

Similarly, if H is normal in G, the same result can be shown by considering the cosets of H. □

Please notice that when H and K are both normal subgroups of G, one can easily verify that HK is also normal:

$$gHK = (Hg)K = H(gK) = HKg\,,\ \ \forall\, g \in G\,.$$

Proposition A.14. *Let H and K be normal subgroups of G with $H \cap K = e$, where e is the identity element of G. Then, the elements of H commute with the elements of K.*

Proof. Consider $h \in H$ and $k \in K$. The condition $H \trianglelefteq G$ implies that $k^{-1}hk$ is in H, as well as $h^{-1}(k^{-1}hk)$. Similarly, being $K \trianglelefteq G$, the conjugate $h^{-1}k^{-1}h$ of k^{-1} is in K, and so is $(h^{-1}k^{-1}h)k$.

Hence, being $H \cap K = e$, one derives

$$h^{-1}k^{-1}hk = (kh)^{-1}(hk) = e \in G\,,$$

what results in the following identity:

$$hk = kh\,, \;\; \forall\, h \in H\,, \; k \in K\,. \qquad \square$$

Definition A.15. Given a normal subgroup N of G, let us denote by G/N the set of all the cosets of N:

$$G/N = \{\, gN = Ng \mid g \in G \,\}\,.$$

Such a set is actually a group, called the *quotient group*, or the *factor group*, of G by N (Lang, 2002).

The multiplication of subgroups provides the law of composition of G/N. Actually, by multiplying two elements aN and bN, one has

$$aNbN = abNN = abN\,, \;\; \forall\, aN, bN \in G/N\,.$$

Since abN is itself a coset of N, G/N is closed under multiplication. Moreover, $N = eN$ is the identity of G/N and the inverse of an element aN is $a^{-1}N$.

The normal subgroups of G are strictly related to the homomorphisms having G as domain.

Let us first consider the groups G and G', with the relevant unit elements e and e', and a group homomorphism $f\colon G \to G'$. The *kernel* of f is the subset of G mapped to the identity element:

$$\mathrm{Ker}(f) = \{\, x \in G \mid f(x) = e' \,\} \subseteq G\,. \tag{A.8}$$

Please observe that $\mathrm{Ker}(f)$ satisfies the group axioms of Definition A.8, which makes the kernel of f a subgroup of G:

- the identity element e is in $\mathrm{Ker}(f)$:

$$f(e) = e' ;$$

- the kernel is closed under the group law of composition:

$$f(xy) = f(x)f(y) = e'e' = e' , \ \forall\, x, y \in \mathrm{Ker}(f) ;$$

- the kernel is closed under inversion:

$$f(x^{-1}) = \big(f(x)\big)^{-1} = (e')^{-1} = e' , \ \forall\, x \in \mathrm{Ker}(f) .$$

Proposition A.16. *Let G and G' be groups, with the identity elements e and e', respectively, and let $f : G \to G'$ be a group homomorphism. Then, $\mathrm{Ker}(f)$ is a normal subgroup of G.*

Proof. Let us set $K = \mathrm{Ker}(f)$ and consider the conjugation gKg^{-1} by an arbitrary element of G. Since f is a group homomorphism, one has

$$f(gKg^{-1}) = f(g)f(K)f(g)^{-1} = f(g)e'f(g)^{-1} = f(g)f(g)^{-1} = e' , \quad \forall\, g \in G .$$

Then, being the preimage of e', the set gKg^{-1} is exactly the kernel of f:

$$K = gKg^{-1} , \ \forall\, g \in G ,$$

implying the invariance of $\mathrm{Ker}(f)$ by conjugation and the normality as a subgroup of G. □

The converse also applies, meaning that any normal subgroup is the kernel of some group homomorphism.

Proposition A.17. *Let G be a group and N a normal subgroup. Then, there exists a group homomorphism whose kernel is given by N.*

Proof. The required group homomorphism is given by

$$\begin{aligned} \pi:\ G &\to G/N \\ g &\mapsto gN\,, \end{aligned} \tag{A.9}$$

where G/N is the quotient group introduced by Definition A.15.

First observe that, since N is normal in G, the map π is actually a group homomorphism:

$$\pi(ab) = abN = abNN = aNbN = \pi(a)\pi(b)\,, \quad \forall\, a, b \in G\,.$$

Then, the kernel of π can be found by considering the elements mapped to $eN = N$:

$$\mathrm{Ker}(\pi) = \{\, g \in G \mid \pi(g) = N \,\}\,.$$

Consequently, since an element g satisfies the condition $\pi(g) = gN = N$ if, and only if, it is in N, the kernel of π is given exactly by N:

$$\mathrm{Ker}(\pi) = N\,. \qquad \square$$

The homomorphism π defined by (A.9) is called the *canonical map* of G to G/N or the *natural projection* of G onto G/N. Clearly, N is also the kernel of any other homomorphism whose codomain is isomorphic to G/N.

Remark. Defining the kernel of the group homomorphism f by (A.8) is consistent with the definition (A.4) as a congruence relation.

In fact, the fiber of f over e', say $K = f^{-1}(e')$, provides the congruence class of the identity e. Then, the class of any other element $g \in G$ is given by the coset $gK = Kg$, whence the entire congruence relation can be recovered.

One can also say that the quotient group G/K is the collection of the fibers of f over the elements of $\mathrm{Im}(f)$. Then, the kernel K is a single element of the group G/K, and the other elements are translates, as cosets, of K.

Let us again consider the group homomorphism $f: G \to G'$, with $K = \mathrm{Ker}(f)$, and let $\pi: G \to G/K$ be the natural projection of G onto G/K.

Then, there exists a unique monomorphism $\bar{f}\colon G/K \to G'$ such that

$$f = \bar{f} \circ \pi\,.$$

We recall that G/K collects the cosets of K and $f(gK) = f(g)$ for any element $g \in G$. Then, the following map is well-defined:

$$\begin{aligned} \bar{f}:\ & G/K \to G' \\ & gK \mapsto f(g) \end{aligned}$$

Since K is normal and being $hK = K$ for all h in K, it results

$$\bar{f}(xKyK) = \bar{f}(xyK) = f(xy) = f(x)f(y) = \bar{f}(xK)\bar{f}(yK)\,, \qquad \forall\, xK, yK \in G/K\,,$$

so that $\bar{f}$ is actually a group homomorphism.

Furthermore, considering the preimage of the identity $e' \in G'$, we derive

$$\mathrm{Ker}(\bar{f}) = \left\{\, gK \in G/K \;\middle|\; \bar{f}(gK) = f(g) = e' \,\right\},$$

and, since the fiber of f over e' is exactly the kernel of f, we obtain $\mathrm{Ker}(\bar{f}) = K$. On the other hand, K is the identity element in the group G/K, so $\bar{f}$ has a trivial kernel and is injective.

The map $\bar{f}$ is the unique homomorphism satisfying the required properties. Actually, because of its injectivity, $\bar{f}$ is the unique map whose fiber over $f(g)$ provides $gK \in G/K$ for any $g \in G$. In addition, considering the restriction of G' to its subgroup $f(G)$, the monomorphism $\bar{f}$ induces the following isomorphism

$$\begin{aligned} \varphi:\ & G/K \to \mathrm{Im}(f) \\ & gK \mapsto \bar{f}(gK) = f(g) \end{aligned}$$

In conclusion, for any group homomorphism $f : G \to G'$, the image is isomorphic to the quotient group of the domain by the kernel:

$$\mathrm{Im}(f) \cong G/\,\mathrm{Ker}(f)\,.$$

Direct and Semidirect Product

Let G_1 and G_2 be groups with operations $*_1$ and $*_2$. The *direct product* of G_1 and G_2 is the group whose underlying set is the Cartesian product $G = G_1 \times G_2$ with the law of composition defined component-wise (Dummit and Foote, 2003):

$$(x_1, x_2) * (y_1, y_2) = (x_1 *_1 y_1, x_2 *_2 y_2)\,, \ \ \forall\, x_1, y_1 \in G_1\,,\ x_2, y_2 \in G_2\,.$$

The above notation can be simplified as follows:

$$(x_1, x_2)(y_1, y_2) = (x_1 y_1, x_2 y_2)\,, \ \ \forall\, x_1, y_1 \in G_1\,,\ x_2, y_2 \in G_2\,,$$

with the clear meaning that each group multiplication makes sense within the relevant algebraic structure.

Furthermore, in order to comply with the group structure consistently with the law of composition, the identity element of G is (e_1, e_2), being e_1 and e_2 the identities in G_1 and G_2, respectively. Also, from the component-wise multiplication, the inverse element of (x_1, x_2) results (x_1^{-1}, x_2^{-2}).

Let us consider the subsets H_1 and H_2 of G obtained by setting $x_2 = e_2$ and $x_1 = e_1$, respectively:

$$H_1 = \{\,(x_1, e_2) \mid x_1 \in G_1\,\} \subset G\,, \quad H_2 = \{\,(e_1, x_2) \mid x_2 \in G_2\,\} \subset G\,.$$

Because of the element-wise definition of the multiplication in G, one can easily verify that both H_1 and H_2 are subgroups of G. It is also clear that H_1 and H_2 are isomorphic with G_1 and G_2, respectively:

$$G_1 \cong H_1 \subset G\,, \quad G_2 \cong H_2 \subset G\,.$$

Moreover, by construction the identity $e = (e_1, e_2)$ is the only element belonging to both the subgroups H_1 and H_2:

$$H_1 \cap H_2 = (e_1, e_2)\,.$$

Let $\varphi_1 \colon G \to H_1$ be the map defined as

$$\varphi_1(x_1, x_2) = (x_1, e_2)\,.$$

By means of the element-wise multiplication, it is possible to prove that φ_1 is a group homomorphism (Dummit and Foote, 2003). Specifically, since for any $(x_1, e_2) \in H_1$ there exist some $(x_1, x_2) \in G$ mapped to (x_1, e_2) by φ_1, it is clear that φ_1 is surjective, that is

$$\mathrm{Im}(\varphi_1) = H_1\,.$$

In addition, it can be noticed that $(x_1, x_2) \in G$ is mapped to the identity $(e_1, e_2) \in H_1$ if, and only if, we set $x_1 = e_1$. This is the same condition identifying the subgroup H_2, so that the kernel of φ_1 coincides with the subgroup H_2:

$$\mathrm{Ker}(\varphi_1) = H_2\,,$$

and, being $\mathrm{Ker}(\varphi_1)$ normal in G, one obtains

$$G/H_2 \cong \mathrm{Im}(\varphi_1) = H_1 \cong G_1\,.$$

An analogous result can be proved by considering the map $\varphi_2 \colon (x_1, x_2) \mapsto (e_1, x_2)$ from G to H_2. Explicitly, the kernel of φ_2 is exactly H_1:

$$\mathrm{Ker}(\varphi_2) = H_1\,,$$

and the quotient group is isomorphic with H_2:

$$G/H_1 \cong \mathrm{Im}(\varphi_2) = H_2 \cong G_2\,.$$

A further property of the subgroups H_1 and H_2 is that every element of the one commutes with the elements of the other:

$$h_1 h_2 = h_2 h_1\,, \;\; \forall\, h_1 \in H_1\,,\; h_2 \in H_2\,.$$

Such a property is a consequence of the element-wise definition for the law of composition in G. In fact, by setting $h_1 = (x_1, e_2)$ and $h_2 = (e_1, x_2)$, one has

$$(x_1, e_2)(e_1, x_2) = (x_1, x_2) = (e_1, x_2)(x_1, e_2)\,, \ \forall\,(x_1, e_2) \in H_1\,, \ (e_1, x_2) \in H_2\,.$$

In summary, constructing a group G as a direct product of two groups G_1 and G_2 implies that there exist two subgroups H_1 and H_2 in G isomorphic to G_1 and G_2, respectively, whose intersection is the identity e and that are both normal in G. Also, each of the subgroups is isomorphic to the quotient group of G by the other one.

We want to point out that the results so far presented concern the direct product of two groups. Similar properties can be proved defining the direct product of an arbitrary number of groups. A discussion of such more general case can be found, e.g., in Dummit and Foote (2003).

Recognizing Direct Products We have seen how a group G can be constructed as the direct product of two given groups G_1 and G_2, along with the relations between the assigned groups and the resulting one.

Now, we want to introduce a criterion to decompose a given group as the direct product of some of its subgroups (Dummit and Foote, 2003).

Lemma A.18. *Let H and K be subgroups of a group G, with identity e, such that $H \cap K = e$. Then, each element of the group multiplication HK can be uniquely written as a product hk, with $h \in H$ and $k \in K$.*

Proof. Let us consider an element $g \in HK$ and let h and k be some elements of H and K, respectively, such that

$$g = hk \in HK\,.$$

Suppose that there exist some $h_1 \in H$ and $k_1 \in K$, other than h and k, whose multiplication provides the element g:

$$hk = h_1 k_1 \in HK\,.$$

Then, the right multiplication by k^{-1} and the left multiplication by h_1^{-1} imply

$$h_1^{-1} h = k_1 k^{-1} \in G\,,$$

where $h_1^{-1}h$ is in H, while k_1k^{-1} is in K.

Since $H \cap K = e$, the above relation holds true if and only if both the left hand side and the right one equal the identity element of G, that is

$$h_1^{-1}h = e\,, \quad k_1k^{-1} = e\,.$$

whence $h_1 = h \in H$ and $k_1 = k \in K$.

In conclusion, the pair of elements $h \in H$ and $k \in K$ providing $g \in HK$ is unique. □

Theorem A.19 (Recognition Theorem)**.** *Let G be a group with subgroups H and K such that H and K are normal in G and $H \cap K = e$, with e the identity element of G. Then, the group multiplication of H and K is isomorphic to the direct product:*

$$HK \cong H \times K\,.$$

Proof. Since $H \cap K = e$, by Lemma A.18 any element of HK can be uniquely written in the form hk, being $h \in H$ and $k \in K$. For this reason, the following map is well-defined:

$$\begin{aligned} \varphi :\ HK &\to H \times K \\ hk &\mapsto (h, k) \end{aligned}$$

In order to show the homomorphic structure of φ, let us consider the elements $h_1, h_2 \in H$ and $k_1, k_2 \in K$, so that both $g_1 = h_1k_1$ and $g_2 = h_2k_2$ are in HK. Moreover, by Proposition A.14, h_1 and h_2 do commute with k_1 and k_2, so that the multiplication of g_1g_2 is the element $(h_1h_2)(k_1k_2) \in HK$. Hence, the image of g_1g_2 is

$$\varphi\big(g_1g_2\big) = \varphi\big((h_1h_2)(k_1k_2)\big) = (h_1, h_2)(k_1, k_2)\,.$$

At the same time, the component-wise multiplication in $H \times K$ applies as follows:

$$(h_1, h_2)(k_1, k_2) = (h_1k_1, h_2k_2) = \varphi(g_1)\varphi(g_2)\,.$$

The comparison of the above relations implies

$$\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2)\,, \;\; \forall\, g_1, g_2 \in HK\,,$$

proving that φ is a homomorphism.

The surjectivity of φ is trivial, since the pairs (h, k) are constructed from the elements of H and K, and the same holds for the members of HK. In addition, the uniqueness of the decomposition hk for any element of HK ensures that φ is injective, proving the theorem. □

Semidirect Product

We have seen how the direct product of two groups allows one to construct a new group. The primary feature emerging from this construction is the normality, up to an isomorphism, of the assigned groups. Conversely, a group can be thought as the direct product of two subgroups, provided that such subgroups are normal.

The notion of semidirect product comes from the idea of relaxing the requirement for the subgroups to be both normal, what can be achieved by properly defining a law of composition involving the action of a subgroup on the other one (Dummit and Foote, 2003).

Definition A.20. Let H and K be groups and let φ be a homomorphism from K into $\mathrm{Aut}(H)$. We define the *semidirect product* of K and H with respect to φ the group, denoted as $G = H \rtimes_\varphi K$, whose underlying set is the Cartesian product $H \times K$ with the following law of composition:

$$(h_1, k_1) * (h_2, k_2) = \big(h_1 *_H \varphi_{k_1}(h_2), k_1 *_K k_2\big)\,, \;\; \forall\, h_1, h_2 \in H\,,\; k_1, k_2 \in K\,,$$

where $\varphi_{k_1} = \varphi(k_1) \in \mathrm{Aut}(H)$ is the representation of the group action φ.

When the group operations are clear from the context, the above notation can be simplified in

$$(h_1, k_1)(h_2, k_2) = (h_1 \varphi_{k_1}(h_2), k_1 k_2)\,, \;\; \forall\, h_1, h_2 \in H\,,\; k_1, k_2 \in K\,.$$

Moreover, if there is no risk of ambiguity, the resulting group can be denoted simply as $G = H \rtimes K$, where the explicit reference to φ is omitted. In any case, the notation points out how the construction of the semidirect product is not symmetric, meaning that the law of composition relies on the group action of K on H.

Let us also observe that one can refer to the semidirect product of K and H by writing $K \ltimes_\varphi H$, or simply $K \ltimes H$. In that case, the group operation is defined by switching the roles of the first and the second component and the resulting group is the same up to an isomorphism.

In order to ensure that Definition A.20 makes sense, it is necessary that the set $H \rtimes_\varphi K$, along with the associated law of composition, actually has a group structure. This is the purpose of the following theorem (Dummit and Foote, 2003).

Theorem A.21. *Let H and K be groups and let $\varphi \colon K \to \mathbf{Aut}(H)$ be a given homomorphism. Consider the set G made by the pairs (h,k), with $h \in H$ and $k \in K$, and define a binary operation $G \times G \to G$ as follows:*

$$(h_1, k_1) * (h_2, k_2) = \big(h_1 \varphi_{k_1}(h_2), k_1 k_2\big)\,, \ \ \forall\, h_1, h_2 \in H\,,\ k_1, k_2 \in K\,,$$

with $\varphi_{k_1} = \varphi(k_1)$.

Then, this operation makes G a group.

Proof. Let (h_1, k_1), (h_2, k_2) and (h_3, k_3) be elements of G and verify the associativity of the introduced operation:

$$\begin{aligned}\big((h_1, k_1) * (h_2, k_2)\big) * (h_3, k_3) &= \big(h_1 \varphi_{k_1}(h_2), k_1 k_2\big) * (h_3, k_3) \\ &= \big(h_1 \varphi_{k_1}(h_2) \varphi_{k_1 k_2}(h_3), k_1 k_2 k_3\big)\,.\end{aligned}$$

The associativity of the second component results from the group structure of K. Regarding the first component, let us observe that the homomorphism φ satisfies

$$\varphi_{k_1 k_2}(h_3) = \big(\varphi_{k_1} \circ \varphi_{k_2}\big)(h_3) = \varphi_{k_1}\big(\varphi_{k_2}(h_3)\big)\,.$$

Then, being $\varphi_{k_1} \in \mathbf{Aut}(H)$, one also has

$$\varphi_{k_1}(h_2) \varphi_{k_1}\big(\varphi_{k_2}(h_3)\big) = \varphi_{k_1}\big(h_2 \varphi_{k_2}(h_3)\big)\,,$$

finally obtaining

$$\begin{aligned}\big((h_1,k_1)*(h_2,k_2)\big)*(h_3,k_3) &= \big(h_1\varphi_{k_1}\big(h_2\varphi_{k_2}(h_3)\big),k_1k_2k_3\big)\\ &= \big(h_1,k_1\big)*\big(h_2\varphi_{k_2}(h_3),k_2k_3\big)\\ &= \big(h_1,k_1\big)*\big((h_2,k_2)*((h_3,k_3)\big)\,.\end{aligned}$$

It is easy to show that the identity element of G is (e_H, e_K), with e_H and e_K the identity elements of H and K, respectively. Actually, for the homomorphic structure of φ, one has $\varphi_{e_K} = \mathbf{id}_H$. On the other hand, since φ_k is an automorphism of H, it also results $\varphi_k(e_H) = e_H$. Then

$$(h,k)*(e_H,e_K) = \big(h\varphi_k(e_H),k\big) = (h,k)\,,\ \ \forall\,(h,k)\in G\,,$$

and also

$$(e_H,e_K)*(h,k) = \big(e_H\varphi_{e_K}(h),k\big) = (h,k)\,,\ \ \forall\,(h,k)\in G\,.$$

Finally, the inverse of an element (h,k) of G results

$$(h,k)^{-1} = \big(\varphi_k^{-1}(h^{-1}),k^{-1}\big)\,,\ \ \forall\,(h,k)\in G\,,$$

what can be proved by applying the definition of inverse:

$$(h,k)*\big(\varphi_k^{-1}(h^{-1}),k^{-1}\big) = \big(h\big(\varphi_k\circ\varphi_k^{-1}\big)(h^{-1}),kk^{-1}\big) = (hh^{-1},e_K) = (e_H,e_K)\,,$$

as well as, recalling that $\varphi_k^{-1} = \varphi_{k^{-1}}$,

$$\big(\varphi_k^{-1}(h^{-1}),k^{-1}\big)*(h,k) = \big(\varphi_k^{-1}(h^{-1})\varphi_{k^{-1}}(h),k^{-1}k\big) = (h^{-1}h,e_K) = (e_H,e_K)\,.\square$$

The semidirect product of two groups has a peculiar structure. Specifically, the resulting group contains isomorphic copies of the assigned groups satisfying specific properties. This is shown in the following theorem.

Theorem A.22. *Let $G = H \rtimes_\varphi K$ be the semidirect product of K and H with respect to φ, and define the subsets $\overline{H}$ and $\overline{K}$ as*

$$\overline{H} = \{\,(h,e_K)\mid h\in H\,\}\subseteq G\,,\quad \overline{K} = \{\,(e_H,k)\mid k\in K\,\}\subseteq G\,.$$

Then, the following statements hold true.

1. $\overline{H}$ *and* $\overline{K}$ *are subgroups of* G *and they are isomorphic with* H *and* K*, respectively.*
2. $\overline{H} \cap \overline{K} = (e_H, e_K) = e$.
3. $\overline{H}$ *is normal in* G.
4. $\overline{K}$ *acts on* $\overline{H}$ *by conjugation and such an action is isomorphic to* φ.

Proof. 1. The subset $\overline{H} \subseteq G$ can be thought as the Cartesian product $H \times \{e_K\}$. Similarly, we can regard the subset $\overline{K} \subseteq G$ as the Cartesian product $\{e_H\} \times K$. Then, the following maps are clearly bijective:

$$\begin{aligned} \pi_H :\ & H \to \overline{H} & \qquad \pi_K :\ & K \to \overline{K} \\ & h \mapsto (h, e_K) & & k \mapsto (e_H, k) \end{aligned} \tag{A.10}$$

Moreover, recalling that the homomorphism φ satisfies the property $\varphi_{e_K} = \mathbf{id}_H$, it is easy to verify that $\overline{H}$ has a group structure under the same law of composition of G:

$$(h_1, e_K)(h_2, e_K) = (h_1 h_2, e_K) \in \overline{H}\,,\ \ \forall\,(h_1, e_K), (h_2, e_K) \in \overline{H}\,.$$

The above relation can also be written in terms of the map π_H as

$$\pi_H(h_1)\pi_H(h_2) = \pi_H(h_1 h_2)\,,\ \ \forall\, h_1, h_2 \in H\,,$$

showing that π_H is a group isomorphism.

On the other hand, since φ_k is an automorphism of H for any k of K, the identity is mapped to itself, i.e. $\varphi_k(e_H) = e_H$. Then, the group operation of G makes $\overline{K}$ also a subgroup, resulting

$$(e_H, k_1)(e_H, k_2) = (e_K, k_1 k_2) \in \overline{K}\,,\ \ \forall\,(e_H, k_1), (e_H, k_2) \in \overline{K}\,,$$

what also proves the homomorphic structure of the bijection π_K:

$$\pi_K(k_1)\pi_K(k_2) = \pi_K(k_1 k_2)\,,\ \ \forall\, k_1, k_2 \in K\,.$$

2. The component-wise comparison of $(h, e_K) \in \overline{H}$ and $(e_H, k) \in \overline{K}$ shows that the condition $(h, e_K) = (e_H, k)$ is satisfied if, and only if, $h = e_H$

and $k = e_K$ both hold. Hence

$$\overline{H} \cap \overline{K} = (e_H, e_K)\,.$$

3. Consider the left coset of $\overline{H}$ by an arbitrary element (h,k) of G:

$$(h,k)\overline{H} = \bigcup_{l \in H} (h,k)(l, e_K) = \bigcup_{l \in H} \big(h\varphi_k(l), k\big)\,.$$

Since φ_k is an automorphism of H, then $\varphi_k(H) = H$ and so

$$\bigcup_{l \in H} h\varphi_k(l) = h\varphi_k(H) = hH = Hh = \bigcup_{l \in H} lh\,,$$

whence

$$(h,k)\overline{H} = \bigcup_{l \in H} (lh, k) = \bigcup_{l \in H} (l, e_K)(h,k) = \overline{H}(h,k)\,.$$

Thus the left and the right cosets of $\overline{H}$ do coincide and, by definition, $\overline{H}$ is normal in G.

4. Let $h \in H$ and $k \in K$ be mapped to $\bar{h} \in \overline{H}$ and $\bar{k} \in \overline{K}$, respectively, through the isomorphisms π_H and π_K defined by (A.10). The conjugate of $\bar{h}$ by $\bar{k}$ is

$$\begin{aligned} C_{\bar{k}}(\bar{h}) &= (e_H, k)(h, e_K)(e_H, k)^{-1} = \big(\varphi_k(h), k\big)\big(e_H, k^{-1}\big) \\ &= \big(\varphi_k(h), e_K\big)\,. \end{aligned}$$

Then, the conjugation $C_{\bar{k}}$ is an automorphism of $\overline{H}$ for any $\bar{k}$ in $\overline{K}$, representing the action of $\overline{K}$ on $\overline{H}$.
In addition, let us observe that

$$C_{\bar{k}}(\bar{h}) = C_{\bar{k}}\big(\pi_H(h)\big) = \big(\varphi_k(h), e_K\big) = \pi_H\big(\varphi_k(h)\big)\,,$$

whence

$$C_{\bar{k}} \circ \pi_H = \pi_H \circ \varphi_k \quad \Leftrightarrow \quad \varphi_k = \pi_H^{-1} \circ C_{\bar{k}} \circ \pi_H\,.$$

This shows how the group action φ of K on H coincides, up to the isomorphisms π_H and π_K, with the action by conjugation of $\overline{K}$ on $\overline{H}$. $\square$

Similarly to the direct product, it is possible recognize the structure of an assigned group depending on the features of its subgroups. To this end, we prove now a recognition theorem for the semidirect product.

Theorem A.23. *Let G be a group with identity e. Suppose H and K are subgroups such that H is normal in G and $H \cap K = e$. Then, the group multiplication of H and K is isomorphic to the semidirect product with respect to the map C:*

$$HK \cong H \rtimes_C K\,,$$

where C is the action by conjugation of K on H:

$$\begin{aligned} C:\ K \times H &\to H \\ (k,h) &\mapsto C_k(h) = khk^{-1}\,. \end{aligned}$$

Proof. Since $H \trianglelefteq G$, the subgroup H is invariant under conjugation (see Definition A.12). Consequently, any representation C_k of the action C is an automorphism of H, resulting

$$C_k(H) = kHk^{-1} = H\,,\ \ \forall\, k \in K \subseteq G\,,$$

and the semidirect product $H \rtimes_C K$ is well-defined.

By Proposition A.13, the normality of H also implies that the group multiplication HK is a subgroup of G. Moreover, being $H \cap K = e$, by Lemma A.18 any element of HK can be written uniquely in the form hk, for some $h \in H$ and $k \in K$. Then, the definition of the following map makes sense:

$$\begin{aligned} \pi:\ HK &\to H \rtimes_C K \\ hk &\mapsto (h,k)\,. \end{aligned}$$

We recall that the underlying set of the group $H \rtimes_C K$ is the Cartesian product $H \times K$. Hence, since any pair of elements $h \in H$ and $k \in K$ uniquely defines the element hk of HK and (h,k) in $H \times K$, the map π is clearly a set bijection.

To prove that π is also a group homomorphism, let us consider the product of two elements h_1k_1 and h_2k_2 of HK:

$$(h_1k_1)(h_2k_2) = h_1k_1h_2(k_1^{-1}k_1)k_2 = h_1(k_1h_2k_1^{-1})k_1k_2 = \big(h_1C_{k_1}(h_2)\big)(k_1k_2)\,,$$

where the resulting elements $h_1C_{k_1}(h_2) \in H$ and $k_1k_2 \in K$ are consistent with the law of composition induced by the action C on the group $H \rtimes_C K$. Then,

$$\pi\big((h_1k_1)(h_2k_2)\big) = \pi\big(h_1C_{k_1}(h_2)\big)\pi(k_1k_2)\,, \ \ \forall\, h_1k_1, h_2k_2 \in HK$$

and the bijection π results a group isomorphism. □

In the same hypotheses of the above theorem, if any element of G can be written in the form hk, with $h \in H$ and $k \in K$, the group G itself coincides with the multiplication HK and is isomorphic with the semidirect product of its subgroups:

$$G \cong H \rtimes K\,,$$

where it is understood the action by conjugation of K on H.

Moreover, each of the subgroups H and K also results the complement of the other one, according to the following definition (Dummit and Foote, 2003).

Definition A.24. Let H be a subgroup of a group G. A subgroup K of G is called a *complement* for H in G if $G = HK$ and $H \cap K = e$.

A.2.3 Rings

A *ring* is a set R along with two laws of composition, called *addition* $(+)$ and *multiplication* $(\cdot)$, such that $(R,+)$ is an abelian group, $(R,\cdot)$ is a monoid and multiplication distributes over addition.

Specifically, the ring structure holds if the laws of composition do fulfill the following properties:

- addition is associative:

$$a + (b + c) = (a + b) + c\,, \ \ \forall\, a, b, c \in R\,;$$

- there exists an element $0 \in R$ called the *additive identity*:

$$a + 0 = 0 + a = a\,, \ \ \forall\, a \in R\,;$$

- for any element there exists an *additive inverse*, or *opposite*, in R:

$$a + (-a) = (-a) + a = 0\,, \ \ \forall\, a \in R\,;$$

- addition is commutative:

$$a + b = b + a\,, \ \ \forall\, a, b \in R\,;$$

- multiplication is associative:

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c\,, \ \ \forall\, a, b, c \in R\,;$$

- there exists $1 \in R$ called the *multiplicative identity*:

$$a \cdot 1 = 1 \cdot a = a\,, \ \ \forall\, a \in R\,;$$

- multiplication is distributive with respect to addition:

$$a \cdot (b + c) = a \cdot b + a \cdot c\,, \ \ (a + b) \cdot c = a \cdot c + b \cdot c\,, \ \ \forall\, a, b, c \in R\,.$$

Multiplication can be represented by juxtaposition and the symbol '$\cdot$' is usually omitted, so that ab stands for $a \cdot b$ with $a, b \in R$.

It is observed in Dummit and Foote (2003) how the commutativity of addition is actually a consequence of distributive laws and existence of 1. In fact, by applying left and right distributivity to $(1 + 1)(a + b)$, one has

$$(1 + 1)(a + b) = (1 + 1)a + (1 + 1)b = a + a + b + b\,,$$
$$(1 + 1)(a + b) = 1(a + b) + 1(a + b) = a + b + a + b\,,$$

whence

$$a + a + b + b = a + b + a + b \,.$$

By adding $-a$ to the left and $-b$ to the right of both members, one can derive the commutative property, i.e. $a + b = b + a$.

A further property, shown in Lang (2002), is that $0a = 0$ for all $a \in R$. Such a property comes from the following relation:

$$0a + a = (0 + 1)a = 1a = a \,.$$

The immediate consequence is that if 0 is both the additive and the multiplicative identity, explicitly $1 = 0$, then

$$0a = a = 0 \,, \;\; \forall \, a \in R \,,$$

and R consists of 0 alone, named the *zero ring*. Hence, $R \neq \{0\}$ requires $1 \neq 0$.

Also, for any $a, b \in R$ we have $(-a)b = a(-b) = -ab$. In fact, left and right distributive property implies

$$ab + a(-b) = a(b - b) = a0 = 0 \,,$$
$$ab + (-a)b = (a - a)b = 0b = 0 \,,$$

so that, by definition, both $a(-b)$ and $(-a)b$ coincide with the opposite of ab. The immediate consequence is $(-a)(-b) = ab$.

It is worth noting that $(R, \cdot)$ is introduced as a (non-commutative) monoid. If commutativity is additionally assumed for multiplication, i.e. $ab = ba$ for all $a, b \in R$, we say that R is a *commutative ring*.

Units and zero divisors

Let R be a non-zero ring and let us consider an element u which has both a left inverse, v, and a right one, w. By associativity, we observe that $vuw = (vu)w = v(uw)$, i.e. $v = w = u^{-1}$. We say that u is a *unit* in R.

The set of all the units in R, denoted by $R^{\times}$, is a group under multiplication and is called the *group of units*, or the *group of invertible elements* of R.

If any non-zero elements of R is invertible, i.e. $R^{\times} = R\backslash\{0\}$, then it is called a *division ring*.

A non-zero element a of R is called a *zero divisor* if there exists an element $b \neq 0$ in R such that $ab = 0$ or $ba = 0$.

Please notice that a zero divisor can never be a unit, and vice-versa. To prove this property, let us consider a unit $a \in R$ and suppose that $ab = 0$ for some $b \neq 0$ in R. Then, there exists $v \in R$ such that $va = 1$ and we can write

$$b = 1b = (va)b = v(ab) = v0 = 0\,,$$

what contradicts the assumption $b \neq 0$.

A.2.4 Fields

A *field* is a commutative division ring. Hence, supposing a set F with addition and multiplication satisfies the ring axioms, it is a field if $(F\backslash\{0\}, \cdot)$ is an abelian group (Dummit and Foote, 2003).

The properties providing the field structure to a set F are here summarized:

- addition is associative:

$$a + (b + c) = (a + b) + c\,,\ \ \forall\, a, b, c \in F\,;$$

- there exists an element $0 \in R$ called the *additive identity*:

$$a + 0 = 0 + a = a\,,\ \ \forall\, a \in F\,;$$

- for any element there exists an *additive inverse*, or *opposite*, in F:

$$a + (-a) = (-a) + a = 0\,,\ \ \forall\, a \in F\,;$$

- addition is commutative:

$$a + b = b + a\,,\ \ \forall\, a, b \in F\,;$$

- multiplication is associative:

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c\,,\ \ \forall\, a, b, c \in F\,;$$

- there exists $1 \in F$ called the *multiplicative identity*:

$$a \cdot 1 = 1 \cdot a = a\,,\ \ \forall\, a \in F\,;$$

- for any non-zero element there exists a *multiplicative inverse*, or simply an *inverse*, in F:

$$a \cdot a^{-1} = a^{-1} \cdot a = 1\,,\ \ \forall\, a \neq 0 \in F\,;$$

- multiplication is commutative:

$$a \cdot b = b \cdot a\,,\ \ \forall\, a, b \in F\,;$$

- multiplication distributes over addition:

$$a \cdot (b + c) = (b + c) \cdot a = a \cdot b + a \cdot c\,,\ \ \forall\, a, b, c \in F\,.$$

A field F inherits from the ring structure all its properties. As an example, $a0 = 0$ for any a in F, as well as $1 \neq 0$ unless $F = \{0\}$. The property $(-a)b = a(-b) = -ab$ also hold for all elements $a, b \in F$.

Ordered Fields

Let $\mathbb{F}$ be a field. An *ordering* of $\mathbb{F}$ is a subset $\mathbb{P}$ of $\mathbb{F}$ such that the following properties are satisfied:

- given any $x \in \mathbb{F}$, either x is in $\mathbb{P}$, or $x = 0$, or $-x$ is in $\mathbb{P}$, and these possibilities are mutually exclusive;

- if x and y are in $\mathbb{P}$, then $x+y$ and xy are both in $\mathbb{P}$.

The subset $\mathbb{P}$ is the set of *positive* elements of $\mathbb{F}$, while $-\mathbb{P}$ is the set of *negative* elements of $\mathbb{F}$, that is the subset of x such that $-x \in \mathbb{P}$. Then, the ordered field $\mathbb{F}$ is the union of the subset $\mathbb{P}$, the additive identity 0, and the subset $-\mathbb{P}$:

$$\mathbb{F} = \mathbb{P} \cup \{0\} \cup (-\mathbb{P}) .$$

Given $x, y \in \mathbb{F}$, the relation $x < y$, or equivalently $y > x$, means that $y - x$ is in $\mathbb{P}$. Then, the positive elements, i.e. $x \in \mathbb{P}$, satisfy $x > 0$, while the negative ones, that is $x \in -\mathbb{P}$, are the elements satisfying $x < 0$.

Please notice that the multiplicative identity 1 is positive. Actually, being other than 0, the identity 1 is either positive or negative, and in both cases one has

$$1^2 = (-1)^2 = 1 > 0 .$$

Moreover, if $x \in \mathbb{P}$ and $x \neq 0$, then $xx^{-1} = 1 > 0$, whence $x^{-1} \in \mathbb{P}$. In addition, if $x \neq 0$, the square x^2 is positive because $x^2 = (-x)^2$ and either $x \in \mathbb{P}$ or $-x \in \mathbb{P}$.

The basic example of an ordered field is the field of real numbers $\mathbb{R}$. Further details about ordered fields can be found, e.g., in Lang (2002).

A.2.5 Vector Spaces

The algebraic structures so far introduced refer to just one set endowed with a number of operations. On the contrary, a vector space represents a composite system involving two sets and two binary operations.

Definition A.25. Let $(\mathcal{V}, +)$ an additive abelian group and $(\mathbb{F}, +, \cdot)$ be a field. We say that $\mathcal{V}$ is an $\mathbb{F}$-*vector space*, or a *vector space* over $\mathbb{F}$, if it is defined a map called *scalar multiplication*

$$\begin{aligned} \mu : \ & \mathbb{F} \times \mathcal{V} \to \mathcal{V} \\ & (a, \mathbf{v}) \mapsto \mu(a, \mathbf{v}) = a\mathbf{v} , \end{aligned}$$

satisfying the following axioms:

- compatibility of scalar multiplication with field multiplication:

$$a(b\mathbf{u}) = (ab)\mathbf{u}\,, \ \ \forall\, \mathbf{u} \in \mathcal{V}\,,\ a, b \in \mathbb{F}\,;$$

- existence of identity element of scalar multiplication:

$$1\mathbf{u} = \mathbf{u}\,, \ \ \forall\, \mathbf{u} \in \mathcal{V}\,,$$

  where 1 is the multiplicative identity in $\mathbb{F}$;

- distributivity of scalar multiplication with respect to addition in $\mathcal{V}$:

$$a(\mathbf{u}+\mathbf{v}) = a\mathbf{u} + a\mathbf{v}\,, \ \ \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{V}\,,\ a \in \mathbb{F}\,;$$

- distributivity of scalar multiplication with respect to field addition:

$$(a+b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}\,, \ \ \forall\, \mathbf{u} \in \mathcal{V}\,,\ a, b \in \mathbb{F}\,.$$

Please notice that we use the same symbol "+" to represent both the addition in the field $\mathbb{F}$ and the group operation of $\mathcal{V}$.

We also say that the elements $a \in \mathbb{F}$ are *scalars*, while the elements $\mathbf{v} \in \mathcal{V}$ are called *vectors* and the law of composition in $\mathcal{V}$ is called *vector addition*.

The properties of the vector addition in $\mathcal{V}$, as an abelian group, are here summarized:

- associativity:

$$\mathbf{u} + (\mathbf{v}+\mathbf{w}) = (\mathbf{u}+\mathbf{v}) + \mathbf{w}\,, \ \ \forall\, \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}\,;$$

- commutativity:

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}\,, \ \ \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{V}\,;$$

- existence of additive identity:

$$\mathbf{u}+\mathbf{o}=\mathbf{o}+\mathbf{u}=\mathbf{u}\,,\ \ \forall\,\mathbf{u}\in\mathcal{V}\,,$$

  where $\mathbf{o}\in\mathcal{V}$ is called the *zero vector* or *null vector*;

- existence of additive inverse:

$$\mathbf{u}+(-\mathbf{u})=-\mathbf{u}+\mathbf{u}=\mathbf{o}\,,\ \ \forall\,\mathbf{u}\in\mathcal{V}\,,$$

  with $-\mathbf{u}\in\mathcal{V}$.

In the sequel we will also adopt the notation $\mathbf{v}-\mathbf{w}$ to represent the addition of $\mathbf{v}$ with the opposite of $\mathbf{w}$:

$$\mathbf{v}-\mathbf{w}=\mathbf{v}+(-\mathbf{w})\,,\ \ \forall\,\mathbf{v},\mathbf{w}\in\mathcal{V}\,,$$

as well as the division by a non-null scalar in $\mathbb{F}$ will stand for the scalar multiplication by its inverse:

$$\frac{\mathbf{v}}{c}=c^{-1}\mathbf{v}\,,\ \ \forall\,\mathbf{v}\in\mathcal{V}\,,\ c\in\mathbb{F}\backslash\{0\}\,.$$

The compatibility of the scalar multiplication with the field structure of $\mathbb{F}$ derives from the ring homomorphism from $\mathbb{F}$ to $\mathsf{End}(\mathcal{V})$.

Proposition A.26. *Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$ and let the null scalar 0 be additive identity of $\mathbb{F}$.*

1. *The scalar multiplication of the of 0 with any vector of $\mathcal{V}$ is the null vector:*

$$0\mathbf{v}=\mathbf{o}\,,\ \ \forall\,\mathbf{v}\in\mathcal{V}\,.$$

2. *The scalar multiplication of any scalar of $\mathbb{F}$ with the null vector $\mathbf{o}$ is the null vector:*

$$c\mathbf{o}=\mathbf{o}\,,\ \ \forall\,c\in\mathbb{F}\,.$$

Proof.

1. Considering an arbitrary vector $\mathbf{v} \in \mathcal{V}$, along with its opposite $-\mathbf{v}$, the following identity holds:

$$0\mathbf{v} = 0\mathbf{v} + \mathbf{o} = 0\mathbf{v} + \mathbf{v} + (-\mathbf{v})\,,$$

whence, applying the property $1\mathbf{v} = \mathbf{v}$ and the distributivity of the scalar multiplication, one finds

$$0\mathbf{v} = (0+1)\mathbf{v} + (-\mathbf{v}) = 1\mathbf{v} + (-\mathbf{v}) = \mathbf{v} + (-\mathbf{v}) = \mathbf{o}\,.$$

2. Let c be an arbitrary scalar of $\mathbb{F}$. By virtue of the above statement, one has

$$c\mathbf{o} = c(0\mathbf{v}) = (c0)\mathbf{v} = 0\mathbf{v} = \mathbf{o}\,,$$

where the compatibility of the scalar multiplication has been exploited.

□

A direct consequence of the above proposition is that the identity $c\mathbf{v} = \mathbf{o}$ holds true if, and only if, either $c = 0$ or $\mathbf{v} = \mathbf{o}$. Actually, we observe that if $c = 0$, the identity $0\mathbf{v} = \mathbf{o}$ is trivially verified. Conversely, if $c \neq 0$ we can multiply by the inverse of c:

$$c^{-1}(c\mathbf{v}) = c^{-1}\mathbf{o}\,,$$

where the first member becomes $(c^{-1}c)\mathbf{v} = 1\mathbf{v} = \mathbf{v}$, and the second one is the null vector, obtaining $\mathbf{v} = \mathbf{o}$.

Please notice that the primary vector space over a field $\mathbb{F}$ is the Cartesian power $\mathbb{F}^n$:

$$\mathbb{F}^n = \{\, (a_1, \ldots, a_n) \mid a_i \in \mathbb{F}\,,\ i = 1, \ldots, n \,\}\,.$$

In fact, since $(\mathbb{F}, +)$ is an abelian group, such a structure is clearly extended to $\mathbb{F}^n$ by the component-wise addition:

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n)\,,$$

being $a_i, b_i \in \mathbb{F}$.

Moreover, exploiting the multiplication in $\mathbb{F}$, the following operation is well-defined

$$c(a_1, \ldots, a_n) = (ca_1, \ldots, ca_n)\,, \ \ \forall\, c \in \mathbb{F}(a_1, \ldots, a_n) \in \mathbb{F}^n\,,$$

what represents a scalar multiplication consistent with the requirements of Definition A.25.

The properties of vector spaces are deeper described in Appendix B.

A.2.6 Lie Algebras

When a vector space is endowed with a bilinear operation, a new algebraic structure results, and it is called an *algebra*.

The bilinear operation introduced here below is the Lie bracket and the resulting algebraic structure is the Lie algebra.

Definition A.27. A *Lie algebra* over a field $\mathbb{F}$ is an $\mathbb{F}$-vector space $\mathfrak{g}$ endowed with a skew-symmetric bilinear map

$$[\cdot,\cdot]\colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}\,,$$

which is called the *Lie bracket*, satisfying the following properties:

- bilinearity:

$$[a\mathbf{u} + b\mathbf{v}, \mathbf{w}] = a[\mathbf{u}, \mathbf{w}] + b[\mathbf{v}, \mathbf{w}]\,, \ \ \forall\, a, b \in fF\,,$$
$$[\mathbf{w}, a\mathbf{u} + b\mathbf{v}] = a[\mathbf{w}, \mathbf{u}] + b[\mathbf{w}, \mathbf{v}]\,, \ \ \forall\, a, b \in fF\,;$$

- antisymmetry:

$$[\mathbf{u}, \mathbf{v}] = -[\mathbf{v}, \mathbf{u}]\,;$$

- Jacobi identity:

$$\big[\mathbf{u}, [\mathbf{v}, \mathbf{w}]\big] + \big[\mathbf{v}, [\mathbf{w}, \mathbf{u}]\big] + \big[\mathbf{w}, [\mathbf{u}, \mathbf{v}]\big] = \mathbf{o}\,.$$

Appendix B

Algebra of Vector Spaces

B.1 Vector Space Structure

The main features of vector spaces and multilinear algebra are recalled.

Remark (Notation)**.** Following the classical references about Differential Geometry and Tensor Analysis (Dimitrienko, 2002, Lee, 2012), from now on we adopt the typical notation compatible with the Einstein summation convention, unless otherwise stated.

Specifically, the *i*-th element in a collection of vectors will be denoted by the subscript 'i':

$$\{\mathbf{v}_i\}_{i=1}^{n}, \quad \text{with } \mathbf{v}_i \in \mathcal{V},$$

and an analogous index will be used as a superscript to refer to the *i*-th scalar:

$$\{a^i\}_{i=1}^{n}, \quad \text{with } a^i \in \mathbb{F}.$$

Then, we will write $a^i \mathbf{v}_i$ to represent a summation with respect to the index i:

$$a^i \mathbf{v}_i = \sum_{i=1}^{n} a^i \mathbf{v}_i = a^1 \mathbf{v}_1 + \ldots + a^n \mathbf{v}_n \in \mathcal{V}.$$

With this convention, it is understood that a summation is performed with respect to any index that appears exactly twice, once in an upper and once in a lower position, in a monomial term. The index that is implicitly summed over is a *dummy index*.

B.1.1 Bases

Definition B.1. Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$ and $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ a subset of $\mathcal{V}$.

1. A *linear combination* of $\mathcal{S}$ is any element of $\mathcal{V}$ expressed in the form

$$a^1\mathbf{v}_1 + \ldots + a^n\mathbf{v}_n\,,$$

with $a^1, \ldots, a^n \in \mathbb{F}$.

2. The *span* of $\mathcal{S}$ is the set of all linear combinations of $\mathcal{S}$ and it is denoted as $\mathsf{Span}(\mathcal{S})$.

We observe that, because of the distributivity properties of the scalar multiplication, as introduced in Definition A.25, any linear combination of a number of vectors of $\mathcal{V}$ is itself a vector of $\mathcal{V}$.

Furthermore, the span of $\mathcal{S}$, along with the operations defined for $\mathcal{V}$, is itself a vector space over $\mathbb{F}$. This is clear from the explicit definition of the span of $\mathcal{S}$:

$$\mathsf{Span}(\mathcal{S}) = \left\{\, a^i\mathbf{v}_i = a^1\mathbf{v}_1 + \ldots + a^n\mathbf{v}_n \;\middle|\; \mathbf{v}_i \in \mathcal{S}\,,\ a^i \in \mathbb{F} \,\right\},$$

whence one can easily see that $\mathsf{Span}(\mathcal{S})$ is closed under addition. Moreover, by setting $a^i = 0$, we find that $\mathbf{o}$ is in $\mathsf{Span}(\mathcal{S})$, while for any element in the form $a^1\mathbf{v}_1 + \ldots + a^n\mathbf{v}_n$ the opposite is given by $(-a^1)\mathbf{v}_1 + \ldots + (-a^n)\mathbf{v}_n$. Then, $\mathsf{Span}(\mathcal{S})$ is an abelian additive group.

Also, the span of $\mathcal{S}$ is closed under the scalar multiplication, resulting

$$c(a^1\mathbf{v}_1 + \ldots + a^n\mathbf{v}_n) = (ca^1)\mathbf{v}_1 + \ldots + (ca^n)\mathbf{v}_n \in \mathsf{Span}(\mathcal{S})\,,\ \ \forall\, c \in \mathbb{F}\,,$$

so that $\mathsf{Span}(\mathcal{S})$ is actually a vector space.

More generally, we say that the span of $\mathcal{S}$ is a subspace of $\mathcal{V}$, according to the following definition.

Definition B.2. Let $\mathcal{V}$ be an $\mathbb{F}$-vector space. We define a *subspace* of $\mathcal{V}$ any (nonempty) subset that is a vector space with the operations inherited from $\mathcal{V}$.

Proposition B.3. *Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$. A subset $\mathcal{W} \neq \{\mathbf{o}\}$ is a subspace of $\mathcal{V}$ if, and only if, it results*

$$a\mathbf{w} + a'\mathbf{w}' \in \mathcal{W}\,, \;\; \forall\, \mathbf{w}, \mathbf{w}' \in \mathcal{W}\,,\; a, a' \in \mathbb{F}\,.$$

Proof. Recalling that any linear combination of vectors is itself a vector of the same space, if $\mathcal{W}$ has a vector space structure, the above requirement is trivially satisfied.

On the contrary, suppose that the above condition holds true. By setting $a = a' = 0$, one verifies that $\mathbf{o}$ is in $\mathcal{W}$. Also, the opposite of any vector in the form $a\mathbf{w} + a'\mathbf{w}'$ is given by $(-a)\mathbf{w} + (-a')\mathbf{w}'$. Then, $\mathcal{W}$ results an additive abelian group.

Furthermore, since the above condition is satisfied for all the elements of $\mathbb{F}$, $\mathcal{W}$ is closed under scalar multiplication, i.e.

$$c(a\mathbf{w} + a'\mathbf{w}') = (ca)\mathbf{w} + (ca')\mathbf{w}' \in \mathcal{W}\,, \;\; \forall\, c \in \mathbb{F}\,,$$

proving the vector space structure of $\mathcal{W}$. □

Definition B.4. Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$. The subset $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathcal{V}$ is *linearly independent* over $\mathbb{F}$ if there exist scalars $a^1, \ldots, a^n \in \mathbb{F}$, not all equal to 0, such that

$$a^1\mathbf{v}_1 + \ldots + a^n\mathbf{v}_n = \mathbf{o}\,.$$

Otherwise, $\mathcal{S}$ is *linearly independent.*

Please notice that any set containing the null vector $\mathbf{o}$ is linearly dependent. As an example, for the set $\{\mathbf{o}, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, the null vector is given by the following linear combination:

$$a^1\mathbf{o} + 0\mathbf{v}_2 + \ldots + 0\mathbf{v}_n = \mathbf{o}\,, \;\; \forall\, a^1 \in \mathbb{F}\,.$$

In addition, for any set $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of vectors of $\mathcal{V}$ linearly independent, the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{w}\}$ is linearly dependent if, and only if, $\mathbf{w}$ is in $\mathsf{Span}(\mathcal{S})$. Actually, if $\mathbf{w}$ is in the span of $\mathcal{S}$, it can be written as a linear combination in the form $\mathbf{w} = a^1\mathbf{v}_1 + \ldots + a^n\mathbf{v}_n$, so that

$$a^1\mathbf{v}_1 + \ldots + a^n\mathbf{v}_n - \mathbf{w} = \mathbf{o}\,,$$

whence the linear dependence.

On the contrary, supposing $\{\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{w}\}$ is linear dependent, one can write

$$a^1\mathbf{v}_1 + \ldots + a^n\mathbf{v}_n + b\mathbf{w} = \mathbf{o}\,,$$

where, being $a^1\mathbf{v}_1 + \ldots + a^n\mathbf{v}_n \neq \mathbf{o}$ for the linear independence of $\mathcal{S}$, the scalar b is non-zero. Then, dividing by b, one finds

$$\mathbf{w} = \left(-\frac{a^1}{b}\right)\mathbf{v}_1 + \ldots + \left(-\frac{a^n}{b}\right)\mathbf{v}_n\,,$$

whence $\mathbf{w} \in \mathsf{Span}(\mathcal{S})$.

Definition B.5. Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$. The subset $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of $\mathcal{V}$ is a *basis* of $\mathcal{V}$ if it is linearly independent and its span coincides with $\mathcal{V}$:

$$\mathsf{Span}(\mathcal{B}) = \mathcal{V}\,.$$

The number of elements of $\mathcal{B}$ is the *dimension* of $\mathcal{V}$.

A vector space spanned by a finite set of vectors is said to be *finite-dimensional*. Otherwise, it is said to be *infinite-dimensional*. If the vector space consists of $\mathbf{o}$ alone, then it does not have a basis and we shall say that its dimension is 0. In the sequel, whenever a vector space is considered, it is assumed to be finite-dimensional.

Proposition B.6. *Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a set of vectors of an $\mathbb{F}$-vector space. Then, $\mathcal{B}$ is a basis for $\mathcal{V}$ if, and only if, any $\mathbf{v} \in \mathcal{V}$ can be written in a unique way as a linear combination of the vectors of $\mathcal{B}$.*

Proof. Supposing $\mathcal{B}$ is a basis for $\mathcal{V}$, any vector $\mathbf{v}$ is in the span of $\mathcal{B}$ and can be written as:

$$\mathbf{v} = v^1\mathbf{b}_1 + \ldots + v^n\mathbf{b}_n\,,$$

where $v^1, \ldots, v^n$ are scalars of $\mathbb{F}$.

Assume there exists another set of scalars of $\mathbb{F}$, say $w^1, \ldots, w^n$, such that

$$\mathbf{v} = w^1\mathbf{b}_1 + \ldots + w^n\mathbf{b}_n\,.$$

Hence, by comparing the above equations, we can write

$$v^1\mathbf{b}_1 + \ldots + v^n\mathbf{b}_n = w^1\mathbf{b}_1 + \ldots + w^n\mathbf{b}_n\,,$$

whence

$$(v^1 - w^1)\mathbf{b}_1 + \ldots + (v^n - w^n)\mathbf{b}_n = \mathbf{o}\,.$$

Since $\mathcal{B}$ is linearly independent, the coefficients providing the null vector vanish, obtaining $v^i = w^i$ with $i = 1, \ldots, n$. Then, the set of scalars $v^1, \ldots, v^n \in \mathbb{F}$ providing the vector $\mathbf{v}$ is unique.

Conversely, suppose that any vector $\mathbf{v}$ of $\mathcal{V}$ can be uniquely written as a linear combination of $\mathcal{B}$:

$$\mathbf{v} = v^1\mathbf{b}_1 + \ldots + v^n\mathbf{b}_n\,,\ \ \forall\,\mathbf{v} \in \mathcal{V}\,.$$

By Definition B.1, the set of linear combinations of the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n$ is the span of $\mathcal{B}$, so that

$$\mathcal{V} = \mathsf{Span}(\mathcal{B})\,.$$

In addition, when the coefficients $v^1, \ldots, v^n$ vanish, the relevant linear combination is the null vector:

$$\mathbf{o} = 0\mathbf{b}_1 + \ldots + 0\mathbf{b}_n \in \mathcal{V}$$

Hence, since such a combination is unique, the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n$ are linearly independent and $\mathcal{B}$ is a basis of $\mathcal{V}$. □

Definition B.7. Let us fix a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ for the vector space $\mathcal{V}$ over the field $\mathbb{F}$. The scalars $v^1, \ldots, v^n \in \mathbb{F}$ providing the vector $\mathbf{v}$, by multiplication with the basis vectors, are called the *coordinates* of $\mathbf{v}$ with respect to the basis $\mathcal{B}$:

$$\mathbf{v} = v^i \mathbf{b}_i \in \mathcal{V} .$$

The coordinates of the vector $\mathbf{v}$ can be gathered into the n-tuple $\bar{\mathbf{v}} = (v^1, \ldots, v^n) \in \mathbb{F}^n$, what represents the *coordinate vector* of $\mathbf{v}$.

Actually, we have observed by the end of Section A.2.5 that $\mathbb{F}^n$ is a vector space over over $\mathbb{F}$. Moreover, since by Proposition B.6 the set of coefficients $v^1, \ldots, v^n \in \mathbb{F}$ is unique for any $\mathbf{v}$ of $\mathcal{V}$, the following map results a vector space isomorphism:

$$\begin{aligned} \varphi_{\mathcal{B}} : \mathcal{V} &\to \mathbb{F}^n & \qquad \varphi_{\mathcal{B}}^{-1} : \mathbb{F}^n &\to \mathcal{V} \\ \mathbf{v} &\mapsto \bar{\mathbf{v}} = (v^1, \ldots, v^n) , & \bar{\mathbf{v}} &\mapsto \mathbf{v} = v^i \mathbf{b}_i . \end{aligned} \tag{B.1}$$

It is worth noting that, from the perspective of the vector space $\mathbb{F}^n$, the scalars $v^1, \ldots, v^n \in \mathbb{F}$ actually represent the coordinates of the vector $\bar{\mathbf{v}}$ with respect to the basis $\{\bar{\mathbf{e}}_1, \ldots, \bar{\mathbf{e}}_n\}$, i.e. the collection of n-tuples $\bar{\mathbf{e}}_i$ with all entries equals to 0, except the i-th equal to 1. Explicitly, we can write

$$\bar{\mathbf{v}} = v^i \bar{\mathbf{e}}_i = v^1 \bar{\mathbf{e}}_1 + \ldots + v^n \bar{\mathbf{e}}_n = v^1 (1, 0, \ldots, 0) + \ldots + v_n (0, \ldots, 0, 1) .$$

Such a basis is called the *standard basis*, or the *natural basis*, of the vector space $\mathbb{F}^n$.

B.1.2 Linear Mappings

In Appendix A.1.3 we introduced a homomorphism between algebraic structures as a map consistent with the operations of such structures. When we refer to vector spaces, the same idea applies and specializes to the notion of linear mapping.

Definition B.8. Let $\mathcal{V}$ and $\mathcal{W}$ be two vector spaces over the same filed $\mathbb{F}$. A *linear mapping* $\boldsymbol{f} : \mathcal{V} \to \mathcal{W}$, also called a *linear map* or a *linear* *transformation*, is a vector space homomorphism such that

- $f(\mathbf{v}+\mathbf{v}') = f(\mathbf{v}) + f(\mathbf{v}')\,,\ \ \forall\, \mathbf{v}, \mathbf{v}' \in \mathcal{V}$;
- $f(c\mathbf{v}) = cf(\mathbf{v})\,,\ \ \forall\, \mathbf{v} \in \mathcal{V}\,,\ c \in \mathbb{F}$.

It is clear that the definition of linear mapping here introduced is fully consistent with the one of a general homomorphism (Definition A.7). In fact, the first requirement establishes that the linear map f preserves the law of composition in $\mathcal{V}$ and $\mathcal{W}$, intended as additive abelian groups, while the second one ensures that the field $\mathbb{F}$ acts consistently over $\mathcal{V}$ and $\mathcal{W}$.

In particular, since the group structure of $\mathcal{V}$ and $\mathcal{W}$ is preserved, the additive identity of the former space is mapped to the identity of the latter one (see Definition A.9). Explicitly, the null vector $\mathbf{o}_{\mathcal{V}}$ of $\mathcal{V}$ is transformed into the null vector $\mathbf{o}_{\mathcal{W}}$ of $\mathcal{W}$:

$$f(\mathbf{o}_{\mathcal{V}}) = \mathbf{o}_{\mathcal{W}}\,. \tag{B.2}$$

Please notice that the above property also results from the linearity of f. In fact, by Part 1 of Proposition A.26, whatever the vector $\mathbf{v} \in \mathcal{V}$ is considered, one has

$$f(\mathbf{o}_{\mathcal{V}}) = f(0\mathbf{v}) = 0f(\mathbf{v}) = \mathbf{o}_{\mathcal{W}}\,.$$

In order to characterize linear mappings, the same nomenclature introduced in Section A.1.3 for general homomorphisms is also applied in this context. Specifically, the linear map $f\colon \mathcal{V} \to \mathcal{W}$ is called a *monomorphism*, an *epimorphism* or an *isomorphism* if it is injective, surjective or bijective, respectively. Furthermore, if f maps the vector space $\mathcal{V}$ to itself, that is $\mathcal{W} = \mathcal{V}$, it is called an *endomorphism*. Finally, when $f\colon \mathcal{V} \to \mathcal{V}$ is bijective, we say that it is an *automorphism*.

Definition B.9. Let $\mathcal{V}$ and $\mathcal{W}$ be $\mathbb{F}$-vector spaces and $f\colon \mathcal{V} \to \mathcal{W}$ a linear map from $\mathcal{V}$ to $\mathcal{W}$.

1. The *kernel* of f as the set of elements $\mathbf{v} \in \mathcal{V}$ mapped to the null vector

of $\mathcal{W}$:

$$\mathrm{Ker}(\boldsymbol{f}) = \{ \mathbf{v} \in \mathcal{V} \mid \boldsymbol{f}(\mathbf{v}) = \mathbf{o}_{\mathcal{W}} \} .$$

2. The *image*, or the *range*, of $\boldsymbol{f}$ is the set of elements $\mathbf{w} \in \mathcal{W}$ which are images of vectors of $\mathcal{V}$:

$$\mathrm{Im}(\boldsymbol{f}) = \{ \mathbf{w} \in \mathcal{W} \mid \mathbf{w} = \boldsymbol{f}(\mathbf{v}) \, , \; \mathbf{v} \in \mathcal{V} \} .$$

It is worth noting that the definition here introduced for the kernel of the linear map $\boldsymbol{f}$ is consistent with the one considered for a general homomorphism in terms of the congruence relation (A.4). Actually, for any pair $(\mathbf{u}, \mathbf{u}') \in \mathcal{V}^2$ satisfying the condition $\boldsymbol{f}(\mathbf{u}) = \boldsymbol{f}(\mathbf{u}')$, it is always possible to find a vector $\mathbf{v} \in \mathcal{V}$ such that $\boldsymbol{f}(\mathbf{v}) = \mathbf{o}_{\mathcal{W}}$, and vice-versa:

$$\boldsymbol{f}(\mathbf{u}) = \boldsymbol{f}(\mathbf{u}') \;\; \Leftrightarrow \;\; \boldsymbol{f}(\mathbf{v}) = \mathbf{o}_{\mathcal{W}} .$$

The direct implication can be simply verified by exploiting the linearity of $\boldsymbol{f}$ and then setting $\mathbf{v} = \mathbf{u} - \mathbf{u}'$. The converse one results observing that any $\mathbf{v}$ in the kernel of $\boldsymbol{f}$ is also in $\mathcal{V}$, which implies, by virtue of the linear space structure, that such a vector can be expressed as the difference of two elements, say $\mathbf{u}$ and $\mathbf{u}'$. Then, the linearity of $\boldsymbol{f}$ ensures the validity of the condition $\boldsymbol{f}(\mathbf{u}) = \boldsymbol{f}(\mathbf{u}')$.

Proposition B.10. *Let $\boldsymbol{f} : \mathcal{V} \to \mathcal{W}$ be a linear map between the vector spaces $\mathcal{V}$ and $\mathcal{W}$ over the field $\mathbb{F}$.*

1. *The kernel of $\boldsymbol{f}$ is a linear subspace of $\mathcal{V}$.*
2. *The image of $\boldsymbol{f}$ is a linear subspace of $\mathcal{W}$.*

Proof. 1. Let $\mathbf{v}$ and $\mathbf{v}'$ be vectors in the kernel of $\boldsymbol{f}$, and consider two arbitrary scalars, say a and a'. Then, resulting $\boldsymbol{f}(\mathbf{v}) = \boldsymbol{f}(\mathbf{v}') = \mathbf{o}_{\mathcal{W}}$ and being $\boldsymbol{f}$ linear, one finds

$$\boldsymbol{f}(a\mathbf{v} + a'\mathbf{v}') = a\boldsymbol{f}(\mathbf{v}) + a'\boldsymbol{f}(\mathbf{v}') = a\mathbf{o}_{\mathcal{W}} + a'\mathbf{o}_{\mathcal{W}} = \mathbf{o}_{\mathcal{W}} ,$$

that is the linear combination $a\mathbf{v} + a'\mathbf{v}'$ is in $\mathrm{Ker}(\boldsymbol{f})$ and, by Proposition B.3, the kernel of $\boldsymbol{f}$ is actually a subspace of $\mathcal{V}$.

2. Consider $\mathbf{w}$ and $\mathbf{w}'$ in the image of f, meaning that there exist $\mathbf{v}, \mathbf{v}' \in \mathcal{V}$ such that $f(\mathbf{v}) = \mathbf{w}$ and $f(\mathbf{v}') = \mathbf{w}'$. Then, for any pair of scalars $a, a' \in \mathbb{F}$, one has

$$f(a\mathbf{v} + a'\mathbf{v}') = af(\mathbf{v}) + a'f(\mathbf{v}') = a\mathbf{w} + a'\mathbf{w}',$$

which means that the linear combination $a\mathbf{w} + a'\mathbf{w}'$ is in the image of f and, again by Proposition B.3, $\mathrm{Im}(f)$ results a subspace of the codomain $\mathcal{W}$. □

The properties of the linear map f are related to the features of its kernel and image as linear subspaces of $\mathcal{V}$ and $\mathcal{W}$, respectively.

Proposition B.11. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the field $\mathbb{F}$ and let the map $f : \mathcal{V} \to \mathcal{W}$ be linear.*

1. *The map f is injective if, and only if,* $\dim(\mathrm{Ker}(f)) = 0$.
2. *The map f is surjective if, and only if,* $\dim(\mathrm{Im}(f)) = \dim(\mathcal{W})$.

Proof. 1. Suppose $\dim(\ker(f)) = 0$, or equivalently $\mathrm{Ker}(f) = \{\mathbf{o}_\mathcal{V}\}$. In order to prove that f is injective, let us consider an arbitrary pair of vectors $\mathbf{v}, \mathbf{v}' \in \mathcal{V}$ such that $f(\mathbf{v}) = f(\mathbf{v}')$. Then, since f is linear, one can write

$$f(\mathbf{v}) - f(\mathbf{v}') = f(\mathbf{v} - \mathbf{v}') = \mathbf{o}_\mathcal{W},$$

which implies that the difference $\mathbf{v} - \mathbf{v}'$ is in the kernel of f. Hence, because of the assumption, one finds $\mathbf{v} - \mathbf{v}' = \mathbf{o}_\mathcal{V}$, that is the condition $\mathbf{v} = \mathbf{v}'$ ensuring the injectivity of f.

Conversely, if f is supposed to be injective, then any vector $\mathbf{v} \in \mathcal{V}$ such that $\mathbf{v} \neq \mathbf{o}_\mathcal{V}$ has an image $f(\mathbf{v}) \neq f(\mathbf{o}_\mathcal{V}) = \mathbf{o}_\mathcal{W}$. This means that the kernel of f is made only of the null vector $\mathbf{o}_\mathcal{V}$, i.e. $\dim(\ker(f)) = 0$.

2. The surjectivity of f means that every element of the codomain $\mathcal{W}$ has some preimages in $\mathcal{V}$, that is the image of f is coincident with the codomain itself and then $\dim(\mathrm{Im}(f)) = \dim(\mathcal{W})$.

On the contrary, since by Part 2 of Proposition B.10 the image of f is a subspace of $\mathcal{W}$, the condition $\dim(\mathrm{Im}(f)) = \dim(\mathcal{W})$ means that

$\mathrm{Im}(\boldsymbol{f})$ is the codomain itself, i.e. $\boldsymbol{f}$ is surjective. □

Proposition B.12. *Given the $\mathbb{F}$-vector spaces $\mathcal{V}$ and $\mathcal{W}$, let $\boldsymbol{f}: \mathcal{V} \to \mathcal{W}$ be an isomorphism from $\mathcal{V}$ to $\mathcal{W}$. Considering a set of k vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of $\mathcal{V}$, the images $\boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_k) \in \mathcal{W}$ are linearly dependent if, and only if, the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are dependent themselves.*

Proof. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{V}$ are linearly dependent. Then, at least one of them can be expressed as a linear combination of the others:

$$\mathbf{v}_i = \sum_{\substack{j=1 \\ j \neq i}}^{k} a^j \mathbf{v}_j \,,$$

and, for the linearity of $\boldsymbol{f}$, one has

$$\boldsymbol{f}(\mathbf{v}_i) = \sum_{\substack{j=1 \\ j \neq i}}^{k} a^j \boldsymbol{f}(\mathbf{v}_j) \,,$$

whence the linear dependence of the images.

On the other hand, assume that $\boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_k) \in \mathcal{W}$ are linearly dependent and that the following relation holds true:

$$\boldsymbol{f}(\mathbf{v}_i) = \sum_{\substack{j=1 \\ j \neq i}}^{k} a^j \boldsymbol{f}(\mathbf{v}_j) \,.$$

Exploiting again the linearity of $\boldsymbol{f}$, one finds

$$\boldsymbol{f}\Big(\mathbf{v}_i - \sum_{\substack{j=1 \\ j \neq i}}^{k} a^j \mathbf{v}_j\Big) = \mathbf{o}_{\mathcal{W}} \,,$$

so that, since $\boldsymbol{f}$ is isomorphic, we have

$$\mathbf{v}_i - \sum_{\substack{j=1 \\ j \neq i}}^{k} a^j \mathbf{v}_j = \mathbf{o}_{\mathcal{V}} \,.$$

Please notice that the null vector of $\mathcal{V}$ has been written as a linear combination with at least a non-null coefficient, i.e. the one relevant to $\mathbf{v}_1$. Hence, the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{V}$ are linearly dependent. □

Proposition B.13. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the same field $\mathbb{F}$ and let $\boldsymbol{f} : \mathcal{V} \to \mathcal{W}$ be a linear map. Then $\boldsymbol{f}$ is injective if, and only if, for any set of linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{V}$ the images $\boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_k) \in \mathcal{W}$ are linearly independent themselves.*

Proof. Assume $\boldsymbol{f}$ is injective and let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be linearly independent vectors of $\mathcal{V}$.

To show the independence of the images $\boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_k) \in \mathcal{W}$, consider a linear combination by means of the coefficients $a^1, \ldots, a^k \in \mathbb{F}$ providing the null vector, that is

$$a^i \boldsymbol{f}(\mathbf{v}_i) = \mathbf{o}_{\mathcal{W}} \,,$$

which, for the linearity of $\boldsymbol{f}$, also means

$$\boldsymbol{f}(a^i \mathbf{v}_i) = \mathbf{o}_{\mathcal{W}} \,.$$

Since $\boldsymbol{f}$ is injective, its kernel contains only the null vector $\mathbf{o}_{\mathcal{V}}$, obtaining

$$a^i \mathbf{v}_i = \mathbf{o}_{\mathcal{V}} \,,$$

where the independence of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ implies the vanishing of the scalars $a^1, \ldots, a^k$ and so the linear independence of $\boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_k) \in \mathcal{W}$.

On the contrary, suppose that for any collection of linearly independent vectors the images are also independent.

To prove the injectivity of $\boldsymbol{f}$, consider its kernel and verify it contains only the null vector $\mathbf{o}_{\mathcal{V}}$. Hence, let $\mathbf{v}$ be a vector in $\mathcal{V}$ such that $\boldsymbol{f}(\mathbf{v}) = \mathbf{o}_{\mathcal{W}}$.

If $\mathbf{v}$ is not in the span of $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, then the vectors $\mathbf{v}, \mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent, and so are the images $\boldsymbol{f}(\mathbf{v}), \boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_k)$. However, this contradicts the assumption $\boldsymbol{f}(\mathbf{v}) = \mathbf{o}_{\mathcal{W}}$.

Hence, $\mathbf{v}$ is spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_k$ and can be expressed as $\mathbf{v} = a^i \mathbf{v}_i$, so that one finds

$$\boldsymbol{f}(\mathbf{v}) = \boldsymbol{f}(a^i \mathbf{v}_i) = a^i \boldsymbol{f}(\mathbf{v}_i) = \mathbf{o}_{\mathcal{W}} \,,$$

where, being $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_k)$ linearly independent, the coefficients $a^1, \ldots, a^k$ do vanish, and so do the vector $\mathbf{v} = a^i \mathbf{v}_i$.

In conclusion, since $f(\mathbf{v}) = \mathbf{o}_{\mathcal{W}}$ implies $\mathbf{v} = \mathbf{o}_{\mathcal{V}}$, the kernel of f consists solely of $\mathbf{o}_{\mathcal{V}}$ and f is actually injective. □

The proposition here above relates dimensions of the kernel and image of the linear map f, as subspaces of $\mathcal{V}$ and $\mathcal{W}$, to the features the map itself. At the same time, such dimensions are also connected with the dimension of the domain $\mathcal{V}$ by means of the following relation:

$$\dim(\mathrm{Ker}(f)) + \dim(\mathrm{Im}(f)) = \dim(\mathcal{V}) \,. \tag{B.3}$$

For the proof of the property here introduced one can refer, among others, to Lang (1987).

A straightforward consequence of the relation (B.3) is the corollary here reported.

Corollary B.14. *Let $f : \mathcal{V} \to \mathcal{W}$ be a linear map between the vector spaces $\mathcal{V}$ and $\mathcal{W}$ over the same field $\mathbb{F}$, and let us assume $\dim(\mathcal{V}) = \dim(\mathcal{W})$. If $\mathrm{Ker}(f) = \{\mathbf{o}_{\mathcal{V}}\}$, or if $\mathrm{Im}(f) = \mathcal{W}$, then the map f is bijective.*

Proof. If $\mathrm{Ker}(f) = \{\mathbf{o}_{\mathcal{V}}\}$, which means $\dim(\mathrm{Ker}(f)) = 0$, the property (B.3) also implies $\dim(\mathrm{Im}(f)) = \dim(\mathcal{V}) = \dim(\mathcal{W})$.

On the other hand, when $\mathrm{Im}(f) = \mathcal{W}$, and so $\dim(\mathrm{Im}(f)) = \dim(\mathcal{W})$, exploiting again the relation (B.3), one also finds $\dim(\mathrm{Ker}(f)) = 0$.

Either way, the linear map f is both injective and surjective by Proposition B.11. □

Corollary B.15. *Let $\mathcal{V}$ and $\mathcal{W}$ be $\mathbb{F}$-vector spaces with the same dimension, i.e. $\dim(\mathcal{V}) = \dim(\mathcal{W})$. A linear maps $f : \mathcal{V} \to \mathcal{W}$ is bijective if, and only if, for any collection of k linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{V}$ the images $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_k) \in \mathcal{W}$ are also linearly independent.*

Proof. If f is a bijection, the hypotheses of Proposition B.12 are fulfilled. Hence, the linear independence of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{V}$ implies the one of the images $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_k) \in \mathcal{W}$.

Conversely, suppose that if vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{V}$ are linearly independent, so are the images $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_k) \in \mathcal{W}$. Then, by Proposition B.13, the linear map f is injective.

This means that $\mathrm{Ker}(f) = \{\mathbf{o}_{\mathcal{V}}\}$ and, by Corollary B.14, f is actually bijective. □

The Algebra of Linear Maps

Definition B.16. Let $\mathcal{V}$ and $\mathcal{W}$ be $\mathbb{F}$-vector spaces. The *space of linear maps* from $\mathcal{V}$ to $\mathcal{W}$ is the set of all the linear transformations from $\mathcal{V}$ to $\mathcal{W}$ and is denoted as $\mathcal{L}(\mathcal{V}, \mathcal{W})$. The symbol $\mathrm{Hom}(\mathcal{V}, \mathcal{W})$ will also be used to point out the homomorphic structure of such mappings.

In addition, we define the *sum* of two linear transformations $f, g \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ as the map $f + g \colon \mathcal{V} \to \mathcal{W}$ satisfying

$$(f+g)(\mathbf{v}) = f(\mathbf{v}) + g(\mathbf{v}), \ \ \forall\, \mathbf{v} \in \mathcal{V}.$$

Moreover, we consider the *multiplication* of $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ by the scalar $c \in \mathbb{F}$ as the map $cf \colon \mathcal{V} \to \mathcal{W}$ such that

$$(cf)(\mathbf{v}) = cf(\mathbf{v}), \ \ \forall\, \mathbf{v} \in \mathcal{V}.$$

By exploiting the point-wise definition of the sum of linear maps, it is easy to see that the sum of linear maps is itself linear, that is

$$f + g \in \mathcal{L}(\mathcal{V}, \mathcal{W}), \ \ \forall\, f, g \in \mathcal{L}(\mathcal{V}, \mathcal{W}),$$

and the same holds for the scalar multiplication:

$$cf \in \mathcal{L}(\mathcal{V}, \mathcal{W}), \ \ \forall\, c \in \mathbb{F}, \ f \in \mathcal{L}(\mathcal{V}, \mathcal{W}).$$

Furthermore, for any map $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, we can define the opposite $-f$, given by the scalar multiplication $(-1)f$. Also, we call the *zero map*, or the *null map*, the linear transformation which maps any element in $\mathcal{V}$ to the null vector of $\mathcal{W}$:

$$\boldsymbol{o}\colon \mathbf{v} \mapsto \boldsymbol{o}(\mathbf{v}) = \mathbf{o}_{\mathcal{W}}\,, \quad \forall\, \mathbf{v} \in \mathcal{V}\,.$$

With the definitions here introduced, the sum of linear transformations represents a commutative law of composition for the set $\mathcal{L}(\mathcal{V},\mathcal{W})$, what makes the space of linear maps an abelian group.

Additionally, one can verify that the scalar multiplication satisfies the properties required in Definition A.25, so that we conclude that $\mathcal{L}(\mathcal{V},\mathcal{W})$ has a vector space structure. Such a result is summarized by the following proposition.

Proposition B.17. *The space of linear maps $\mathcal{L}(\mathcal{V},\mathcal{W})$ between two vector spaces $\mathcal{V}$ to $\mathcal{W}$ over the same field $\mathbb{F}$ is a vector space.*

Composing linear transformations

Definition B.18. Let $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$ be vector spaces over the field $\mathbb{F}$ and $\boldsymbol{f} \in \mathcal{L}(\mathcal{U},\mathcal{V})$, $\boldsymbol{g} \in \mathcal{L}(\mathcal{V},\mathcal{W})$ linear maps. We define the *product* of the linear maps $\boldsymbol{g}$ and $\boldsymbol{f}$ the map $\boldsymbol{gf}$ such that

$$(\boldsymbol{gf})(\mathbf{u}) = \boldsymbol{g}\big(\boldsymbol{f}(\mathbf{u})\big)\,, \quad \forall\, \mathbf{u} \in \mathcal{U}\,.$$

Please notice that the product of linear transformations is given, de facto, by the map composition. Then it is well-defined only when the involved maps are composable. Explicitly, it is required that

$$\boldsymbol{f}\colon \mathcal{U} \to \mathcal{V}\,, \quad \boldsymbol{g}\colon \mathcal{V} \to \mathcal{W}\,,$$

and their product is a map between $\mathcal{U}$ and $\mathcal{W}$:

$$\boldsymbol{gf}\colon \mathcal{U} \to \mathcal{W}\,.$$

Proposition B.19. *Let $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$ be $\mathbb{F}$-vector spaces and let $\boldsymbol{f}$ and $\boldsymbol{g}$ be linear maps in $\mathcal{L}(\mathcal{U},\mathcal{V})$ and $\mathcal{L}(\mathcal{V},\mathcal{W})$, respectively. Then, the product map $\boldsymbol{gf}$ is linear.*

Proof. To prove the linearity of gf, we verify that the requirements of Definition B.8 are satisfied.

Since both f and g are linear, the point-wise evaluation of gf at arbitrary elements $\mathbf{u}$ and $\mathbf{u}'$ of $\mathcal{U}$, by Definition B.18, reads

$$(gf)(\mathbf{u}+\mathbf{u}') = g\big(f(\mathbf{u}+\mathbf{u}')\big) = g\big(f(\mathbf{u})+f(\mathbf{u}')\big) = g\big(f(\mathbf{u})\big)+g\big(f(\mathbf{u}')\big)\,,$$

whence

$$(gf)(\mathbf{u}+\mathbf{u}') = (gf)(\mathbf{u}) + (gf)(\mathbf{u}')\,,\ \ \forall\,\mathbf{u},\mathbf{u}' \in \mathcal{U}\,.$$

Also, for any scalar $c \in \mathbb{F}$ and an arbitrary element $\mathbf{u}$ of $\mathcal{U}$, one has

$$(gf)(c\mathbf{u}) = g(f(c\mathbf{u})) = g(cf(\mathbf{u})) = cg(f(\mathbf{u}))\,,$$

obtaining

$$(gf)(c\mathbf{u}) = (cgf)(\mathbf{u})\,,\ \ \forall\,\mathbf{u} \in \mathcal{U}\,,\ c \in \mathbb{F}\,.$$

□

Proposition B.20. *Let $f \in \mathcal{L}(\mathcal{V},\mathcal{W})$ be an isomorphism between the vector spaces $\mathcal{V}$ and $\mathcal{W}$ over $\mathbb{F}$. Then, the inverse $f^{-1}\colon \mathcal{W} \to \mathcal{V}$ is linear.*

Proof. Since we have identified the map composition with the product of maps, the inverse of f is such that

$$f^{-1}f = \mathrm{id}_{\mathcal{V}}\ .$$

Then, since f is linear, for an arbitrary pair of vectors $\mathbf{v},\mathbf{v}' \in \mathcal{V}$ the following relation holds:

$$f^{-1}\big(f(\mathbf{v})+f(\mathbf{v}')\big) = f^{-1}\big(f(\mathbf{v}+\mathbf{v}')\big) = \mathbf{v}+\mathbf{v}' = f^{-1}\big(f(\mathbf{v})\big)+f^{-1}\big(f(\mathbf{v}')\big)\,,$$

which, by setting $\mathbf{w} = f(\mathbf{v})$ and $\mathbf{w}' = f(\mathbf{v}')$, can be written as

$$f^{-1}(\mathbf{w}+\mathbf{w}') = f^{-1}(\mathbf{w}) + f^{-1}(\mathbf{w}')\,,\ \ \forall\,\mathbf{w},\mathbf{w}' \in \mathcal{W}\,.$$

Similarly, considering an arbitrary scalar c of the field $\mathbb{F}$ and a vector

$\mathbf{v} \in \mathcal{V}$, one has

$$f^{-1}\big(cf(\mathbf{v})\big) = f^{-1}\big(f(c\mathbf{v})\big) = c\mathbf{v} = cf^{-1}\big(f(\mathbf{v})\big)\,,$$

that is, setting again $\mathbf{w} = f(\mathbf{v})$,

$$f^{-1}(c\mathbf{w}) = cf^{-1}(\mathbf{w})\,,\ \ \forall\, \mathbf{w} \in \mathcal{W}\,,\ c \in \mathbb{F}\,.$$

Since both the properties of Definition B.8 are verified, the linearity of f^{-1} is proved. □

Change of Basis Map

Let us introduce a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ for the $\mathbb{F}$-vector space $\mathcal{V}$ and consider the map $\varphi_{\mathcal{B}} \colon \mathcal{V} \to \mathbb{F}^n$ defined by (B.1).

It is worth noting that $\varphi_{\mathcal{B}}$ is not only a vector space isomorphism, but it is also a linear transformation.

In fact, by expressing any vector of $\mathcal{V}$ with respect to the basis $\mathcal{B}$, both the vector addition and the scalar multiplication are reflected in the relevant operations on the coordinates, i.e. on the images through the map $\varphi_{\mathcal{B}}$.

Explicitly, observing that

$$\mathbf{v} + \mathbf{v}' = v^i\mathbf{b}_i + v'^i\mathbf{b}_i = (v^i + v'^i)\mathbf{b}_i\,,\ \ \forall\, \mathbf{v}, \mathbf{v}' \in \mathcal{V}\,,$$

as well as

$$c\mathbf{v} = c\, v^i\mathbf{b}_i = (cv^i)\mathbf{b}_i\,,\ \ \forall\, \mathbf{v} \in \mathcal{V}\,,\ c \in \mathbb{F}\,,$$

one can easily derive

$$\varphi_{\mathcal{B}}(\mathbf{v} + \mathbf{v}') = \varphi_{\mathcal{B}}(\mathbf{v}) + \varphi_{\mathcal{B}}(\mathbf{v}')\,,\ \ \forall\, \mathbf{v}, \mathbf{v}' \in \mathcal{V}\,,$$

and also

$$\varphi_{\mathcal{B}}(c\mathbf{v}) = c\varphi_{\mathcal{B}}(\mathbf{v})\,,\ \ \forall\, \mathbf{v} \in \mathcal{V}\,,\ c \in \mathbb{F}\,.$$

Then the map $\varphi_{\mathcal{B}}$ is actually linear in accordance with Definition B.8 and the same applies to the inverse map $\varphi_{\mathcal{B}}^{-1}$ (cf. Proposition B.20).

Let us now consider two distinct bases. Then, it is possible to uniquely define a linear transformation mapping each vector of one basis to the relevant one of the other basis.

Proposition B.21. *Let $\mathcal{V}$ be an $\mathbb{F}$-vector space and let $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be two distinct bases for $\mathcal{V}$. Then, there exists a unique automorphism $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}} : \mathcal{V} \to \mathcal{V}$ mapping each vector of $\mathcal{A}$ into the relevant element in $\mathcal{B}$:*

$$\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}(\mathbf{a}_i) = \mathbf{b}_i\,, \;\; \forall\, i = 1, \ldots, n\,. \tag{B.4}$$

Proof. Let $\varphi_{\mathcal{A}}$ and $\varphi_{\mathcal{B}}$ be the specifications of the map defined by (B.1) to the bases $\mathcal{A}$ and $\mathcal{B}$, respectively.

In other words, the maps $\varphi_{\mathcal{A}}$ and $\varphi_{\mathcal{B}}$ provide the coordinate vectors $\bar{\mathbf{v}}_{\mathcal{A}}$ and $\bar{\mathbf{v}}_{\mathcal{B}}$ of the same vector $\mathbf{v} \in \mathcal{V}$ with respect to the basis $\mathcal{A}$ and $\mathcal{B}$, respectively.

Then, the map $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}$ can be easily be constructed by composing the inverse of $\varphi_{\mathcal{B}}$ with $\varphi_{\mathcal{A}}$:

$$\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}} = \varphi_{\mathcal{B}}^{-1} \circ \varphi_{\mathcal{A}}\,. \tag{B.5}$$

In fact, since the i-th coordinate of $\mathbf{a}_i$ with respect to $\mathcal{A}$ is 1 and all the other ones do vanish, applying the map $\varphi_{\mathcal{A}}$ provides the i-th element of the standard basis of $\mathbb{F}^n$:

$$\varphi_{\mathcal{A}}(\mathbf{a}_i) = \bar{\mathbf{e}}_i = (0, \ldots, 1, \ldots, 0)\,, \;\; \forall\, i = 1, \ldots, n\,.$$

The same applies to the i-th vector of the basis $\mathcal{B}$ with respect to the map $\varphi_{\mathcal{B}}$, so that the inverse transformation $\varphi_{\mathcal{B}}^{-1}$ maps $\bar{\mathbf{e}}_i \in \mathbb{F}^n$ to the vector $\mathbf{b}_i \in \mathcal{V}$:

$$\varphi_{\mathcal{B}}^{-1}(\bar{\mathbf{e}}_i) = \mathbf{b}_i\,, \;\; \forall\, i = 1, \ldots, n\,.$$

At this point, the following relation can be trivially verified:

$$\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}(\mathbf{a}_i) = (\varphi_{\mathcal{B}}^{-1} \circ \varphi_{\mathcal{A}})(\mathbf{a}_i) = \varphi_{\mathcal{B}}^{-1}(\bar{\mathbf{e}}_i) = \mathbf{b}_i\,, \;\; \forall\, i = 1, \ldots, n\,.$$

The uniqueness of $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}$ is a direct consequence of the one of $\varphi_{\mathcal{A}}$, as well as of $\varphi_{\mathcal{B}}$, since for any vector of $\mathcal{V}$ there exists a unique set of coordinates with respect to any fixed basis (cf. Proposition B.6).

Finally, being the composition of linear isomorphisms, the map $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}$ is itself a bijective linear transformation. □

The map $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}$, defined by the property (B.4), is called the *change of basis map* from $\mathcal{A}$ to $\mathcal{B}$.

Clearly, the inverse map ${\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}}^{-1} = \boldsymbol{h}_{\mathcal{B}}^{\mathcal{A}}$ provides the change of basis from $\mathcal{B}$ to $\mathcal{A}$:

$$ {\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}}^{-1}(\mathbf{b}_i) = \boldsymbol{h}_{\mathcal{B}}^{\mathcal{A}}(\mathbf{b}_i) = \mathbf{a}_i\,, \ \ \forall\, i = 1, \ldots, n\,. $$

B.1.3 Dual Vector Space

Let $\mathcal{V}$ be an n-dimensional vector space over the field $\mathbb{F}$ and consider a linear map from $\mathcal{V}$ to $\mathbb{F}$. We call such a map a *linear functional*, or also a *linear form*, from $\mathcal{V}$ to $\mathbb{F}$.

We recall that the field $\mathbb{F}$ can be seen as a one-dimensional vector space over $\mathbb{F}$ itself. Then, the set of the linear functionals of V actually represents the space of homomorphisms $\mathrm{Hom}(\mathcal{V}, \mathbb{F})$, which, by Proposition B.17, is a vector space over $\mathbb{F}$.

In order to emphasize the vector space structure of $\mathrm{Hom}(\mathcal{V}, \mathbb{F})$ and how this is related with the one of $\mathcal{V}$, the space of functionals over $\mathcal{V}$ is more properly called the *dual space* of $\mathcal{V}$ and is usually denoted as $\mathcal{V}^*$.

Definition B.22. Let $\mathcal{V}$ be an $\mathbb{F}$-vector space. The *dual space* of $\mathcal{V}$ is the vector space $\mathcal{V}^*$ of the linear functionals from $\mathcal{V}$ to $\mathbb{F}$.

The elements of the dual space $\mathcal{V}^*$ are the linear maps $\mathbf{w}^*\colon \mathcal{V} \to \mathbb{F}$ and are called *dual vectors*, *covectors* or *linear forms*.

Remark (Notation)**.** In order to distinguish the covectors from the vectors in $\mathcal{V}$, we usually adopt the superscript '$*$' (e.g., $\mathbf{w}^* \in \mathcal{V}^*$). However, when a covector is intended as an element of a numerable set, the relevant index is itself used as a superscript, such as $\mathbf{w}^i \in \mathcal{V}^*$.

Consistently, if we regard $\mathbb{F}$ as the field of the dual space $\mathcal{V}^*$, the i-th element within a collection of scalars is identified by the relevant subscript (e.g., $a_i \in \mathbb{F}$), and the Einstein convention still applies:

$$a_i \mathbf{w}^i = \sum_{i=1}^{n} a_i \mathbf{w}^i = a_1 \mathbf{w}^1 + \ldots + a_n \mathbf{w}^n \in \mathcal{V}^* .$$

We recall that by fixing a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ for the vector space $\mathcal{V}$, any vector $\mathbf{v}$ is completely defined by the n-tuple of scalars $(v^1, \ldots, v^n)$. The component v^i, that is the i-th coordinate of $\mathbf{v}$ with respect to the basis $\mathcal{B}$, can be seen as the image of $\mathbf{v}$ through the *coordinate function* $\mathbf{b}^i$ (see Lang (1987)):

$$\begin{aligned} \mathbf{b}^i :\ & \mathcal{V} \to \mathbb{F} \\ & \mathbf{v} \mapsto v^i = \mathbf{b}^i(\mathbf{v}) . \end{aligned} \tag{B.6}$$

Moreover, any of the n maps $\mathbf{b}^i$ is clearly linear, and hence it is in $\mathcal{V}^*$.

Proposition B.23. *Let $\mathcal{V}$ be an $\mathbb{F}$-vector space and $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ a basis. Then, the set $\mathcal{B}^* = \{\mathbf{b}^1, \ldots, \mathbf{b}^n\}$ of the coordinate functions is a basis for the dual space $\mathcal{V}^*$.*

Proof. We first prove that any covector $\mathbf{w}^* \in \mathcal{V}^*$ is in the span of $\mathcal{B}^*$. Actually, for the linearity of $\mathbf{w}^*$ and using the coordinate function (B.6), one has

$$\mathbf{w}^*(\mathbf{v}) = \mathbf{w}^*(v^i \mathbf{b}_i) = v^i \mathbf{w}^*(\mathbf{b}_i) = \mathbf{b}^i(\mathbf{v})\, \mathbf{w}^*(\mathbf{b}_i) , \ \ \forall\, \mathbf{v} \in \mathcal{V} ,$$

Moreover, denoting by w_i the image of $\mathbf{b}_i$ through the functional $\mathbf{w}^*$, i.e. $w_i = \mathbf{w}^*(\mathbf{b}_i) \in \mathbb{F}$, the above equation can be written as

$$\mathbf{w}^*(\mathbf{v}) = w_i \mathbf{b}^i(\mathbf{v}) , \ \ \forall\, \mathbf{v} \in \mathcal{V} ,$$

and, for the arbitrariness of $\mathbf{v}$, one finally has

$$\mathbf{w}^* = w_i \mathbf{b}^i \in \mathsf{Span}(\mathcal{B}^*) .$$

To prove the linear independence of $\mathcal{B}^*$, we consider a linear combination providing the null vector $\mathbf{o}^*$ of $\mathcal{V}^*$, i.e.

$$a_i \mathbf{b}^i = \mathbf{o}^* .$$

Since $\mathbf{o}^*$ is the null functional, its value vanishes at any vector of $\mathcal{V}$. Specifically, for the vectors $\mathbf{b}_j$ of the basis $\mathcal{B}$ one has

$$\mathbf{o}^*(\mathbf{b}_j) = a_i \mathbf{b}^i(\mathbf{b}_j) = 0 , \quad j = 1, \ldots, n .$$

In addition, observing that the j-th coordinate of $\mathbf{b}_j$ is exactly 1 while all the others vanish, the coordinate functions $\mathbf{b}^i$ give null contributions in the above summations except when $i = j$, obtaining

$$a_j = 0 , \quad j = 1, \ldots, n .$$

Hence, all the coefficients providing the null covector $\mathbf{o}^*$ vanish and the linear independence of $\mathcal{B}^*$ is proved. □

The straightforward corollary of Proposition B.23 is that, since the number of the coordinate functions $\mathbf{b}^i$ is exactly n, the dual space $\mathcal{V}^*$ has the same dimension as $\mathcal{V}$:

$$\dim(\mathcal{V}^*) = \dim(\mathcal{V}) . \tag{B.7}$$

It is worth emphasizing that the covectors $\mathbf{b}^i$ of $\mathcal{B}^*$ are strictly related to the basis $\mathcal{B}$ introduced for the vector space $\mathcal{V}$, because of their role as coordinate functions specified by (B.6).

Notably, each covector $\mathbf{b}^i$ is the only linear functional on $\mathcal{V}$ providing the value 1 when applied to $\mathbf{b}_i$ and 0 if evaluated at the other elements of the basis $\mathcal{B}$. For this reason the following definition is introduced.

Definition B.24. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for the $\mathbb{F}$-vector space $\mathcal{V}$ and let $\mathcal{B}^* = \{\mathbf{b}^1, \ldots, \mathbf{b}^n\}$ be the basis of the dual space $\mathcal{V}^*$, made of the coordinate functions of $\mathcal{V}$ with respect of the basis $\mathcal{B}$. The set $\mathcal{B}^*$ is called the *dual basis* to $\mathcal{B}$ and its covectors $\mathbf{b}^i$ are characterized by the following property:

$$\mathbf{b}^i(\mathbf{b}_j) = \delta^i_j\,. \tag{B.8}$$

In the above definition, δ^i_j is the Kronecker symbol, whose value is 1 if $i = j$ and 0 otherwise:

$$\delta^i_j = \begin{cases} 1 & \text{if } i = j\,, \\ 0 & \text{if } i \neq j\,. \end{cases} \tag{B.9}$$

Please notice again that the basis vector $\mathbf{b}_i$ plays for the covectors in $\mathcal{V}^*$ a comparable role to the one of the coordinate function $\mathbf{b}^i$, defined by (B.6), for the vectors of $\mathcal{V}$.

In fact, any covector $\mathbf{w}^*$ can be identified by the n-tuple $(w_1, \ldots, w_n)$, which collects the coordinates of $\mathbf{w}^*$ with respect to the dual basis $\mathcal{B}^*$, that is $\mathbf{w}^* = w_i\mathbf{b}^i$. Then, applying the covector $\mathbf{w}^*$ to the vector $\mathbf{b}_i$ of the basis $\mathcal{B}$ provides exactly the i-th coordinate of $\mathbf{w}^*$ with respect to $\mathcal{B}^*$:

$$\mathbf{w}^*(\mathbf{b}_i) = w_i\,. \tag{B.10}$$

Such a property comes trivially from Definition B.24, resulting

$$\mathbf{w}^*(\mathbf{b}_i) = w_j\mathbf{b}^j(\mathbf{b}_i) = w_j\delta^j_i = w_i\,.$$

A further consequence of the relation between the basis $\mathcal{B}$ of $\mathcal{V}$ and its dual basis $\mathcal{B}^*$ is the following proposition.

Proposition B.25. *Let $\mathcal{V}$ be an $\mathbb{F}$-vector space and $\mathcal{V}^*$ its dual space.*

1. *For any non-null vector $\mathbf{v} \in \mathcal{V}$, there exists a covector $\mathbf{w}^* \in \mathcal{V}^*$ such that $\mathbf{w}^*(\mathbf{v}) \neq 0$. Equivalently,*

$$\mathbf{w}^*(\mathbf{v}) = 0\,,\ \ \forall\,\mathbf{w}^* \in \mathcal{V}^* \quad\Leftrightarrow\quad \mathbf{v} = \mathbf{o}\,.$$

2. *For any non-null covector $\mathbf{w}^* \in \mathcal{V}^*$, there exists a vector $\mathbf{v} \in \mathcal{V}$ such that $\mathbf{w}^*(\mathbf{v}) \neq 0$. Equivalently,*

$$\mathbf{w}^*(\mathbf{v}) = 0\,,\ \ \forall\,\mathbf{v} \in \mathcal{V} \quad\Leftrightarrow\quad \mathbf{w}^* = \mathbf{o}^*\,.$$

Proof. Consider the basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of $\mathcal{V}$ and the dual basis $\mathcal{B}^* = \{\mathbf{b}^1, \ldots, \mathbf{b}^n\}$ of $\mathcal{V}^*$.

1. Once the basis $\mathcal{B}^*$ has been fixed, we can write $\mathbf{w}^* = w_i \mathbf{b}^i \in \mathcal{V}^*$ and the arbitrariness of $\mathbf{w}^*$ results in that of the coordinates w_i. Then, applying (B.6), one has

$$w_i \mathbf{b}^i(\mathbf{v}) = w_i v^i = 0\,, \ \ \forall\, w_i \in \mathbb{F}\,, \ i = 1, \ldots, n\,,$$

which implies $v^i = 0$ for any $i = 1, \ldots, n$, that is

$$\mathbf{v} = v^i \mathbf{b}_i = \mathbf{o}\,.$$

2. Proceeding as above, we consider an arbitrary set of coordinates v^i, with respect to the basis $\mathcal{B}$, providing a vector $\mathbf{v} = v^i \mathbf{b}_i$. Then, for the linearity of $\mathbf{w}^* \in \mathcal{V}^*$ and using (B.10), one easily finds

$$\mathbf{w}^*(v^i \mathbf{b}_i) = v^i \mathbf{w}^*(\mathbf{b}_i) = v^i w_i = 0\,, \ \ \forall\, v^i \in \mathbb{F}\,, \ i = 1, \ldots, n\,,$$

whence, for the arbitrariness of $v^i \in \mathbb{F}$, one infers $w_i = 0$ for any $i = 1, \ldots, n$ and then

$$\mathbf{w}^* = w_i \mathbf{b}^i = \mathbf{o}^*\,.$$

□

Transpose of a Linear Map

Consider now two vector spaces over the same field, along with the relevant dual spaces. If the vector spaces are related by a linear transformation, a further linear map is induced between the relevant dual spaces. Such a property is formally defined here below.

Definition B.26. Let $\mathcal{V}$ and $\mathcal{W}$ be $\mathbb{F}$-vector spaces and $\boldsymbol{f}\colon \mathcal{V} \to \mathcal{W}$ a linear map. The *transpose*, or the *dual map*, of $\boldsymbol{f}$ is the linear map $\boldsymbol{f}^{\mathsf{T}}\colon \mathcal{W}^* \to \mathcal{V}^*$ such that

$$\boldsymbol{f}^{\mathsf{T}}(\mathbf{w}^*) = \mathbf{w}^* \circ \boldsymbol{f}\,, \ \ \forall\, \mathbf{w}^* \in \mathcal{W}^*\,, \tag{B.11}$$

and the functional $\boldsymbol{f}^{\mathsf{T}}(\mathbf{w}^*) \in \mathcal{V}^*$ is the *pullback* of $\mathbf{w}^*$ along $\boldsymbol{f}$.

The above definition implies the following identity:

$$\big(\boldsymbol{f}^{\mathsf{T}}(\mathbf{w}^*)\big)(\mathbf{v}) = \mathbf{w}^*\big(\boldsymbol{f}(\mathbf{v})\big)\,, \ \ \forall\,\mathbf{v} \in \mathcal{V}\,,\ \mathbf{w}^* \in \mathcal{W}^*\,, \tag{B.12}$$

which also allows one to easily verify the linearity of $\boldsymbol{f}^{\mathsf{T}}$.

Actually, considering a linear combination $a\mathbf{w}^* + a'\mathbf{w}'^* \in \mathcal{W}^*$ and an arbitrary vector $\mathbf{v} \in \mathcal{V}$, the linearity of the map $\boldsymbol{f}$ and the functionals $\mathbf{w}^*$ and $\mathbf{w}'^*$ implies

$$\big(\boldsymbol{f}^{\mathsf{T}}(a\mathbf{w}^* + a'\mathbf{w}'^*)\big)(\mathbf{v}) = (a\mathbf{w}^* + a'\mathbf{w}'^*)\big(\boldsymbol{f}(\mathbf{v})\big) = a\mathbf{w}^*\big(\boldsymbol{f}(\mathbf{v})\big) + a'\mathbf{w}'^*\big(\boldsymbol{f}(\mathbf{v})\big)\,,$$

which, by using again identity (B.12), can be written as

$$\big(\boldsymbol{f}^{\mathsf{T}}(a\mathbf{w}^* + a'\mathbf{w}'^*)\big)(\mathbf{v}) = \big(a\boldsymbol{f}^{\mathsf{T}}(\mathbf{w}^*)\big)(\mathbf{v}) + \big(a'\boldsymbol{f}^{\mathsf{T}}(\mathbf{w}'^*)\big)(\mathbf{v})\,,$$

and, for the arbitrariness of $\mathbf{v}$, one finally finds

$$\boldsymbol{f}^{\mathsf{T}}(a\mathbf{w}^* + a'\mathbf{w}'^*) = a\boldsymbol{f}^{\mathsf{T}}(\mathbf{w}^*) + a'\boldsymbol{f}^{\mathsf{T}}(\mathbf{w}'^*)\,.$$

When two composable linear maps are considered, say $\boldsymbol{f}\colon \mathcal{U} \to \mathcal{V}$ and $\boldsymbol{g}\colon \mathcal{V} \to \mathcal{W}$, the transpose of their product is given by the reversed product of the relevant dual maps, that is

$$(\boldsymbol{g}\boldsymbol{f})^{\mathsf{T}} = \boldsymbol{f}^{\mathsf{T}}\boldsymbol{g}^{\mathsf{T}}\,. \tag{B.13}$$

The above statement can be easily verified by applying the property defined by (B.11) to an arbitrary functional $\mathbf{w}^* \in \mathcal{W}^*$ and simply exploiting the associativity of map composition:

$$(\boldsymbol{g}\boldsymbol{f})^{\mathsf{T}}(\mathbf{w}^*) = \mathbf{w}^* \circ (\boldsymbol{g}\boldsymbol{f}) = (\mathbf{w}^* \circ \boldsymbol{g}) \circ \boldsymbol{f} = \big(\boldsymbol{g}^{\mathsf{T}}(\mathbf{w}^*)\big) \circ \boldsymbol{f} = \boldsymbol{f}^{\mathsf{T}}\big(\boldsymbol{g}^{\mathsf{T}}(\mathbf{w}^*)\big)\,,$$

which can be concisely written as

$$(\boldsymbol{g}\boldsymbol{f})^{\mathsf{T}}(\mathbf{w}^*) = (\boldsymbol{f}^{\mathsf{T}}\boldsymbol{g}^{\mathsf{T}})(\mathbf{w}^*)\,, \ \ \forall\,\mathbf{w}^* \in \mathcal{W}^*\,.$$

Then, for the arbitrariness of $\mathbf{w}^*$, identity (B.13) is proved.

Double Dual Space

The vector space structure of the dual space $\mathcal{V}^*$ leads us to consider the linear maps defined on it. Specifically, the set of the linear functionals of $\mathcal{V}^*$ provides its own dual space.

The dual space of $\mathcal{V}^*$ is called the *double dual space*, or sometimes the *second dual space*, of $\mathcal{V}$ and it is denoted as $\mathcal{V}^{**}$.

We recall that a covector $\mathbf{w}^* \in \mathcal{V}^*$ is a linear functional on $\mathcal{V}$, meaning that $\mathbf{w}^*(\mathbf{v})$ represents the image of the map $\mathbf{w}^* : \mathcal{V} \to \mathbb{F}$ evaluated at $\mathbf{v} \in \mathcal{V}$. However, if the vector $\mathbf{v}$ is hold fixed, the same scalar $\mathbf{w}^*(\mathbf{v})$ can be thought as the image of a different map, whose domain is $\mathcal{V}^*$, evaluated at $\mathbf{w}^*$.

Then, for each $\mathbf{v} \in \mathcal{V}$, we can define the following functional of $\mathcal{V}^*$:

$$\begin{aligned} \xi_{\mathbf{v}} :\ & \mathcal{V}^* \to \mathbb{F} \\ & \mathbf{w}^* \mapsto \xi_{\mathbf{v}}(\mathbf{w}^*) = \mathbf{w}^*(\mathbf{v}) \,. \end{aligned} \tag{B.14}$$

Because of the algebraic structure of $\mathcal{V}^*$, the map $\xi_{\mathbf{v}}$ is clearly linear. Actually, considering an arbitrary pair of covectors $\mathbf{w}^*, \mathbf{w}'^* \in \mathcal{V}^*$, one has

$$\xi_{\mathbf{v}}(\mathbf{w}^* + \mathbf{w}'^*) = (\mathbf{w}^* + \mathbf{w}'^*)(\mathbf{v}) = \mathbf{w}^*(\mathbf{v}) + \mathbf{w}'^*(\mathbf{v}) = \xi_{\mathbf{v}}(\mathbf{w}^*) + \xi_{\mathbf{v}}(\mathbf{w}'^*) \,.$$

Similarly, for any scalar $c \in \mathbb{F}$ and any $\mathbf{w}^* \in \mathcal{V}^*$, one also has

$$\xi_{\mathbf{v}}(c\mathbf{w}^*) = (c\mathbf{w}^*)(\mathbf{v}) = c\mathbf{w}^*(\mathbf{v}) = c\xi_{\mathbf{v}}(\mathbf{w}^*) \,,$$

proving the linearity of $\xi_{\mathbf{v}}$.

For this reason, we conclude that, for each vector $\mathbf{v}$ of $\mathcal{V}$, the map $\xi_{\mathbf{v}} : \mathcal{V}^* \to \mathbb{F}$ is in the double dual space $\mathcal{V}^{**}$.

Proposition B.27. *Let $\mathcal{V}$ be an n-dimensional vector space over $\mathbb{F}$ and $\mathcal{V}^{**}$ its double dual. Then, $\mathcal{V}$ and $\mathcal{V}^{**}$ are isomorphic under the following vector space isomorphism:*

$$\begin{aligned} \xi :\ & \mathcal{V} \to \mathcal{V}^{**} \\ & \mathbf{v} \mapsto \xi_{\mathbf{v}} \,, \end{aligned} \tag{B.15}$$

where $\xi_{\mathbf{v}} : \mathcal{V}^* \to \mathbb{F}$ is the functional such that $\xi_{\mathbf{v}}(\mathbf{w}^*) = \mathbf{w}^*(\mathbf{v})$ for any

$\mathbf{w}^* \in \mathcal{V}^*$.

Proof. We first prove that ξ is a linear map. To this aim, let $a\mathbf{v} + a'\mathbf{v}'$ be an arbitrary linear combination of two vectors in $\mathcal{V}$. Applying (B.14) to an arbitrary $\mathbf{w}^* \in \mathcal{V}^*$, one has

$$\xi_{a\mathbf{v}+a'\mathbf{v}'}(\mathbf{w}^*) = \mathbf{w}^*(a\mathbf{v} + a'\mathbf{v}') = a\mathbf{w}^*(\mathbf{v}) + a'\mathbf{w}^*(\mathbf{v}')\,,$$

where the linearity of the functional $\mathbf{w}^*$ has been used. Then, again by (B.14), one finds

$$\xi_{a\mathbf{v}+a'\mathbf{v}'}(\mathbf{w}^*) = a\xi_{\mathbf{v}}(\mathbf{w}^*) + a'\xi_{\mathbf{v}'}(\mathbf{w}^*) = (a\xi_{\mathbf{v}} + a'\xi_{\mathbf{v}'})(\mathbf{w}^*)\,,$$

and, for the arbitrariness of $\mathbf{w}^* \in \mathcal{V}^*$, the linearity of ξ is proved:

$$\xi_{a\mathbf{v}+a'\mathbf{v}'} = a\xi_{\mathbf{v}} + a'\xi_{\mathbf{v}'}\,.$$

Now we prove that ξ is a vector space isomorphism. In this respect, we first observe that the domain $\mathcal{V}$ and the codomain $\mathcal{V}^{**}$ have the same dimension. Actually, by (B.7), the dual space $\mathcal{V}^*$ has the dimension of $\mathcal{V}$, and the same goes for $\mathcal{V}^{**}$ as dual of $\mathcal{V}^*$.

Furthermore, denoted by $\mathbf{o}^{**}$ the null vector of $\mathcal{V}^{**}$, let us consider the condition $\xi_{\mathbf{v}} = \mathbf{o}^{**}$ characterizing the kernel of ξ. As a functional of $\mathcal{V}^*$, the map $\mathbf{o}^{**}\colon \mathcal{V}^* \to \mathbb{F}$ satisfies

$$\mathbf{o}^{**}(\mathbf{w}^*) = 0\,, \ \ \forall\, \mathbf{w}^* \in \mathcal{V}^*\,,$$

so that, using (B.14), the vectors $\mathbf{v}$ of $\mathrm{Ker}(\xi)$ do fulfill the following condition:

$$\xi_{\mathbf{v}}(\mathbf{w}^*) = \mathbf{w}^*(\mathbf{v}) = 0\,, \ \ \forall\, \mathbf{w}^* \in \mathcal{V}^*\,,$$

which, by Part 1 of Proposition B.25, is satisfied if, and only if, $\mathbf{v} = \mathbf{o}_{\mathcal{V}}$.

Hence, we obtain $\mathrm{Ker}(\xi) = \{\mathbf{o}_{\mathcal{V}}\}$ and, by Corollary B.14, we conclude that ξ is a vector space isomorphism. □

The map ξ, defined by (B.15), represents the *canonical isomorphism* between $\mathcal{V}$ and $\mathcal{V}^{**}$, meaning that each vector $\mathbf{v}$ is naturally associated with the functional $\xi_{\mathbf{v}}$ without reference to any basis (Lee (2012)). By virtue of such an isomorphism, whose algebraic meaning relies on (B.14), it is possible to unambiguously identify the double dual $\mathcal{V}^{**}$ with the vector space $\mathcal{V}$ itself:

$$\mathcal{V}^{**} \cong \mathcal{V} . \tag{B.16}$$

Consequently, the vector space $\mathcal{V}$ can itself be considered the dual of $\mathcal{V}^*$, in the sense that any vector $\mathbf{v} \in \mathcal{V}$ can be thought as a linear functional acting on $\mathcal{V}^*$ as follows:

$$\begin{aligned} \mathbf{v} :\ & \mathcal{V}^* \to \mathbb{F} \\ & \mathbf{w}^* \mapsto \mathbf{v}(\mathbf{w}^*) = \mathbf{w}^*(\mathbf{v}) . \end{aligned} \tag{B.17}$$

On the other hand, a similar identification of the dual space $\mathcal{V}^*$ with the vector space $\mathcal{V}$ is not possible. Actually, even if $\mathcal{V}^*$ is isomorphic with $\mathcal{V}$, since they have the same dimension, there is not a canonical isomorphism.

B.2 Multilinear Forms and Tensors

B.2.1 Tensor Product

The dual space $\mathcal{V}^*$ has been introduced considering the linear functionals of the vector space $\mathcal{V}$.

Definition B.28. Let $\mathcal{V}_1, \dots, \mathcal{V}_k$ and $\mathcal{W}$ be vector space over the field $\mathbb{F}$. A *multilinear map* is mapping $\mathbf{A} : \mathcal{V}_1 \times \cdots \times \mathcal{V}_k \to \mathcal{W}$ that is linear as a function of each variable, separately, when the others are all held fixed. We also say that the map $\mathbf{A}$ is *k-linear*.

Based on Definition B.8, the multilinear map $\mathbf{A}$ is characterized by the following conditions:

- $\mathbf{A}(\mathbf{v}_1, \dots, \mathbf{v}_i + \mathbf{v}'_i, \dots, \mathbf{v}_k) = \mathbf{A}(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_k) + \mathbf{A}(\mathbf{v}_1, \dots, \mathbf{v}'_i, \dots, \mathbf{v}_k) , \ \forall\, \mathbf{v}_i, \mathbf{v}'_i \in \mathcal{V}_i , \ i = 1, \dots, k ;$

- $\mathbf{A}(\mathbf{v}_1, \ldots, c\mathbf{v}_i, \ldots, \mathbf{v}_k) = c\mathbf{A}(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_k)\,,$
$$\forall\, \mathbf{v}_i \in \mathcal{V}_i\,,\ i = 1, \ldots, k\,.$$

Please notice that the multilinear maps have the same algebraic properties as the linear transformations between two vector spaces (see Section B.1.2). Specifically, the set of the multilinear maps from $\mathcal{V}_1 \times \cdots \times \mathcal{V}_k$ to $\mathcal{W}$, denoted by $\mathcal{L}(\mathcal{V}_1, \ldots, \mathcal{V}_k; \mathcal{W})$, can be endowed by the usual sum and scalar multiplication.

In particular, the *sum* of $\mathbf{A}, \mathbf{A}' \in \mathcal{L}(\mathcal{V}_1, \ldots, \mathcal{V}_k; \mathcal{W})$ is the k-linear map $\mathbf{A} + \mathbf{A}'$ satisfying

$$(\mathbf{A} + \mathbf{A}')(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \mathbf{A}(\mathbf{v}_1, \ldots, \mathbf{v}_k) + \mathbf{A}(\mathbf{v}_1, \ldots, \mathbf{v}_k)\,, \qquad \forall\, \mathbf{v}_i \in \mathcal{V}_i\,,\ i = 1, \ldots, k\,, \tag{B.18}$$

and the *multiplication* of $\mathbf{A} \in \mathcal{L}(\mathcal{V}_1, \ldots, \mathcal{V}_k; \mathcal{W})$ by the scalar $c \in \mathbb{F}$ is the multilinear map $c\mathbf{A}$ such that

$$(c\mathbf{A})(\mathbf{v}_1, \ldots, \mathbf{v}_k) = c\mathbf{A}(\mathbf{v}_1, \ldots, \mathbf{v}_k)\,,\ \ \forall\, \mathbf{v}_i \in \mathcal{V}_i\,,\ i = 1, \ldots, k\,. \tag{B.19}$$

It is easy to verify that the sum and the scalar multiplication here defined are linear and make $\mathcal{L}(\mathcal{V}_1, \ldots, \mathcal{V}_k; \mathcal{W})$ a vector space.

Moreover, extending again the notions introduced for linear transformations, when the codomain of the k-linear map $\mathbf{A}$ is the field $\mathbb{F}$ itself, that is $\mathbf{A} : \mathcal{V}_1 \times \cdots \times \mathcal{V}_k \to \mathbb{F}$, we more properly say that $\mathbf{A}$ is a *multilinear functional*, or a *multilinear form* (see Halmos (1958)).

Following Lee (2012), as well as Halmos (1958), multilinear forms represent the starting point for introducing the tensor product of transformations.

Definition B.29. Let $\mathcal{V}_1, \ldots, \mathcal{V}_k$ and $\mathcal{W}_1, \ldots, \mathcal{W}_l$ be vector spaces over the field $\mathbb{F}$, and suppose $\mathbf{A} \in \mathcal{L}(\mathcal{V}_1, \ldots, \mathcal{V}_k; \mathbb{F})$ and $\mathbf{B} \in \mathcal{L}(\mathcal{W}_1, \ldots, \mathcal{W}_l; \mathbb{F})$. The *tensor product* of $\mathbf{A}$ and $\mathbf{B}$ is the multilinear map $\mathbf{A} \otimes \mathbf{B}$ such that

$$\begin{aligned} \mathbf{A} \otimes \mathbf{B} :\ & \mathcal{V}_1, \times \cdots \times \mathcal{V}_k \times \mathcal{W}_1 \times \cdots \times \mathcal{W}_l \to \mathbb{F} \\ & (\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{w}_1, \ldots, \mathbf{w}_l) \mapsto \mathbf{A}(\mathbf{v}_1, \ldots, \mathbf{v}_k)\mathbf{B}(\mathbf{w}_1, \ldots, \mathbf{w}_l) \end{aligned} \tag{B.20}$$

One can readily observe that, because of the multilinearity of both $\mathbf{A}$ and $\mathbf{B}$, their tensor product linearly depends on each of the $k+l$ variables. Then, $\mathbf{A} \otimes \mathbf{B}$ is in $\mathcal{L}(\mathcal{V}_1, \dots, \mathcal{V}_k, \mathcal{W}_1, \dots, \mathcal{W}_l; \mathbb{F})$.

Moreover, considering a further multilinear form $\mathbf{C} \in \mathcal{L}(\mathcal{U}_1, \dots, \mathcal{U}_h; \mathbb{F})$, where $\mathcal{U}_1, \dots, \mathcal{U}_h$ are themselves vector spaces over the same field $\mathbb{F}$, we can verify that associativity holds for the tensor product.

Actually, by applying Definition B.29 at arbitrary sets of vectors, say $(\mathbf{u}_1, \dots, \mathbf{u}_h)$, $(\mathbf{v}_1, \dots, \mathbf{v}_k)$ and $(\mathbf{w}_1, \dots, \mathbf{w}_l)$, we observe that the image is a scalar. Hence, the associativity of the multiplication in $\mathbb{F}$ allows one to write

$$\begin{aligned}
\big((\mathbf{C} \otimes \mathbf{A}) \otimes \mathbf{B}\big)&(\mathbf{u}_1, \dots, \mathbf{u}_h, \mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_1, \dots, \mathbf{w}_l) \\
&= (\mathbf{C} \otimes \mathbf{A})(\mathbf{u}_1, \dots, \mathbf{u}_h, \mathbf{v}_1, \dots, \mathbf{v}_k)\, \mathbf{B}(\mathbf{w}_1, \dots, \mathbf{w}_l) \\
&= \mathbf{C}(\mathbf{u}_1, \dots, \mathbf{u}_h)\mathbf{A}(\mathbf{v}_1, \dots, \mathbf{v}_k)\mathbf{B}(\mathbf{w}_1, \dots, \mathbf{w}_l) \\
&= \mathbf{C}(\mathbf{u}_1, \dots, \mathbf{u}_h)\,(\mathbf{A} \otimes \mathbf{B})(\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_1, \dots, \mathbf{w}_l) \\
&= \big(\mathbf{C} \otimes (\mathbf{A} \otimes \mathbf{B})\big)(\mathbf{u}_1, \dots, \mathbf{u}_h, \mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_1, \dots, \mathbf{w}_l)\,.
\end{aligned}$$

which, for the arbitrariness of the considered vectors, implies

$$(\mathbf{C} \otimes \mathbf{A}) \otimes \mathbf{B} = \mathbf{C} \otimes (\mathbf{A} \otimes \mathbf{B}) = \mathbf{C} \otimes \mathbf{A} \otimes \mathbf{B}\,. \tag{B.21}$$

Furthermore, the tensor product can be readily extended to an arbitrary number of multilinear functionals. Specifically, given r functionals $\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(r)}$ depending on $k_1, \dots, k_r$ variables, respectively, their tensor product is the multilinear form $\mathbf{A}$ defined as

$$\mathbf{A} = \bigotimes_{j=1}^{r} \mathbf{A}^{(j)} = \mathbf{A}^{(1)} \otimes \cdots \otimes \mathbf{A}^{(r)}\,. \tag{B.22}$$

Please notice that we are using the notation "$(\cdot)$" as a superscript in order distinguish the functionals involved in the tensor product. Then, being k_j the arity of the j-th functional, we write $\mathbf{A}^{(j)} \in \mathcal{L}\big(\mathcal{V}_1^{(j)}, \dots, \mathcal{V}_{k_j}^{(j)}; \mathbb{F}\big)$, with $j = 1, \dots, r$. In addition, by setting $k = k_1 + \cdots + k_r$, it is clear that $\mathbf{A}$ is a k-linear functional, that is $\mathbf{A} \in \mathcal{L}\big(\mathcal{V}_1^{(1)}, \dots, \mathcal{V}_{k_1}^{(1)}, \dots, \mathcal{V}_1^{(r)}, \dots, \mathcal{V}_{k_r}^{(r)}; \mathbb{F}\big)$.

We further point out that, within each space $\mathcal{L}\big(\mathcal{V}_1^{(j)}, \dots, \mathcal{V}_{k_j}^{(j)}; \mathbb{F}\big)$, the sum of functionals and the scalar multiplication do apply in accordance with (B.18) and (B.19), respectively. Then, considering two arbitrary k_j-linear forms, say $\mathbf{A}^{(j)}$ and $\mathbf{A}^{(j)\prime}$, the following identity holds true:

$$\begin{aligned}&\mathbf{A}^{(1)} \otimes \cdots \otimes \left(\mathbf{A}^{(j)} + \mathbf{A}^{(j)\prime}\right) \otimes \cdots \otimes \mathbf{A}^{(r)} \\ &\quad = \mathbf{A}^{(1)} \otimes \cdots \otimes \mathbf{A}^{(j)} \otimes \cdots \otimes \mathbf{A}^{(r)} + \mathbf{A}^{(1)} \otimes \cdots \otimes \mathbf{A}^{(j)\prime} \otimes \cdots \otimes \mathbf{A}^{(r)} .\end{aligned}$$

Also, for any $c \in \mathbb{F}$, the following relation applies:

$$\mathbf{A}^{(1)} \otimes \cdots \otimes \left(c\mathbf{A}^{(j)}\right) \otimes \cdots \otimes \mathbf{A}^{(r)} = c\left(\mathbf{A}^{(1)} \otimes \cdots \otimes \mathbf{A}^{(j)} \otimes \cdots \otimes \mathbf{A}^{(r)}\right) .$$

The identities here above introduced can be easily proved by considering arbitrary sets of vectors $\left(\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_{k_1}^{(1)}\right), \ldots, \left(\mathbf{v}_1^{(r)}, \ldots, \mathbf{v}_{k_r}^{(r)}\right)$ as arguments of the relevant functionals. Then, since the images are scalars, the properties of the multiplication within the field $\mathbb{F}$ can be applied, i.e. distributivity over addition and commutativity, and both relations are point-wise fulfilled.

Such a result proves, in practice, the linearity of the tensor product with respect to the involved functionals, which is summarized by the following proposition.

Proposition B.30. *Let $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(r)}$ be multilinear forms of arity $k_1, \ldots, k_r$, respectively, defined on the $\mathbb{F}$-vector spaces $\mathcal{V}_1^{(j)}, \ldots, \mathcal{V}_{k_j}^{(j)}$, with $j = 1, \ldots, r$. Then, the tensor product $\mathbf{A}^{(1)} \otimes \cdots \otimes \mathbf{A}^{(r)}$ depends linearly on each multilinear functional.*

Tensor Product of Covectors and Vectors

Let us introduce k vector spaces over the field $\mathbb{F}$, say $\mathcal{V}_1, \ldots, \mathcal{V}_k$, and consider a covector for each of them. Since any covector $\mathbf{w}^*_{(j)}$ is a linear functional for the j-th vector space, we can apply (B.22) to define the tensor product of the covectors $\mathbf{w}^*_{(1)}, \ldots, \mathbf{w}^*_{(k)}$.

Definition B.31. Let $\mathcal{V}_1, \ldots, \mathcal{V}_k$ be $\mathbb{F}$-vector spaces and let $\mathcal{V}_1^*, \ldots, \mathcal{V}_k^*$ be the relevant dual spaces. The *tensor product of the covectors* $\mathbf{w}^*_{(1)}, \ldots, \mathbf{w}^*_{(k)}$

is the multilinear form of $\mathcal{L}(\mathcal{V}_1, \dots, \mathcal{V}_k; \mathbb{F})$ defined as

$$\begin{aligned} \mathbf{w}^*_{(1)} \otimes \cdots \otimes \mathbf{w}^*_{(k)} :\ & \mathcal{V}_1 \times \cdots \times \mathcal{V}_k \to \mathbb{F} \\ & (\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(k)}) \mapsto \mathbf{w}^*_{(1)}(\mathbf{v}^{(1)}) \cdots \mathbf{w}^*_{(k)}(\mathbf{v}^{(k)}) \,, \end{aligned} \tag{B.23}$$

where $\mathbf{w}^*_{(j)} \in \mathcal{V}_j^*$, with $j = 1, \dots, k$.

Remark. Since distinct vector spaces are involved, the index "(j)" will be used to refer to the j-th vector space. We will write, for example, $\mathbf{v}_1^{(j)}, \dots, \mathbf{v}_{n_j}^{(j)}$ to represent n_j distinct vectors of $\mathcal{V}_j$, while $a^1_{(j)}, \dots, a^{n_j}_{(j)}$ will denote a set of scalars of $\mathbb{F}$ by which a linear combination can be performed. Then, consistently with the Einstein summation convention, we will write

$$\mathbf{v}^{(j)} = a^1_{(j)} \mathbf{v}_1^{(j)} + \cdots + a^{n_j}_{(j)} \mathbf{v}_{n_j}^{(j)} = a^{i_j}_{(j)} \mathbf{v}_{i_j}^{(j)} \in \mathcal{V}_j \,.$$

Similarly, when we refer to a linear combination within the dual space $\mathcal{V}_j^*$, we will write

$$\mathbf{w}^*_{(j)} = b_1^{(j)} \mathbf{w}^1_{(j)} + \cdots + b_{n_j}^{(j)} \mathbf{w}^{n_j}_{(j)} = b_{i_j}^{(j)} \mathbf{w}^{i_j}_{(j)} \in \mathcal{V}_j^* \,.$$

The tensor product introduced in Definition B.31 provides a relation between a set of k covectors acting on $\mathcal{V}_1, \dots, \mathcal{V}_k$ to a k-linear functional defined on $\mathcal{V}_1 \times \cdots \times \mathcal{V}_k$. However, in general, such relation is not one-to-one.

In fact, we firstly underline that a multilinear form in $\mathcal{L}(\mathcal{V}_1, \dots, \mathcal{V}_k; \mathbb{F})$ is not always representable as the tensor product of covectors. Moreover, even when a representation of this kind is possible, the relevant expression is not unique. For example, because of the linearity stated in Proposition B.30, the tensor product $c(\mathbf{w}^*_{(1)} \otimes \cdots \otimes \mathbf{w}^*_{(k)})$ can be obtained by multiplying any of the covectors $\mathbf{w}^*_{(1)}, \dots, \mathbf{w}^*_{(k)}$ by the scalar c.

At the same time, we have already seen that the set of multilinear functionals $\mathcal{L}(\mathcal{V}_1, \dots, \mathcal{V}_k; \mathbb{F})$, endowed with the sum (B.18) and the scalar multiplication (B.19), has a vector space structure. In particular, its null element is the functional $\mathbf{0}$ providing the null scalar of $\mathbb{F}$ whatever are the input vectors, that is

$$\mathbf{0}\big(\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(k)}\big)=0\,,\ \ \forall\,\mathbf{v}^{(j)}\in\mathcal{V}_j\,,\ j=1,\ldots,k\,. \tag{B.24}$$

So now we are going to show how the tensor product can be exploited to construct a basis for $\mathcal{L}(\mathcal{V}_1,\ldots,\mathcal{V}_k;\mathbb{F})$ starting from the bases of the vector spaces $\mathcal{V}_1,\ldots,\mathcal{V}_k$ themselves.

Proposition B.32. *Let $\mathcal{V}_1,\ldots,\mathcal{V}_k$ be vector spaces over $\mathbb{F}$ with dimensions $n_1,\ldots,n_k$, respectively, and let $\mathcal{L}(\mathcal{V}_1,\ldots,\mathcal{V}_k;\mathbb{F})$ be the space of the multilinear forms on $\mathcal{V}_1,\ldots,\mathcal{V}_k$. Also, let $\mathcal{B}_j=\big\{\mathbf{b}^{(j)}_1,\ldots,\mathbf{b}^{(j)}_{n_j}\big\}$ a basis for $\mathcal{V}_j$ and $\mathcal{B}^*_j=\big\{\mathbf{b}^1_{(j)},\ldots,\mathbf{b}^{n_j}_{(j)}\big\}$ the relevant dual basis for $\mathcal{V}^*_j$, with $j=1,\ldots,k$. Then, a basis for the space $\mathcal{L}(\mathcal{V}_1,\ldots,\mathcal{V}_k;\mathbb{F})$ is given by*

$$\mathcal{B}_\otimes=\big\{\,\mathbf{b}^{i_1}_{(1)}\otimes\cdots\otimes\mathbf{b}^{i_k}_{(k)}\;\big|\;i_j=1,\ldots,n_j\,,\ j=1,\ldots,k\,\big\}\,. \tag{B.25}$$

Proof. We first prove that the space $\mathcal{L}(\mathcal{V}_1,\ldots,\mathcal{V}_k;\mathbb{F})$ is spanned by $\mathcal{B}_\otimes$. To this end, let us consider a multilinear functional $\mathbf{A}$ applied at the arbitrary set of vectors $(\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(k)})$.

Since a basis $\mathcal{B}_j$ has been fixed, we can express the vector $\mathbf{v}^{(j)}$ of the j-th vector space $\mathcal{V}_j$ as $\mathbf{v}_j=v^{i_j}_{(j)}\mathbf{b}^{(j)}_{i_j}$. Then, for the multilinearity of $\mathbf{A}$, we have

$$\mathbf{A}(\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(k)})=\mathbf{A}\big(v^{i_1}_{(1)}\mathbf{b}^{(1)}_{i_1},\ldots,v^{i_k}_{(k)}\mathbf{b}^{(k)}_{i_k}\big)=v^{i_1}_{(1)}\cdots v^{i_k}_{(k)}\,\mathbf{A}\big(\mathbf{b}^{(1)}_{i_1},\ldots,\mathbf{b}^{(k)}_{i_k}\big)\,.$$

We call $A_{i_1\ldots i_k}$ the value assumed by the functional $\mathbf{A}$ at the vectors of bases $\mathcal{B}_1,\ldots,\mathcal{B}_k$, that is

$$A_{i_1\ldots i_k}=\mathbf{A}\big(\mathbf{b}^{(1)}_{i_1},\ldots,\mathbf{b}^{(k)}_{i_k}\big)\,, \tag{B.26}$$

so that the above equation becomes

$$\mathbf{A}(\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(k)})=v^{i_1}_{(1)}\cdots v^{i_k}_{(k)}\,A_{i_1\ldots i_k}\,.$$

At the same time, any scalar $v^{i_j}_{(j)}$ represents the result of the coordinate function $\mathbf{b}^{i_j}_{(j)}$ applied at $\mathbf{v}^{(j)}$ according to (B.6), that is $v^{i_j}_{(j)}=\mathbf{b}^{i_j}_{(j)}(\mathbf{v}^{(j)})$. Then, applying the property defined by (B.23), one can find

$$\begin{aligned}\mathbf{A}(\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(k)}) &= A_{i_1\ldots i_k}\,\mathbf{b}^{i_1}_{(1)}(\mathbf{v}^{(1)})\cdots\mathbf{b}^{i_k}_{(k)}(\mathbf{v}^{(k)})\\ &= A_{i_1\ldots i_k}\left(\mathbf{b}^{i_1}_{(1)}\otimes\cdots\otimes\mathbf{b}^{i_k}_{(k)}\right)(\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(k)})\,,\end{aligned}$$

and, for the arbitrariness of the vectors $\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(k)}$, one finally obtains the expression of the required linear combination:

$$\mathbf{A} = A_{i_1\ldots i_k}\,\mathbf{b}^{i_1}_{(1)}\otimes\cdots\otimes\mathbf{b}^{i_k}_{(k)}\,.$$

To prove the linear independence of $\mathcal{B}_\otimes$, we consider a linear combination such that

$$A_{i_1\ldots i_k}\,\mathbf{b}^{i_1}_{(1)}\otimes\cdots\otimes\mathbf{b}^{i_k}_{(k)} = \mathbf{0}\,,$$

where $\mathbf{0}\in\mathcal{L}(\mathcal{V}_1,\ldots,\mathcal{V}_k;\mathbb{F})$ is the null multilinear functional defined by the property (B.24).

Hence, using (B.23) to evaluate the above combination at an arbitrary set of vectors $(\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(k)})$, and recalling again that each covector $\mathbf{b}^{i_j}_{(j)}$ represents a coordinate function, defined by (B.6), within the vector space $\mathcal{V}_j$, we have

$$\begin{aligned}A_{i_1\ldots i_k}\left(\mathbf{b}^{i_1}_{(1)}\otimes\cdots\otimes\mathbf{b}^{i_k}_{(k)}\right)\left(\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(k)}\right) &= A_{i_1\ldots i_k}\,\mathbf{b}^{i_1}_{(1)}(\mathbf{v}^{(1)})\cdots\mathbf{b}^{i_k}_{(k)}(\mathbf{v}^{(k)})\\ &= A_{i_1\ldots i_k}\,v^{i_1}_{(1)}\cdots v^{i_k}_{(k)} = 0\,.\end{aligned}$$

Since the arbitrariness of the vectors $\mathbf{v}^{(j)}$ is reflected in the coordinates $v^{i_j}_{(j)}\in\mathbb{F}$, the linear combination here above vanishes only if all the coefficients $A_{i_1\ldots i_k}$ are simultaneously null, that is

$$A_{i_1\ldots i_k} = 0\,,\ \ \forall\, i_j = 1,\ldots,n_j\,,\ j = 1,\ldots,k\,,$$

which proves the linear independence of $\mathcal{B}_\otimes$.

In conclusion, $\mathcal{B}_\otimes$ is actually a basis for $\mathcal{L}(\mathcal{V}_1,\ldots,\mathcal{V}_k;\mathbb{F})$ and the scalar $A_{i_1\ldots i_k}$, evaluated by (B.26), provides the coordinate of the functional $\mathbf{A}$ relevant to the element $\mathbf{b}^{i_1}_{(1)}\otimes\cdots\otimes\mathbf{b}^{i_k}_{(k)}$ of this basis. □

We want to emphasize that any element $\mathbf{b}^{i_1}_{(1)}\otimes\cdots\otimes\mathbf{b}^{i_k}_{(k)}$ of the basis $\mathcal{B}_\otimes$

in (B.25) is uniquely identified once the k indices $i_1, \dots, i_k$ are fixed. This means that all the elements of $\mathcal{B}_\otimes$ can be obtained by varying the functional at the j-th position of the tensor product from $\mathbf{b}^1_{(j)}$ to $\mathbf{b}^{n_j}_{(j)}$, where n_j is the dimension of the space $\mathcal{V}_j$, while all the other $k-1$ functionals are fixed. Explicitly, we can write

$$\mathcal{B}_\otimes = \left\{\mathbf{b}^1_{(1)} \otimes \cdots \otimes \mathbf{b}^1_{(k)},\ \mathbf{b}^2_{(1)} \otimes \cdots \otimes \mathbf{b}^1_{(k)}, \right. \\ \left. \dots,\ \mathbf{b}^{n_1}_{(1)} \otimes \cdots \otimes \mathbf{b}^{n_k-1}_{(k)},\ \mathbf{b}^{n_1}_{(1)} \otimes \cdots \otimes \mathbf{b}^{n_k}_{(k)}\right\},$$

whence it is also clear that the number of elements in $\mathcal{B}_\otimes$, as well as the dimension of $\mathcal{L}(\mathcal{V}_1, \dots, \mathcal{V}_k; \mathbb{F})$ as vector space, is the product $n = n_1 \cdots n_k$.

Similarly to the vector space $\mathcal{V}_j$, the dual space $\mathcal{V}_j^*$ also has its own linear functionals, collected in the double dual space $\mathcal{V}_j^{**}$. For this reason, focusing on the vector structure of the dual spaces, it makes sense to consider the space $\mathcal{L}(\mathcal{V}_1^*, \dots, \mathcal{V}_k^*; \mathbb{F})$ gathering the multilinear forms on the dual spaces $\mathcal{V}_1^*, \dots, \mathcal{V}_k^*$.

In addition, since any double dual space $\mathcal{V}_j^{**}$ is identified with $\mathcal{V}_j$ by the canonical isomorphism (B.15), any vector $\mathbf{v}^{(j)} \in \mathcal{V}_j$ can be seen as a linear functional on the dual space $\mathcal{V}_j^*$.

Consequently, we can introduce the following definition of tensor product of vectors.

Definition B.33. Given the vector spaces $\mathcal{V}_1, \dots, \mathcal{V}_k$ over the field $\mathbb{F}$, with the relevant dual spaces $\mathcal{V}_1^*, \dots, \mathcal{V}_k^*$, we define the *tensor product of the vectors* $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(k)}$ as the multilinear functional in $\mathcal{L}(\mathcal{V}_1^*, \dots, \mathcal{V}_k^*; \mathbb{F})$ such that

$$\begin{aligned} \mathbf{v}^{(1)} \otimes \cdots \otimes \mathbf{v}^{(k)} :\ & \mathcal{V}_1^* \times \cdots \times \mathcal{V}_k^* \to \mathbb{F} \\ & \left(\mathbf{w}^*_{(1)}, \dots, \mathbf{w}^*_{(k)}\right) \mapsto \mathbf{w}^*_{(1)}\left(\mathbf{v}^{(1)}\right) \cdots \mathbf{w}^*_{(k)}\left(\mathbf{v}^{(k)}\right), \end{aligned} \tag{B.27}$$

where $\mathbf{v}^{(j)}$ is an element of the vector space $\mathcal{V}_j$, for any j from 1 to k.

Of course, by virtue of the linearity of the spaces and the maps involved in the above definition, the same properties specified for the tensor product of covectors also apply to the product of vectors. Specifically, a basis for the linear space $\mathcal{L}(\mathcal{V}_1^*, \dots, \mathcal{V}_k^*; \mathbb{F})$ is given by

$$\mathcal{B}'_{\otimes} = \left\{ \mathbf{b}_{i_1}^{(1)} \otimes \cdots \otimes \mathbf{b}_{i_k}^{(k)} \,\middle|\, i_j = 1, \dots, n_j \,,\; j = 1, \dots, k \right\}, \tag{B.28}$$

where $\mathbf{b}_{i_j}^{(j)}$ is the i_j-th vector of the basis $\mathcal{B}_j = \left\{\mathbf{b}_1^{(j)}, \dots, \mathbf{b}_{n_j}^{(j)}\right\}$ for $\mathcal{V}_j$.

In order to emphasize the structure of $\mathcal{L}(\mathcal{V}_1^*, \dots, \mathcal{V}_k^*; \mathbb{F})$ as the space spanned by the tensor products $\mathbf{b}_{i_1}^{(1)} \otimes \cdots \otimes \mathbf{b}_{i_k}^{(k)}$, it will be referred to as the *tensor product space* of $\mathcal{V}_1, \dots, \mathcal{V}_k$, to be denoted as $\mathcal{V}_1 \otimes \cdots \otimes \mathcal{V}_k$.

Similarly, the space of linear functionals $\mathcal{L}(\mathcal{V}_1, \dots, \mathcal{V}_k; \mathbb{F})$ can be intended as the tensor product space of the dual spaces $\mathcal{V}_1^*, \dots, \mathcal{V}_k^*$, and the notation $\mathcal{V}_1^* \otimes \cdots \otimes \mathcal{V}_k^*$ will be used.

Remark. The tensor product of the dual vector spaces $\mathcal{V}_1^*, \dots, \mathcal{V}_k^*$ provides, strictly speaking, an algebraic structure distinct from the space of multilinear forms, namely $\mathcal{L}(\mathcal{V}_1, \dots, \mathcal{V}_k; \mathbb{F})$. To distinguish these two types of algebraic structures, in Lee (2012) the nomenclature *abstract tensor product space* is used to refer to $\mathcal{V}_1^* \otimes \cdots \otimes \mathcal{V}_k^*$.

At the same times, it can be proved that the two spaces are canonically isomorphic, justifying the identification $\mathcal{V}_1^* \otimes \cdots \otimes \mathcal{V}_k^* = \mathcal{L}(\mathcal{V}_1, \dots, \mathcal{V}_k; \mathbb{F})$ previously mentioned. The same also applies to the tensor product of the vector spaces $\mathcal{V}_1, \dots, \mathcal{V}_k$, so that it is legitimate to set $\mathcal{V}_1 \otimes \cdots \otimes \mathcal{V}_k = \mathcal{L}(\mathcal{V}_1^*, \dots, \mathcal{V}_k^*; \mathbb{F})$.

Further details about the abstract and concrete definitions of the tensor product can be found, among others, in Dimitrienko (2002), Halmos (1958), Lee (2012), Mac Lane and Birkhoff (1999), Winitzki (2020).

From a more general perspective, since any dual vector space has itself a linear space structure, the following identification can be applied:

$$\mathcal{W}_1 \otimes \cdots \otimes \mathcal{W}_k = \mathcal{L}(\mathcal{W}_1^*, \dots, \mathcal{W}_k^*; \mathbb{F}), \tag{B.29}$$

where each $\mathcal{W}_i$ can be intended as an arbitrary vector space $\mathcal{V}_i$, and than $\mathcal{W}_i^* = \mathcal{V}_i^*$ is the i-th dual space, or as a dual space $\mathcal{V}_i^*$, in which case $\mathcal{W}_i^*$ represents the double dual $\mathcal{V}_i^{**} \cong \mathcal{V}_i$.

Consistently, generalizing Proposition B.32, a basis for the tensor space in (B.29) is provided as

$$\mathcal{B}_{\otimes} = \left\{ \mathbf{b}_{i_1} \otimes \cdots \otimes \mathbf{b}_{i_k} \,\middle|\, i_j = 1, \ldots, n_j \,,\ j = 1, \ldots, k \right\}. \tag{B.30}$$

where each vector $\mathbf{b}_{i_j}$ should be intended as the j-th vector $\mathbf{b}_{(i)}^{i_j}$ of the i-th basis $\mathcal{B}_i$ when $\mathcal{W}_i = \mathcal{V}_i$, or as the j-th covector $\mathbf{b}_{i_j} = \mathbf{b}_{i_j}^{(i)}$ of the i-th dual basis $\mathcal{B}_i^*$ when $\mathcal{W}_i = \mathcal{V}_i^*$.

Tensors on a Vector Space

Suppose a vector space $\mathcal{V}$ is assigned. The tensor product of covector can be defined considering k copies of $\mathcal{V}$, along with its dual space $\mathcal{V}^*$.

Definition B.34. Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$ and let $\mathcal{V}^*$ be its dual space. A *covariant k-tensor on $\mathcal{V}$* is a multilinear form whose domain is the k-th Cartesian power of $\mathcal{V}$:

$$\mathbf{T} \colon \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_{k \text{ times}} \to \mathbb{F}\,,$$

and k is the *rank* of the covariant tensor $\mathbf{T}$.

The set collecting all the covariant tensors of rank k, denoted as $\mathcal{T}^k(\mathcal{V}^*)$, is the tensor product space given by k copies of $\mathcal{V}^*$:

$$\mathcal{T}^k(\mathcal{V}^*) = \underbrace{\mathcal{V}^* \otimes \cdots \otimes \mathcal{V}^*}_{k \text{ times}}\,. \tag{B.31}$$

Similarly, we can introduce a multilinear form defined on the k copies of the dual space $\mathcal{V}^*$.

Definition B.35. Let $\mathcal{V}$ be an $\mathbb{F}$-vector space, with the dual space $\mathcal{V}^*$. A *contravariant k-tensor on $\mathcal{V}$* is a multilinear functional defined on the k-th Cartesian power of $\mathcal{V}^*$:

$$\mathbf{T} \colon \underbrace{\mathcal{V}^* \times \cdots \times \mathcal{V}^*}_{k \text{ times}} \to \mathbb{F}\,,$$

and we say that the contravariant tensor $\mathbf{T}$ has *rank* k.

Recalling the role of $\mathcal{V}$ as the space of functionals on $\mathcal{V}^*$, up to a canonical isomorphism, the tensor product space given by k copies of $\mathcal{V}$ provides the space of the contravariant k-tensors. Such a space is referred to using the notation $\mathcal{T}^k(\mathcal{V})$:

$$\mathcal{T}^k(\mathcal{V}) = \underbrace{\mathcal{V} \otimes \cdots \otimes \mathcal{V}}_{k \text{ times}} . \tag{B.32}$$

Both the spaces of covariant and contravariant tensors on $\mathcal{V}$ can be seen as the specification of a more general space. Actually, given the non-negative integers p and q such that $p+q=k$, we introduce the space of *mixed tensors on* $\mathcal{V}$ *of rank* (p,q). Such a space, denoted as $\mathcal{T}^{(p,q)}$, represents the tensor product space of p copies of $\mathcal{V}$ and q copies of $\mathcal{V}^*$:

$$\mathcal{T}^{(p,q)}(\mathcal{V}) = \underbrace{\mathcal{V} \otimes \cdots \otimes \mathcal{V}}_{p \text{ times}} \otimes \underbrace{\mathcal{V}^* \otimes \cdots \otimes \mathcal{V}^*}_{q \text{ times}} . \tag{B.33}$$

Any tensor $\mathbf{T} \in \mathcal{T}^{(p,q)}$ is k-linear functional defined on the Cartesian product of p times $\mathcal{V}^*$ and q times $\mathcal{V}$:

$$\mathbf{T} \colon \underbrace{\mathcal{V}^* \times \cdots \times \mathcal{V}^*}_{p \text{ times}} \times \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_{q \text{ times}} \to \mathbb{F} ,$$

where p and q are said the *order of covariance* and the *order of contravariance* of $\mathbf{T}$, respectively.

Comparing (B.33) with (B.31) and (B.32), it is clear that a tensor with rank $(k,0)$ is a contravariant k-tensor:

$$\mathcal{T}^{(k,0)}(\mathcal{V}) = \mathcal{T}^k(\mathcal{V}) ,$$

as well as the rank $(0,k)$ represents a covariant k-tensor:

$$\mathcal{T}^{(0,k)}(\mathcal{V}) = \mathcal{T}^k(\mathcal{V}^*) .$$

Please notice that when $k=1$, the tensor product space defined by (B.31) specializes to

$$\mathcal{T}^{(0,1)}(\mathcal{V}) = \mathcal{T}^1(\mathcal{V}^*) = \mathcal{V}^* ,$$

as also (B.32) implies

$$\mathcal{T}^{(1,0)}(\mathcal{V}) = \mathcal{T}^1(\mathcal{V}) = \mathcal{V}\,,$$

meaning that the covectors in $\mathcal{V}^*$ are covariant 1-tensors on $\mathcal{V}$, while the vectors in $\mathcal{V}$ are contravariant 1-tensors.

Finally, it is assumed by convention that when the rank vanishes, both the spaces of covariant and contravariant tensors coincide with the field $\mathbb{F}$:

$$\mathcal{T}^{(0,0)}(\mathcal{V}) = \mathcal{T}^0(\mathcal{V}^*) = \mathcal{T}^0(\mathcal{V}) = \mathbb{F}\,,$$

so that any 0-tensor, being a functional of arity $k = 0$, can be thought as a constant in $\mathbb{F}$.

It is worth pointing out that $\mathcal{T}^{(p,q)}(\mathcal{V})$, representing a specialization of the tensor product space of k distinct vector spaces, inherits the property of being a linear space itself. In particular, Proposition B.32 readily results in the following corollary.

Corollary B.36. *Let $\mathcal{V}$ be an n-dimensional vector space over the field $\mathbb{F}$ and let $\mathcal{V}^*$ be the dual space. Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis for $\mathcal{V}$ and $\mathcal{B}^* = \{\mathbf{b}^1, \ldots, \mathbf{b}^n\}$ is the relevant dual basis for $\mathcal{V}^*$. Then, a basis for the space $\mathcal{T}^{(p,q)}(\mathcal{V})$ of the tensors on $\mathcal{V}$ of rank (p, q) is given by*

$$\begin{aligned}\mathcal{B}^{(p,q)} = \big\{\, &\mathbf{b}_{i_1} \otimes \cdots \otimes \mathbf{b}_{i_p} \otimes \mathbf{b}^{j_1} \otimes \cdots \otimes \mathbf{b}^{j_q} \;\big|\; \\ &i_1, \ldots, i_p = 1, \ldots, n\,,\; j_1, \ldots, j_q = 1, \ldots, n \,\big\}\,,\end{aligned} \tag{B.34}$$

and the dimension of $\mathcal{T}^{(p,q)}(\mathcal{V})$ is n^k, with $k = p + q$.

With explicit reference to the space of covariant and contravariant tensors, a basis for $\mathcal{T}^k(\mathcal{V}^*)$ is

$$\mathcal{B}^{(0,k)} = \big\{\, \mathbf{b}^{j_1} \otimes \cdots \otimes \mathbf{b}^{j_k} \;\big|\; j_1, \ldots, j_k = 1, \ldots, n \,\big\}\,, \tag{B.35}$$

as well as a basis for $\mathcal{T}^k(\mathcal{V})$ is

$$\mathcal{B}^{(k,0)} = \big\{\, \mathbf{b}_{i_1} \otimes \cdots \otimes \mathbf{b}_{i_k} \;\big|\; i_1, \ldots, i_k = 1, \ldots, n \,\big\}\,. \tag{B.36}$$

It is also relevant to observe that, following the proof of Proposition B.32, any tensor $\mathbf{T} \in \mathcal{T}^{(p,q)}$ can be expressed in the form

$$\mathbf{T} = T^{i_1 \dots i_p}{}_{j_1 \dots j_q} \mathbf{b}_{i_1} \otimes \cdots \otimes \mathbf{b}_{i_p} \otimes \mathbf{b}^{j_1} \otimes \cdots \otimes \mathbf{b}^{j_q} \,,$$

where one can derive the n^k coordinates $T^{i_1 \dots i_p}{}_{j_1 \dots j_q}$ with respect to $\mathcal{B}^{(p,q)}$ by applying the tensor $\mathbf{T}$ to a collection of p covectors of the dual basis $\mathcal{B}^*$ and q vectors of the basis $\mathcal{B}$:

$$T^{i_1 \dots i_p}{}_{j_1 \dots j_q} = \mathbf{T}(\mathbf{b}^{i_1}, \dots, \mathbf{b}^{i_p}, \mathbf{b}_{j_1}, \dots, \mathbf{b}_{j_q}) \,. \tag{B.37}$$

B.2.2 Symmetric and Alternating Tensors

We have seen that a tensor on a linear space $\mathcal{V}$ is a multilinear map to be applied to an ordered set of vectors which, with reference to covariant or contravariant tensors, are elements of $\mathcal{V}$ or $\mathcal{V}^*$, respectively. When the order of the input vectors is changed, the new result is not generally related with the former one.

However, in some special cases, it is possible that the scalar resulting from applying a tensor does not change when the arguments are switched, or only its sign changes.

Such properties represent the basis for the definition of symmetric and alternating tensors (see, e.g., Dimitrienko (2002), Lee (2012)).

Definition B.37. Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$ and consider the k-tensors in $\mathcal{T}^k(\mathcal{V}^*)$ and $\mathcal{T}^k(\mathcal{V})$.

- A k-tensor is called *symmetric* if it is invariant under any permutation of its arguments.
- A k-tensor is *alternating*, also called *antisymmetric* or *skew-symmetric*, if any interchange of two arguments results in a change of the sign.

Since the possibility to recognize symmetric and alternating tensors relies on how their value changes when the input vectors are interchanged, in order to give a formal characterization some preliminary notions are required.

Moreover, we first proceed with reference to covariant tensors and then we will specify the analogous notions to contravariant tensors.

Permutations on Tensors

Given the set of integers $\{1, \ldots, k\}$, we consider the group of permutations of its elements, denoted as S_k, which is called the *symmetric group on k elements*.

For clarity, we recall that a permutation on $\{1, \ldots, k\}$ is a bijective map in the form $\sigma : i \mapsto j = \sigma(i)$, where j is itself an integer between 1 and k. Moreover, a permutation interchanging two distinct elements of $\{1, \ldots, k\}$, while all the others remain fixed, is called a *transposition*.

Since a transposition represents the simplest transformation changing the order of k elements, any permutation $\sigma \in S_k$ can be decomposed into a finite sequence of transpositions. Moreover, the parity of the number of transpositions providing σ defines the property of such permutation of being even or odd (see, e.g., Dummit and Foote (2003), Lang (2002)).

Consequently, a sign function is associated with any permutation σ of S_k such that

$$\mathrm{sgn}(\sigma) = \begin{cases} +1 & \text{if } \sigma \text{ is even}\,, \\ -1 & \text{if } \sigma \text{ is odd}\,. \end{cases} \tag{B.38}$$

Permutations on Covariant Tensors With reference to the Cartesian power $\mathcal{V}^k$, a permutation of indices by $\sigma \in S_k$ results in rearranging the entries of $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ to obtain the new ordered set $(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(k)})$.

Then, any permutation $\sigma \in S_k$ induces a map

$$\begin{aligned} \varphi_\sigma :\ \mathcal{T}^k(\mathcal{V}^*) &\to \mathcal{T}^k(\mathcal{V}^*) \\ \mathbf{T} &\mapsto {}^\sigma\mathbf{T}\,, \end{aligned} \tag{B.39}$$

such that the resulting tensor ${}^\sigma\mathbf{T}$ is defined by the following property:

$${}^\sigma\mathbf{T}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \mathbf{T}(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(k)})\,,\ \ \forall\,(\mathbf{v}_1, \ldots, \mathbf{v}_k) \in \mathcal{V}^k\,. \tag{B.40}$$

Suppose to express a covariant tensor in $\mathcal{T}^k(\mathcal{V}^*)$ in the form of the tensor product of the covectors $\mathbf{w}^1, \ldots, \mathbf{w}^k \in \mathcal{V}^*$. Applying the permutation map φ_σ to the tensor $\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k$, and using the defining property (B.23) at an arbitrary set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$, one finds

$$\begin{aligned} {}^\sigma(\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k)(\mathbf{v}_1, \ldots, \mathbf{v}_k) &= (\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k)(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(k)}) \\ &= \mathbf{w}^1(\mathbf{v}_{\sigma(1)}) \cdots \mathbf{w}^k(\mathbf{v}_{\sigma(k)}) \,. \end{aligned}$$

Since the multiplication of scalars is commutative, the last expression can be written with the factors in a different order, which means permuting the indices $1, \ldots, k$ and $\sigma(1), \ldots, \sigma(k)$ of the covectors and the vectors, respectively Specifically, using the inverse permutation σ^{-1}, one obtains

$$\mathbf{w}^1(\mathbf{v}_{\sigma(1)}) \cdots \mathbf{w}^k(\mathbf{v}_{\sigma(k)}) = \mathbf{w}^{\sigma^{-1}(1)}(\mathbf{v}_1) \cdots \mathbf{w}^{\sigma^{-1}(k)}(\mathbf{v}_k) \,,$$

whence it is easy to find

$${}^\sigma(\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k)(\mathbf{v}_1, \ldots, \mathbf{v}_k) = (\mathbf{w}^{\sigma^{-1}(1)} \otimes \cdots \otimes \mathbf{w}^{\sigma^{-1}(k)})(\mathbf{v}_1, \ldots, \mathbf{v}_k) \,,$$

and, for the arbitrariness of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$, one finally has

$${}^\sigma(\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k) = \mathbf{w}^{\sigma^{-1}(1)} \otimes \cdots \otimes \mathbf{w}^{\sigma^{-1}(k)} \,. \tag{B.41}$$

Let us now introduce the basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ for the vector space $\mathcal{V}$, along with the basis $\mathcal{B}^* = \{\mathbf{b}^1, \ldots, \mathbf{b}^n\}$ for the dual space $\mathcal{V}^*$. We recall that a basis for the tensor space $\mathcal{T}^k(\mathcal{V}^*)$ is given by (B.35) and collects the tensors $\mathbf{b}^{j_1} \otimes \cdots \otimes \mathbf{b}^{j_k}$, with $j_1, \ldots, j_k$ varying from 1 to n.

Also, the coordinates of a tensor $\mathbf{T} \in \mathcal{T}^k(\mathcal{V}^*)$ are the scalars $T_{j_1 \ldots j_k}$, which can be obtained, specializing the property (B.37) to tensors of rank $(0, k)$, by evaluating $\mathbf{T}$ at the vectors $\mathbf{b}_{j_1}, \ldots, \mathbf{b}_{j_k}$ of the basis $\mathcal{B}$:

$$T_{j_1 \ldots j_k} = \mathbf{T}(\mathbf{b}_{j_1}, \ldots, \mathbf{b}_{j_k}) \,.$$

The same property applies to the permuted tensor ${}^\sigma\mathbf{T}$, that is

$${}^\sigma T_{j_1 \ldots j_k} = {}^\sigma\mathbf{T}(\mathbf{b}_{j_1}, \ldots, \mathbf{b}_{j_k}) \,,$$

so that, being ${}^{\sigma}\mathbf{T}(\mathbf{b}_{j_1}, \ldots, \mathbf{b}_{j_k}) = \mathbf{T}(\mathbf{b}_{\sigma(j_1)}, \ldots, \mathbf{b}_{\sigma(j_k)})$, we find

$$ {}^{\sigma}T_{j_1 \ldots j_k} = T_{\sigma(j_1) \ldots \sigma(j_k)} \,, \tag{B.42} $$

meaning that the coordinates of the permuted tensor ${}^{\sigma}\mathbf{T}$ can be obtained by permuting the ones of $\mathbf{T}$.

Permutations on Contravariant Tensors Given a contravariant tensor $\mathbf{T} \in \mathcal{T}^k(\mathcal{V})$, we denote as ${}^{\sigma}\mathbf{T}$ the relevant tensor resulting from a permutation $\sigma \in S_k$ acting on the arguments, that is

$$ {}^{\sigma}\mathbf{T}(\mathbf{w}^1, \ldots, \mathbf{w}^k) = \mathbf{T}(\mathbf{w}^{\sigma(1)}, \ldots, \mathbf{w}^{\sigma(k)}) \,, \;\; \forall\, (\mathbf{w}^1, \ldots, \mathbf{w}^k) \in {\mathcal{V}^*}^k \,, \tag{B.43} $$

which is reflected in the following transformation of coordinates:

$$ {}^{\sigma}T^{j_1 \ldots j_k} = T^{\sigma(j_1) \ldots \sigma(j_k)} \,. \tag{B.44} $$

Also, when a contravariant tensor is provided by the product of k vectors of $\mathcal{V}$, say $\mathbf{v}_1, \ldots, \mathbf{v}_k$, the action of $\sigma \in S_k$ results in the inverse permutation of the factors:

$$ {}^{\sigma}(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k) = \mathbf{v}_{\sigma^{-1}(1)} \otimes \cdots \otimes \mathbf{v}_{\sigma^{-1}(k)} \,. \tag{B.45} $$

Symmetric and Alternating Covariant Tensors

In this section we introduce a formal expression for both symmetric and alternating tensors by exploiting the permuting map defined by (B.39) and the properties of permutations on covariant tensors.

Symmetric Covariant Tensors With regard to the first part of Definition B.37, we say that a covariant k-tensor $\mathbf{S} \in \mathcal{T}^k(\mathcal{V}^*)$ is symmetric if it satisfies the condition

$$ {}^{\sigma}\mathbf{S} = \mathbf{S} \,, \;\; \forall\, \sigma \in S_k \,, \tag{B.46} $$

where ${}^{\sigma}\mathbf{S}$ is the tensor resulting from the permutation σ, as defined by (B.40), and S_k is the group of permutations on k elements.

Explicitly, the value of $\mathbf{S}$ at an arbitrary ordered set $(\mathbf{v}_1, \ldots, \mathbf{v}_k) \in \mathcal{V}^k$ remains the same if any pair of entries are switched:

$$\mathbf{S}(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_k) = \mathbf{S}(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_k)\,, \qquad \forall\, i, j = 1, \ldots, k\,.$$

Moreover, in terms of coordinates with respect to a basis $\mathcal{B}^{(0,k)}$ for $\mathcal{T}^k(\mathcal{V}^*)$, by means of (B.42), the defining property (B.46) reads

$$S_{\sigma(j_1)\ldots\sigma(j_k)} = S_{j_1\ldots j_k}\,, \ \ \forall\, \sigma \in S_k\,. \tag{B.47}$$

At the same time, if the symmetric tensor $\mathbf{S}$ can be represented as the product of k covectors, say $\mathbf{w}^1, \ldots, \mathbf{w}^k \in \mathcal{V}^*$, the following property holds:

$$\mathbf{w}^{\sigma(1)} \otimes \cdots \otimes \mathbf{w}^{\sigma(k)} = \mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k\,, \ \ \forall\, \sigma \in S_k\,, \tag{B.48}$$

which comes from the property (B.41) and observing that, since any permutation in S_k is considered, both σ and the inverse σ^{-1} are included.

The set of the symmetric covariant k-tensors on the vector space $\mathcal{V}$ is denoted as $\Sigma^k(\mathcal{V}^*)$. Such a space is not only a subset of $\mathcal{T}^k(\mathcal{V}^*)$, but actually a linear subspace.

To prove the vector space structure of $\Sigma^k(\mathcal{V}^*)$, we show that it is closed under the sum and the scalar multiplication, i.e. that such operations are both consistent with the defining property (B.46).

In fact, using the linearity of the covariant tensors, it is easy to verify that the sum of any pair of symmetric tensors $\mathbf{S}, \mathbf{S}' \in \Sigma^k(\mathcal{V}^*)$ is symmetric, that is

$$^{\sigma}(\mathbf{S} + \mathbf{S}') = {}^{\sigma}\mathbf{S} + {}^{\sigma}\mathbf{S}' = \mathbf{S} + \mathbf{S}'\,, \ \ \forall\, \sigma \in S_k\,,$$

and that the product of a symmetric tensor $\mathbf{S} \in \Sigma^k(\mathcal{V}^*)$ by a scalar $c \in \mathbb{F}$ is itself symmetric, i.e.

$$^{\sigma}(c\mathbf{S}) = c\,{}^{\sigma}\mathbf{S} = c\mathbf{S}\,, \ \ \forall\, \sigma \in S_k\,.$$

The map projecting of a covariant k-tensor $\mathbf{T}$ from $\mathcal{T}^k(\mathcal{V}^*)$ to $\Sigma^k(\mathcal{V}^*)$ is defined the *symmetrization* of $\mathbf{T}$:

$$\begin{aligned} \mathsf{Sym}: \mathcal{T}^k(\mathcal{V}^*) &\to \Sigma^k(\mathcal{V}^*) \\ \mathbf{T} \mapsto \mathsf{Sym}\,\mathbf{T} &= \frac{1}{k!}\sum_{\sigma\in S_k} {}^{\sigma}\mathbf{T}\,, \end{aligned} \tag{B.49}$$

which explicitly means

$$(\mathsf{Sym}\,\mathbf{T})(\mathbf{v}_1,\ldots,\mathbf{v}_k) = \frac{1}{k!}\sum_{\sigma\in S_k} \mathbf{T}(\mathbf{v}_{\sigma(1)},\ldots,\mathbf{v}_{\sigma(k)})\,,\ \ \forall\,(\mathbf{v}_1,\ldots,\mathbf{v}_k)\in\mathcal{V}^k\,.$$

Alternating Covariant Tensors We formalize the second point of Definition B.37 by saying that an alternating covariant k-tensor $\mathbf{W}\in\mathcal{T}^k(\mathcal{V}^*)$ is such that

$${}^{\sigma}\mathbf{W} = \mathsf{sgn}(\sigma)\,\mathbf{W}\,,\ \ \forall\,\sigma\in S_k\,, \tag{B.50}$$

where ${}^{\sigma}\mathbf{W}$ comes from the application of the permutation map (B.39) to the k-tensor $\mathbf{W}$, and $\mathsf{sgn}(\sigma)$ is the sign of the permutation σ defined by (B.38).

Another way to characterize the alternating tensor $\mathbf{W}$ consists in considering an ordered set $(\mathbf{v}_1,\ldots,\mathbf{v}_k)\in\mathcal{V}^k$ and switching an arbitrary pair of distinct vectors. Then, the values provided by $\mathbf{W}$ are the one the opposite of the other:

$$\mathbf{W}(\mathbf{v}_1,\ldots,\mathbf{v}_i,\ldots,\mathbf{v}_j,\ldots,\mathbf{v}_k) = -\mathbf{W}(\mathbf{v}_1,\ldots,\mathbf{v}_j,\ldots,\mathbf{v}_i,\ldots,\mathbf{v}_k)\,,\qquad \forall\,i,j=1,\ldots,k\,,\ i\neq j\,.$$

In addition, when we refer to a basis $\mathcal{B}^{(0,k)}$ of $\mathcal{T}^k(\mathcal{V}^*)$, the coordinates of the alternating tensor $\mathbf{W}$ satisfies the following condition:

$$W_{\sigma(j_1)\ldots\sigma(j_k)} = \mathsf{sgn}(\sigma)\,W_{j_1\ldots j_k}\,,\ \ \forall\,\sigma\in S_k\,, \tag{B.51}$$

which comes from applying the property (B.42) to the definition (B.50).

We observe as well that when an alternating k-tensor can be expressed as the tensor product of $\mathbf{w}^1,\ldots,\mathbf{w}^k\in\mathcal{V}^*$, the property (B.41) implies

$$\mathbf{w}^{\sigma(1)}\otimes\cdots\otimes\mathbf{w}^{\sigma(k)} = \mathsf{sgn}(\sigma)\,(\mathbf{w}^1\otimes\cdots\otimes\mathbf{w}^k)\,,\ \ \forall\,\sigma\in S_k\,, \tag{B.52}$$

which also accounts that for any permutation in S_k the inverse map σ^{-1} has the same sign as σ.

Similarly to the symmetric tensors, the set of the alternating covariant k-tensors, denoted as $\Lambda^k(\mathcal{V}^*)$, has a vector space structure and is a subspace of $\mathcal{T}^k(\mathcal{V}^*)$ as well. Actually, exploiting the linearity of the covariant tensors, it can be proved that $\Lambda^k(\mathcal{V}^*)$ is closed under the sum and the scalar multiplication.

Specifically, one can verify that the sum of any pair of alternating tensors, say $\mathbf{W}$ and $\mathbf{W}'$, is in $\Lambda^k(\mathcal{V}^*)$, resulting

$$^{\sigma}(\mathbf{W}+\mathbf{W}') = {}^{\sigma}\mathbf{W} + {}^{\sigma}\mathbf{W}' = \operatorname{sgn}(\sigma)\,(\mathbf{W}+\mathbf{W}')\,,\ \ \forall\,\sigma \in S_k\,,$$

as well as the product of a tensor $\mathbf{W} \in \Lambda^k(\mathcal{V}^*)$ by any scalar $c \in \mathbb{F}$ is also alternating:

$$^{\sigma}(c\mathbf{W}) = c\,{}^{\sigma}\mathbf{W} = \operatorname{sgn}(\sigma)\,(c\mathbf{W})\,,\ \ \forall\,\sigma \in S_k\,.$$

In addition, it is possible to construct an alternating tensor by means of the following map:

$$\begin{aligned} \operatorname{Alt}:\ & \mathcal{T}^k(\mathcal{V}^*) \to \Lambda^k(\mathcal{V}^*) \\ & \mathbf{T} \mapsto \operatorname{Alt}\mathbf{T} = \frac{1}{k!}\sum_{\sigma\in S_k}\operatorname{sgn}(\sigma)\,{}^{\sigma}\mathbf{T}\,, \end{aligned} \tag{B.53}$$

which provides the projection from $\mathcal{T}^k(\mathcal{V}^*)$ to $\Lambda^k(\mathcal{V}^*)$ and is called the *alternation* of $\mathbf{T}$.

Then, the alternation of $\mathbf{T}$ is the k-tensor satisfying

$$(\operatorname{Alt}\mathbf{T})(\mathbf{v}_1,\ldots,\mathbf{v}_k) = \frac{1}{k!}\sum_{\sigma\in S_k}\operatorname{sgn}(\sigma)\,\mathbf{T}(\mathbf{v}_{\sigma(1)},\ldots,\mathbf{v}_{\sigma(k)})\,, \quad \forall\,(\mathbf{v}_1,\ldots,\mathbf{v}_k)\in\mathcal{V}^k\,. \tag{B.54}$$

Please notice that when k covectors $\mathbf{w}^1,\ldots,\mathbf{w}^k \in \mathcal{V}^*$ are considered, the alternation of the resulting covariant tensor can be expressed as

$$\mathrm{Alt}(\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k) = \frac{1}{k!} \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma)\, \mathbf{w}^{\sigma(1)} \otimes \cdots \otimes \mathbf{w}^{\sigma(k)}\,, \tag{B.55}$$

which comes from (B.53) observing that the permutation by σ of the tensor product is provided by (B.41) as $\mathbf{w}^{\sigma^{-1}(1)} \otimes \cdots \otimes \mathbf{w}^{\sigma^{-1}(k)}$ and that, in considering all the permutations in S_k, both σ and the inverse σ^{-1} are included.

Remark. Both the symmetrization and the alternation of a tensor $\mathbf{T}$, defined by (B.49) and (B.53), respectively, are characterized by a summation to be performed on all the permutations in S_k.

In fact, the summation has the role of taking into account all the possible interchanges of the k arguments when the tensors $\mathsf{Sym}\,\mathbf{T}$ and $\mathsf{Alt}\,\mathbf{T}$ are evaluated. Since the total number of such permutations is exactly $k!$, the symmetrization of $\mathbf{T}$, as well as its alternation, is actually defined as the average of all its possible permuted versions.

The only difference to be pointed out is that in the alternation map each permutation of $\mathbf{T}$ is considered along with its parity, so that a skew-symmetric tensor is actually obtained.

Symmetric and Alternating Contravariant Tensors

The definitions and the properties so far introduced for the covariant k-tensors on $\mathcal{V}$ can be easily extended to the contravariant ones. Here we report the main features, which can be verified proceeding similarly to the covariant tensors.

Specializing Definition B.37 to contravariant k-tensors, we say that the tensor $\mathbf{S} \in \mathcal{T}^k(\mathcal{V})$ is symmetric if the following property applies:

$${}^{\sigma}\mathbf{S} = \mathbf{S}\,, \ \ \forall\, \sigma \in S_k\,, \tag{B.56}$$

or, in terms of coordinates with respect to a basis $\mathcal{B}^{(k,0)}$,

$$S^{\sigma(i_1)\ldots\sigma(i_k)} = S^{i_1\ldots i_k}\,, \ \ \forall\, \sigma \in S_k\,. \tag{B.57}$$

A further property of the a symmetric contravariant tensor arises when it can be written as the tensor product of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathcal{V}$:

$$\mathbf{v}_{\sigma(1)} \otimes \cdots \otimes \mathbf{v}_{\sigma(k)} = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k\,, \quad \forall\, \sigma \in S_k\,. \tag{B.58}$$

Furthermore, recalling that the sign of a permutation is given by (B.38), a contravariant k-tensor $\mathbf{W} \in \mathcal{T}^k(\mathcal{V})$ is alternating if it satisfies the condition

$$^{\sigma}\mathbf{W} = \operatorname{sgn}(\sigma)\,\mathbf{W}\,, \quad \forall\, \sigma \in S_k\,, \tag{B.59}$$

which is reflected in the coordinates as follows:

$$W^{\sigma(j_1)\dots\sigma(j_k)} = \operatorname{sgn}(\sigma)\,W^{j_1\dots j_k}\,, \quad \forall\, \sigma \in S_k\,. \tag{B.60}$$

Additionally, supposing that a tensor is given by the product of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathcal{V}$, the property of being alternating becomes

$$\mathbf{v}_{\sigma(1)} \otimes \cdots \otimes \mathbf{v}_{\sigma(k)} = \operatorname{sgn}(\sigma)\,(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k)\,, \quad \forall\, \sigma \in S_k\,. \tag{B.61}$$

We denote as $\Sigma^k(\mathcal{V})$ and $\Lambda^k(\mathcal{V})$ the spaces of the symmetric and the alternating contravariant k-tensors, respectively. They are both vector subspaces of $\mathcal{T}^k(\mathcal{V})$.

Finally, the symmetrization of a contravariant tensor $\mathbf{T} \in \mathcal{T}^k(\mathcal{V})$ is given by the projection onto $\Sigma^k(\mathcal{V})$, that is

$$\begin{aligned} \operatorname{Sym}:\ & \mathcal{T}^k(\mathcal{V}) \to \Sigma^k(\mathcal{V}) \\ & \mathbf{T} \mapsto \operatorname{Sym}\mathbf{T} = \frac{1}{k!}\sum_{\sigma \in S_k} {}^{\sigma}\mathbf{T}\,, \end{aligned} \tag{B.62}$$

as well as the projection onto $\Lambda^k(\mathcal{V})$ provides the alternation of $\mathbf{T}$:

$$\begin{aligned} \operatorname{Alt}:\ & \mathcal{T}^k(\mathcal{V}) \to \Lambda^k(\mathcal{V}) \\ & \mathbf{T} \mapsto \operatorname{Alt}\mathbf{T} = \frac{1}{k!}\sum_{\sigma \in S_k} \operatorname{sgn}(\sigma)\,{}^{\sigma}\mathbf{T}\,. \end{aligned} \tag{B.63}$$

In addition, for a contravariant tensor resulting from the product of k

vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{V}$, the alternating map is provided as

$$\mathrm{Alt}(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma)\, \mathbf{v}_{\sigma(1)} \otimes \cdots \otimes \mathbf{v}_{\sigma(k)}\,. \tag{B.64}$$

Please observe that, recalling (B.45), the term $\mathbf{v}_{\sigma(1)} \otimes \cdots \otimes \mathbf{v}_{\sigma(k)}$ is actually the permutation ${}^{\sigma^{-1}}(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k)$. However, since for any σ appearing in the summation the inverse map σ^{-1} is also included, the expression (B.64) is, in fact, the specification of the general definition (B.63).

B.2.3 The Space of Alternating Tensors

We have seen in the previous section how $\Lambda^k(\mathcal{V}^*)$ and $\Lambda^k(\mathcal{V})$, i.e. the spaces of alternating covariant and contravariant tensors on a vector space $\mathcal{V}$, are subspaces of the tensor spaces $\mathcal{T}^k(\mathcal{V}^*)$ and $\mathcal{T}^k(\mathcal{V})$, respectively.

Moreover, the structure of such subspaces is related with a multilinear map defined on an ordered set of vectors or covectors, which is the exterior product, similarly to how the spaces $\mathcal{T}^k(\mathcal{V}^*)$ and $\mathcal{T}^k(\mathcal{V})$ are related with the tensor product.

For this reason, alternating covariant k-tensors are sometimes called *exterior forms*.

On the other hand, in order to emphasize the vector space structure of $\Lambda^k(\mathcal{V}^*)$, exterior forms are also referred to as k*-covectors*.

Similarly, alternating contravariant k-tensors are called k*-vectors*.

Exterior Product of Vectors and Covectors

Following Dimitrienko (2002), we introduce the exterior product of vectors and covectors by exploiting the alternation maps defined by (B.53) and (B.63).

Definition B.38. Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$ and let $\mathcal{V}^*$ be its dual. We define the *exterior product*, or the *wedge product*, of k vectors as the alternation of their tensor product. The same holds for the product of k covectors.

For the exterior product of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{V}$, we adopt the notation

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k = \mathrm{Alt}(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k) \in \Lambda^k(\mathcal{V}) \,, \tag{B.65}$$

as well as, for the covectors $\mathbf{w}^1, \ldots, \mathbf{w}^k \in \mathcal{V}^*$, we write

$$\mathbf{w}^1 \wedge \cdots \wedge \mathbf{w}^k = \mathrm{Alt}(\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k) \in \Lambda^k(\mathcal{V}^*) \,. \tag{B.66}$$

Recalling that each element of $\Lambda^k(\mathcal{V})$ is an alternating contravariant k-tensor, the exterior product $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$ represents the k-linear functional on $\mathcal{V}^{*k}$ fulfilling the following properties:

- $\mathbf{v}_1 \wedge \cdots \wedge (\mathbf{v}_i + \mathbf{v}_i') \wedge \cdots \wedge \mathbf{v}_k = (\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_i \wedge \cdots \wedge \mathbf{v}_k) + (\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_i' \wedge \cdots \wedge \mathbf{v}_k) \,, \;\; \forall\, \mathbf{v}_i, \mathbf{v}_i' \in \mathcal{V} \,,\; i = 1, \ldots, k \,;$
- $c(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k) = \mathbf{v}_1 \wedge \cdots \wedge (c\mathbf{v}_i) \wedge \cdots \wedge \mathbf{v}_k \,, \;\; \forall\, c \in \mathbb{F} \,,\; i = 1, \ldots, k \,;$
- $\mathbf{v}_{\sigma(1)} \wedge \cdots \wedge \mathbf{v}_{\sigma(k)} = \mathrm{sgn}(\sigma)\,(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k) \,, \;\; \forall\, \sigma \in S_k \,.$

Please notice that the first two properties are inherited from the tensor product (cf. Proposition B.30), while the third one is the specific feature of the alternating forms, as expressed by (B.61).

The straightforward implication of being alternating is the following Lemma.

Lemma B.39. *The exterior product of k vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of the $\mathbb{F}$-vector space $\mathcal{V}$ vanishes if one the entries is repeated:*

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v}_k = \mathbf{0} \,, \tag{B.67}$$

Proof. The statement can be easily verified observing that switching the vector $\mathbf{v}$ implies a change of sign and the resulting tensor equals its opposite:

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v}_k = -(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v}_k) \,. \square$$

Clearly, when k covectors are considered, the above lemma reads

$$\mathbf{w}^1 \wedge \cdots \wedge \mathbf{w}^* \wedge \cdots \wedge \mathbf{w}^* \wedge \cdots \wedge \mathbf{w}^k = \mathbf{0} \,, \tag{B.68}$$

Proposition B.40. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and let $\mathcal{V}^*$ be the dual space. Then, for any pair of sets of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of $\mathcal{V}$ and covectors $\mathbf{w}^1, \ldots, \mathbf{w}^k$ of $\mathcal{V}^*$ the following identity holds true:*

$$(\mathbf{w}^1 \wedge \cdots \wedge \mathbf{w}^k)(\mathbf{v}_1, \ldots, \mathbf{v}_k) = (\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k)(\mathbf{w}^1, \ldots, \mathbf{w}^k)\,. \tag{B.69}$$

Proof. We recall that the exterior product $\mathbf{w}^1 \wedge \cdots \wedge \mathbf{w}^k$ is the alternation of $\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k$. Hence, applying the property (B.54) implies

$$\begin{aligned}(\mathbf{w}^1 \wedge \cdots \wedge \mathbf{w}^k)(\mathbf{v}_1, \ldots, \mathbf{v}_k) &= \mathrm{Alt}(\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k)(\mathbf{v}_1, \ldots, \mathbf{v}_k) \\ &= \frac{1}{k!} \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma)\, (\mathbf{w}^1 \otimes \cdots \otimes \mathbf{w}^k)(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(n)}) \\ &= \frac{1}{k!} \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma)\, \mathbf{w}^1(\mathbf{v}_{\sigma(1)}) \cdots \mathbf{w}^k(\mathbf{v}_{\sigma(k)})\,.\end{aligned}$$

Moreover, considering the map (B.17) defining the action of a vector on a covector, each term can be written as $\mathbf{w}^i(\mathbf{v}_{\sigma(i)}) = \mathbf{v}_{\sigma(i)}(\mathbf{w}^i)$, with $i = 1, \ldots, k$, obtaining

$$\begin{aligned}(\mathbf{w}^1 \wedge \cdots \wedge \mathbf{w}^k)(\mathbf{v}_1, \ldots, \mathbf{v}_k) &= \frac{1}{k!} \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma)\, \mathbf{v}_{\sigma(1)}(\mathbf{w}^1) \cdots \mathbf{v}_{\sigma(k)}(\mathbf{w}^k) \\ &= \frac{1}{k!} \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma)\, (\mathbf{v}_{\sigma(1)} \otimes \cdots \otimes \mathbf{v}_{\sigma(k)})(\mathbf{w}^1, \ldots, \mathbf{w}^k)\,,\end{aligned}$$

where the summation here above, recalling (B.64), provides exactly the alternation of the tensor $\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k$, that is the exterior product $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$, proving the identity (B.69). □

We remark that the exterior product of vectors, defined by (B.65), is introduced as a multilinear map from $\mathcal{V}^k$ to $\Lambda^k(\mathcal{V})$, represented through the symbol '$\wedge$', just as the tensor product maps a set of vectors from $\mathcal{V}^k$ to $\mathcal{T}^k(\mathcal{V})$ by means of the symbol '$\otimes$'. Correspondingly, the exterior product of covectors represents a map from ${\mathcal{V}^*}^k$ to $\Lambda^k(\mathcal{V}^*)$.

Moreover, just like a tensor in $\mathcal{T}^k(\mathcal{V})$ in general cannot be reduced to the tensor product of vectors, analogously not all the k-vectors in $\Lambda^k(\mathcal{V})$, or the k-covectors in $\Lambda^k(\mathcal{V}^*)$, can be expressed in the form of an exterior product.

Specifically, when a k-vector $\mathbf{W} \in \Lambda^k(\mathcal{V})$ results from the wedge product of k elements $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of $\mathcal{V}$, we say that $\mathbf{W}$ is *simple*, or also *decomposable*, that is $\mathbf{W} = \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$. Clearly, the same name applies to k-covectors.

However, the exterior product can be exploited to define a basis for the space $\Lambda^k(\mathcal{V})$, in a similar manner to what has been shown in Proposition B.32 in respect of the tensor product.

Proposition B.41. *Let $\mathcal{V}$ be an n-dimensional vector space over the field $\mathbb{F}$ and let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis. Then, a basis for the space $\Lambda^k(\mathcal{V})$ of the alternating k-tensors is given by*

$$\mathcal{B}^{\wedge k} = \left\{ \mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k} \,\middle|\, 1 \leq i_1 < \ldots < i_k \leq n \right\}. \tag{B.70}$$

Proof. To prove that $\Lambda^k(\mathcal{V})$ is spanned by $\mathcal{B}^{\wedge k}$, consider an arbitrary alternating k-tensor $\mathbf{W} \in \Lambda^k(\mathcal{V})$.

We recall that $\mathbf{W}$ satisfies ${}^{\sigma}\mathbf{W} = \operatorname{sgn}(\sigma)\,\mathbf{W}$ for any permutation $\sigma \in S_k$. Hence, applying the definition (B.63), it is easy to see that $\mathbf{W}$ equals its own alternation:

$$\mathbf{W} = \operatorname{Alt}(\mathbf{W})\,.$$

At the same time, since $\Lambda^k(\mathcal{V})$ is a subspace, the tensor $\mathbf{W}$ also results in $\mathcal{T}^k(\mathcal{V})$ and can be expressed, with respect to the basis $\mathcal{B}^{(k,0)}$ given by (B.36), as

$$\mathbf{W} = W^{i_1 \ldots i_k} \mathbf{b}_{i_1} \otimes \cdots \otimes \mathbf{b}_{i_k}\,,$$

so that, observing that the alternation is clearly linear with respect to the input tensor, one finds

$$\begin{aligned} \mathbf{W} = \operatorname{Alt}(W^{i_1 \ldots i_k} \mathbf{b}_{i_1} \otimes \cdots \otimes \mathbf{b}_{i_k}) &= W^{i_1 \ldots i_k} \operatorname{Alt}(\mathbf{b}_{i_1} \otimes \cdots \otimes \mathbf{b}_{i_k}) \\ &= W^{i_1 \ldots i_k} \mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k}\,, \end{aligned}$$

where Definition B.38 of the wedge product has been used.

We recall that, having used the Einstein summation convention in the above relation, each of the indices $i_1, \ldots, i_k$ in turn assumes all the values from 1 to n while the remaining $k-1$ indices are fixed. Explicitly, the tensor $\mathbf{W}$ can be written as

$$\mathbf{W} = \sum_{i_1,\ldots,i_k=1}^{n} W^{i_1\ldots i_k} \mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k} \, .$$

Recalling the property (B.67), in the equation here above the terms with repeating indices do vanish. Moreover, the remaining addends can be gathered taking the indices in ascending order, which ensures that only the non-null contributions are actually considered, and applying all their possible permutations:

$$\mathbf{W} = \sum_{\substack{1 \le i_1 < \ldots < i_k \le n \\ \sigma \in S_k}} W^{\sigma(i_1)\ldots\sigma(i_k)} \mathbf{b}_{\sigma(i_1)} \wedge \cdots \wedge \mathbf{b}_{\sigma(i_k)} \, , \tag{B.71}$$

The resulting relation may look cumbersome, but it simplifies if one takes into account that the following identity holds true for each addend:

$$W^{\sigma(i_1)\ldots\sigma(i_k)} \mathbf{b}_{\sigma(i_1)} \wedge \cdots \wedge \mathbf{b}_{\sigma(i_k)} = W^{i_1\ldots i_k} \mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k} \, , \quad \text{(no summation)}$$

which comes from the condition $\mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k} = \operatorname{sgn}(\sigma) \mathbf{b}_{\sigma(i_1)} \wedge \cdots \wedge \mathbf{b}_{\sigma(i_k)}$, along with the property (B.60) for the coordinates of alternating tensors. Noting also that the total number of distinct permutations in S_k is $k!$, the equation (B.71) finally becomes:

$$\mathbf{W} = k! \sum_{1 \le i_1 < \ldots < i_k \le n} W^{i_1\ldots i_k} \mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k} \, ,$$

which proves that the tensor $\mathbf{W}$ of $\Lambda^k(\mathcal{V})$ is in the span of the set $\mathcal{B}^{\wedge k}$.

Now we prove the linear independence of the elements collected in $\mathcal{B}^{\wedge k}$ by assuming that the above linear combination equals the null tensor $\mathbf{0}$ of $\Lambda^k(\mathcal{V})$:

$$k! \sum_{1 \le i_1 < \ldots < i_k \le n} W^{i_1\ldots i_k} \mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k} = \mathbf{0} \, .$$

Applying once again Definition B.38 of the exterior product, and using the linearity of the alternation provided by (B.64), one has

$$
\begin{aligned}
k! \sum_{1 \le i_1 < \ldots < i_k \le n} W^{i_1 \ldots i_k} \mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k} &= k! \sum_{1 \le i_1 < \ldots < i_k \le n} W^{i_1 \ldots i_k} \operatorname{Alt}(\mathbf{b}_{i_1} \otimes \cdots \otimes \mathbf{b}_{i_k}) \\
&= k! \sum_{1 \le i_1 < \ldots < i_k \le n} W^{i_1 \ldots i_k} \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) \left(\mathbf{b}_{\sigma(i_1)} \otimes \cdots \otimes \mathbf{b}_{\sigma(i_k)}\right) \\
&= \sum_{\substack{1 \le i_1 < \ldots < i_k \le n \\ \sigma \in S_k}} \left(\operatorname{sgn}(\sigma) W^{i_1 \ldots i_k}\right) \left(\mathbf{b}_{\sigma(i_1)} \otimes \cdots \otimes \mathbf{b}_{\sigma(i_k)}\right) = \mathbf{0}
\end{aligned}
$$

Please observe that the indices $i_1, \ldots, i_k$ differ from each other, so that the contravariant tensors $\mathbf{b}_{\sigma(i_1)} \otimes \cdots \otimes \mathbf{b}_{\sigma(i_k)}$ are distinct themselves. Moreover, since such tensors are all included in the basis $\mathcal{B}^{(k,0)}$ of $\mathcal{T}^k(\mathcal{V})$, as expressed by (B.36), they are certainly linearly independent.

Then, the null tensor $\mathbf{0}$ can be obtained only by vanishing the coefficients appearing in the above linear combination, that is

$$
\operatorname{sgn}(\sigma) W^{i_1 \ldots i_k} = 0\,, \quad 1 \le i_1 < \ldots < i_k \le n\,, \ \forall \sigma \in S_k\,,
$$

which implies the condition $W^{i_1 \ldots i_k} = 0$ for all the indices $i_1 < \ldots < i_k$, and hence the linear independence of the tensors $\mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k}$. □

Evidently, a basis for the space $\Lambda^k(\mathcal{V}^*)$ of the k-covectors on $\mathcal{V}$ is

$$
\mathcal{B}^{\wedge k *} = \left\{ \mathbf{b}^{i_1} \wedge \cdots \wedge \mathbf{b}^{i_k} \;\middle|\; 1 \le i_1 < \ldots < i_k \le n \right\}, \tag{B.72}
$$

where $\mathbf{b}^{i_j}$ is an element of the dual basis $\mathcal{B}^* = \{\mathbf{b}^1, \ldots, \mathbf{b}^n\}$.

Furthermore, since in constructing the basis $\mathcal{B}^{\wedge k}$ the indices $i_1, \ldots, i_k$ must be distinct, each tensor in the form $\mathbf{b}_{i_1} \wedge \cdots \wedge \mathbf{b}_{i_k}$ is given by a combination of n elements, i.e. the vectors $\mathbf{b}_i$ of the basis $\mathcal{B}$, taken k at a time without repetition. The total number of such combinations provides the number of elements of the basis $\mathcal{B}^{\wedge k}$.

The same is true for the number of tensors in $\mathcal{B}^{\wedge k *}$, so that the dimension of the spaces $\Lambda^k(\mathcal{V})$ and $\Lambda^k(\mathcal{V}^*)$ is

$$
\dim \Lambda^k(\mathcal{V}) = \dim \Lambda^k(\mathcal{V}^*) = \binom{n}{k} = \frac{n!}{k!(n-k)!}\,. \tag{B.73}
$$

Linear Dependence of Vectors

A consequence of Proposition B.41 is that the space of alternating tensors on an n-dimensional vector space $\mathcal{V}$ is well-defined only for $k \leq n$.

Actually, when k is greater than n, the binomial coefficient $\binom{n}{k}$ in (B.73) is zero by convention. In this case, the dimension of both $\Lambda^k(\mathcal{V})$ and $\Lambda^k(\mathcal{V}^*)$ vanishes, and such vector spaces coincide with trivial ones $\{\mathbf{0}\}$ and $\{\mathbf{0}^*\}$, respectively.

In addition, the exterior product of k vectors or k covectors, providing an alternating tensor in $\Lambda^k(\mathcal{V})$ and $\Lambda^k(\mathcal{V}^*)$, respectively, vanishes for $k > n$:

$$\mathbf{v}_1 \wedge \cdots \mathbf{v}_n \wedge \mathbf{v}_{n+1} \wedge \cdots \wedge \mathbf{v}_k = \mathbf{0}\,, \quad \forall\, \mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{V}\,,\ k > n\,, \tag{B.74}$$

and also

$$\mathbf{w}^1 \wedge \cdots \mathbf{w}^n \wedge \mathbf{w}^{n+1} \wedge \cdots \wedge \mathbf{w}^k = \mathbf{0}^*\,, \quad \forall\, \mathbf{w}^1, \ldots, \mathbf{w}^k \in \mathcal{V}^*\,,\ k > n\,. \tag{B.75}$$

Let us observe that the result in (B.74), as well as the one in (B.75), can also be seen as the specification of a more general property, which associates the vanishing of the exterior product with the linear dependence of the involved vectors.

From this perspective, the wedge products in (B.74) and (B.75) do vanish since any collection of vectors gathering more than n elements of an n-dimensional vector space is clearly linearly dependent.

The connection of the exterior product with the linear dependence of vectors is formalized in the following proposition.

Proposition B.42. *The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of the $\mathbb{F}$-vector space $\mathcal{V}$ are linearly dependent if, and only if, their exterior product is null:*

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k = \mathbf{0}\,.$$

Proof. If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly dependent, there is a vector $\mathbf{v}_i$ among them which can be expressed as a linear combination of the other ones, that is:

$$\mathbf{v}_i = \sum_{\substack{j=1 \\ j\neq i}}^{k} a^j \mathbf{v}_j \,.$$

Then, the multilinearity of the exterior product implies

$$\begin{aligned}\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_i \wedge \cdots \wedge \mathbf{v}_k &= \mathbf{v}_1 \wedge \cdots \wedge \sum_{\substack{j=1 \\ j\neq i}}^{k} a^j \mathbf{v}_j \wedge \cdots \wedge \mathbf{v}_k \\ &= \sum_{\substack{j=1 \\ j\neq i}}^{k} a^j \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_j \wedge \cdots \wedge \mathbf{v}_k = \mathbf{0}\,,\end{aligned}$$

where the j-th addend, having the vector $\mathbf{v}_j$ repeated, do vanishes (cf. Lemma B.39), and hence the summation is itself null.

Let us now show that the vanishing of the exterior product implies the linear dependence of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

If k is greater than the dimension n of the vector space, the result is trivial, since the maximum number of linearly independent vectors in $\mathcal{V}$ is exactly n.

Suppose $k \leq n$ and proceed by contradiction assuming that the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are independent.

We can collect such vectors along with additional $n-k$ elements of $\mathcal{V}$ to obtain a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$. Then, by Proposition B.41, a basis for $\Lambda^k(\mathcal{V})$ is given by

$$\left\{ \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_k} \;\middle|\; 1 \leq i_1 < \ldots < i_k \leq n \right\}.$$

Then, the exterior product $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$ is an element of this basis and cannot be null.

Such a result contradicts the hypothesis $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k = \mathbf{0}$, concluding that the k vectors are actually dependent. □

Determinant of an Endomorphism

Let us focus on the alternating tensors with rank $k = n$.

By equation (B.73), the space $\Lambda^n(\mathcal{V})$ is one-dimensional and then it is isomorphic with the field $\mathbb{F}$ itself.

Consequently, the non-null elements in $\Lambda^n(\mathcal{V})$ are proportional to each other and any of them can be used as a basis tensor.

As an example, by virtue of Proposition B.41, a basis for $\Lambda^n(\mathcal{V})$ is given by $\mathcal{B}^{\wedge n} = \{\mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_n\}$, so that an alternating tensor $\omega \in \Lambda^n(\mathcal{V})$ can be identified by a scalar $c \in \mathbb{F}$, i.e.

$$\omega = c\, \mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_n \,.$$

This representation shows that any n-vector is decomposable in the form of the exterior product of n vectors, but the decomposition is not unique. In fact, for the linearity of the wedge product, one can equally express

$$\omega = (c\mathbf{b}_1) \wedge \cdots \wedge \mathbf{b}_n = \mathbf{b}_1 \wedge \cdots (c\mathbf{b}_i) \cdots \wedge \mathbf{b}_n = \mathbf{b}_1 \wedge \cdots \wedge (c\mathbf{b}_n) \,.$$

At the same time, any set of linearly independent vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ can serve as a basis for $\mathcal{V}$, obtaining a new basis tensor $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n$ for $\Lambda^n(\mathcal{V})$ and so a further possible representation of the multivector ω.

We also remark that the isomorphism of $\Lambda^n(\mathcal{V})$ with $\mathbb{F}$ is not canonical, meaning that it depends on the basis tensor, and hence on the choice of a basis for $\mathcal{V}$.

Let us now consider a linear map in $\mathsf{End}(\mathcal{V})$. Any linear transformation $\boldsymbol{f}\colon \mathcal{V} \to \mathcal{V}$ induces a map $\Lambda^n_{\boldsymbol{f}}\colon \Lambda^n(\mathcal{V}) \to \Lambda^n(\mathcal{V})$ defined as follows:

$$\Lambda^n_{\boldsymbol{f}}(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n) = \boldsymbol{f}(\mathbf{v}_1) \wedge \cdots \wedge \boldsymbol{f}(\mathbf{v}_n) \,. \tag{B.76}$$

Please notice that the linearity of the wedge product, along with the one of the map $\boldsymbol{f}$, makes $\Lambda^n_{\boldsymbol{f}}$ a linear transformation, that is $\Lambda^n_{\boldsymbol{f}} \in \mathsf{End}\, \Lambda^n(\mathcal{V})$.

Moreover, recalling that all the tensors in $\Lambda^n(\mathcal{V})$ are proportional to each other, the image $\Lambda^n_{\boldsymbol{f}}(\omega)$ of a multivector ω results from ω itself by a scalar multiplication.

The scalar relating $\Lambda^n_{\boldsymbol{f}}(\omega)$ with the preimage ω is the determinant of the linear map $\boldsymbol{f}$ (see, e.g. Winitzki (2020)).

Definition B.43. Let $\mathcal{V}$ be an $\mathbb{F}$-vector space and $\boldsymbol{f} \in \mathsf{End}\,\mathcal{V}$ a linear map. The *determinant* of $\boldsymbol{f}$ is the scalar $\mathsf{det}(\boldsymbol{f}) \in \mathbb{F}$ such that

$$\Lambda^n_{\boldsymbol{f}}(\omega) = \mathsf{det}(\boldsymbol{f})\,\omega\,, \tag{B.77}$$

where $\omega \in \Lambda^n(\mathcal{V})$ is an arbitrary non-null n-vector.

Equivalently, by the property (B.76) of the map $\Lambda^n_{\boldsymbol{f}}$, the determinant of $\boldsymbol{f}$ also satisfies

$$\boldsymbol{f}(\mathbf{v}_1) \wedge \cdots \wedge \boldsymbol{f}(\mathbf{v}_n) = \mathsf{det}(\boldsymbol{f})\,\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n\,, \tag{B.78}$$

which holds true for any set of linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathcal{V}$.

As could be expected, the number $\mathsf{det}(\boldsymbol{f})$ satisfying (B.77) does not depends on the multivector ω. In fact, since any tensor in $\Lambda^n(\mathcal{V})$ can be expressed as $\omega' = c\omega$, for some $c \in \mathbb{F}$, the linearity of $\Lambda^n_{\boldsymbol{f}}$ implies

$$\Lambda^n_{\boldsymbol{f}}(\omega') = \Lambda^n_{\boldsymbol{f}}(c\omega) = c\Lambda^n_{\boldsymbol{f}}(\omega) = c\,\mathsf{det}(\boldsymbol{f})\omega = \mathsf{det}(\boldsymbol{f})\omega'\,,\ \ \forall\,\omega' \in \Lambda^n(\mathcal{V})\,.$$

On the other side, the determinant of $\boldsymbol{f}$ clearly depends on the transformation $\boldsymbol{f}$. In particular, the value of $\mathsf{det}(\boldsymbol{f})$ characterizes $\boldsymbol{f}$ as an isomorphism.

Proposition B.44. *Let $\boldsymbol{f}$ be an endomorphism of the $\mathbb{F}$-vector space $\mathcal{V}$. Then, the determinant of $\boldsymbol{f}$ is non-null if, and only if, the map $\boldsymbol{f}$ is bijective.*

Proof. Consider an arbitrary set of n linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathcal{V}$, so that, by Proposition B.42, the exterior product $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n$ is non-null.

Consequently, using the property (B.78), the product $\boldsymbol{f}(\mathbf{v}_1) \wedge \cdots \wedge \boldsymbol{f}(\mathbf{v}_n)$ vanishes if, and only if, $\mathsf{det}(\boldsymbol{f})$ is the zero scalar of $\mathbb{F}$.

At the same time, recalling Corollary B.15, the map $\boldsymbol{f}$ is bijective if, and only if, the images $\boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_n)$ are linearly independent.

Hence, if $\mathsf{det}(\boldsymbol{f}) \neq 0$, the product $\boldsymbol{f}(\mathbf{v}_1) \wedge \cdots \wedge \boldsymbol{f}(\mathbf{v}_n)$ is other than the null tensor of $\Lambda^n(\mathcal{V})$, the vectors $\boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_n)$ are linearly independent and the endomorphism $\boldsymbol{f}$ is bijective.

Conversely, when $\boldsymbol{f}$ is bijective, the images $\boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_n)$ are linearly independent, the product $\boldsymbol{f}(\mathbf{v}_1) \wedge \cdots \wedge \boldsymbol{f}(\mathbf{v}_n)$ is non-null, and so is the determinant of $\boldsymbol{f}$. □

Proposition B.45. *Let $\mathcal{V}$ be an n-dimensional vector space over the field $\mathbb{F}$. Then, the following properties hold true.*

1. *The determinant of the identity map* $\mathrm{id}_{\mathcal{V}}$ *is the multiplicative identity of* $\mathbb{F}$*:*

$$\det(\mathrm{id}_{\mathcal{V}}) = 1\,. \tag{B.79}$$

2. *The determinant of the product of endomorphisms equals the multiplication of the determinants of the maps:*

$$\det(\boldsymbol{f}\boldsymbol{g}) = \det(\boldsymbol{f}) \det(\boldsymbol{g})\,, \quad \forall\, \boldsymbol{f}, \boldsymbol{g} \in \operatorname{End} \mathcal{V}\,. \tag{B.80}$$

3. *The determinant of the inverse map of an automorphism is the inverse of the determinant of the map:*

$$\det(\boldsymbol{f}^{-1}) = \det(\boldsymbol{f})^{-1}\,, \quad \forall\, \boldsymbol{f} \in \operatorname{Aut} \mathcal{V}\,. \tag{B.81}$$

4. *The determinant of the scalar multiplication of an endomorphism by a number is the determinant of the map multiplied by the n-th power of the scalar:*

$$\det(c\boldsymbol{f}) = c^n \det(\boldsymbol{f})\,, \quad \forall\, \boldsymbol{f} \in \operatorname{End} \mathcal{V}\,,\ c \in \mathbb{F}\,. \tag{B.82}$$

Proof. 1. Using the property (B.78) with $\boldsymbol{f} = \mathrm{id}_{\mathcal{V}}$ reads

$$\mathrm{id}_{\mathcal{V}}(\mathbf{v}_1) \wedge \cdots \wedge \mathrm{id}_{\mathcal{V}}(\mathbf{v}_n) = \det(\mathrm{id}_{\mathcal{V}})\, \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n\,.$$

Moreover, being $\mathrm{id}_{\mathcal{V}}(\mathbf{v}) = \mathbf{v}$ for all the vectors in $\mathcal{V}$, one also has

$$\mathrm{id}_{\mathcal{V}}(\mathbf{v}_1) \wedge \cdots \wedge \mathrm{id}_{\mathcal{V}}(\mathbf{v}_n) = \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n\,,$$

whence, by comparison, the property $\det(\mathrm{id}_{\mathcal{V}}) = 1$ is verified.

2. Since $(\boldsymbol{f}\boldsymbol{g})(\mathbf{v}) = \boldsymbol{f}\big(\boldsymbol{g}(\mathbf{v})\big)$ for any $\mathbf{v} \in \mathcal{V}$, one can verify

$$\begin{aligned}(\boldsymbol{f}\boldsymbol{g})(\mathbf{v}_1) \wedge \cdots \wedge (\boldsymbol{f}\boldsymbol{g})(\mathbf{v}_n) &= \boldsymbol{f}\big(\boldsymbol{g}(\mathbf{v}_1)\big) \wedge \cdots \wedge \boldsymbol{f}\big(\boldsymbol{g}(\mathbf{v}_n)\big) \\ &= \det(\boldsymbol{f})\, \boldsymbol{g}(\mathbf{v}_1) \wedge \cdots \wedge \boldsymbol{g}(\mathbf{v}_n) \\ &= \det(\boldsymbol{f}) \det(\boldsymbol{g})\, \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n\,,\end{aligned}$$

where the property (B.78) has been first applied to the map $\boldsymbol{f}$ and the vectors $\boldsymbol{g}(\mathbf{v}_1), \ldots, \boldsymbol{g}(\mathbf{v}_n)$, and then to $\boldsymbol{g}$ and the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. At the same time, the determinant of $\boldsymbol{f}\boldsymbol{g}$ is such that

$$(\boldsymbol{f}\boldsymbol{g})(\mathbf{v}_1) \wedge \cdots \wedge (\boldsymbol{f}\boldsymbol{g})(\mathbf{v}_n) = \det(\boldsymbol{f}\boldsymbol{g})\, \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n\,,$$

which, by comparing with the above relation, implies

$$\det(\boldsymbol{f}\boldsymbol{g}) = \det(\boldsymbol{f}) \det(\boldsymbol{g})\,.$$

3. We recall that the inverse $\boldsymbol{f}^{-1}$ of an automorphism $\boldsymbol{f}$ is such that

$$\boldsymbol{f}\boldsymbol{f}^{-1} = \mathrm{id}_{\mathcal{V}}\,,$$

whence, by applying parts 2. and 1. of the proposition to the two sides, respectively, one finds

$$\det(\boldsymbol{f}) \det(\boldsymbol{f}^{-1}) = 1\,,$$

and then $\det(\boldsymbol{f}^{-1}) = \det(\boldsymbol{f})^{-1}$.

4. Specializing the property (B.78) to the map $c\boldsymbol{f}$, one has

$$(c\boldsymbol{f})(\mathbf{v}_1) \wedge \cdots \wedge (c\boldsymbol{f})(\mathbf{v}_n) = \det(c\boldsymbol{f})\, \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n\,.$$

At the same time, because of the linearity of the exterior product, as well as of the map $\boldsymbol{f}$, one also finds

$$\begin{aligned}(c\boldsymbol{f})(\mathbf{v}_1) \wedge \cdots \wedge (c\boldsymbol{f})(\mathbf{v}_n) &= c^n \boldsymbol{f}(\mathbf{v}_1) \wedge \cdots \wedge \boldsymbol{f}(\mathbf{v}_n) \\ &= c^n \det(\boldsymbol{f})\, \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n\,,\end{aligned}$$

so that the comparison with the previous expression proves the identity

$$\det(c\boldsymbol{f}) = c^n \det(\boldsymbol{f}). \qquad \square$$

An additional feature of the determinant relates an endomorphism with its transpose.

Proposition B.46. *Let $\boldsymbol{f} : \mathcal{V} \to \mathcal{V}$ be an endomorphism of the $\mathbb{F}$-vector space $\mathcal{V}$ and let $\boldsymbol{f}^\top : \mathcal{V}^* \to \mathcal{V}^*$ be the transpose map. Then, the determinant of $\boldsymbol{f}^\top$ is equal to the one of $\boldsymbol{f}$:*

$$\det(\boldsymbol{f}^\top) = \det(\boldsymbol{f}) . \tag{B.83}$$

Proof. Consider an arbitrary set of linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathcal{V}$ and covectors $\mathbf{w}^1, \ldots, \mathbf{w}^n$ of the dual space $\mathcal{V}^*$

The exterior product of the images $\boldsymbol{f}^\top(\mathbf{w}^1), \ldots, \boldsymbol{f}^\top(\mathbf{w}^n)$ provides an alternating n-covector of $\Lambda^n(\mathcal{V}^*)$ to be applied to the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Subsequently, recalling the property (B.54) of the alternation map, one has

$$\begin{aligned}
&\big(\boldsymbol{f}^\top(\mathbf{w}^1) \wedge \cdots \wedge \boldsymbol{f}^\top(\mathbf{w}^n)\big)(\mathbf{v}_1, \ldots, \mathbf{v}_n) \\
&\qquad = \frac{1}{n!} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \left(\boldsymbol{f}^\top(\mathbf{w}^1) \otimes \cdots \otimes \boldsymbol{f}^\top(\mathbf{w}^n)\right)(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(n)}) \\
&\qquad = \frac{1}{n!} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \left(\boldsymbol{f}^\top(\mathbf{w}^1)\right)(\mathbf{v}_{\sigma(1)}) \cdots \left(\boldsymbol{f}^\top(\mathbf{w}^n)\right)(\mathbf{v}_{\sigma(n)}) .
\end{aligned}$$

Moreover, by means of (B.12), the i-th term of each addend satisfies the following identity

$$\big(\boldsymbol{f}^\top(\mathbf{w}^i)\big)(\mathbf{v}_{\sigma(i)}) = \big(\boldsymbol{f}(\mathbf{v}_{\sigma(i)})\big)(\mathbf{w}^i) ,$$

so that one also can write

$$\begin{aligned}
&\big(\boldsymbol{f}^\top(\mathbf{w}^1) \wedge \cdots \wedge \boldsymbol{f}^\top(\mathbf{w}^n)\big)(\mathbf{v}_1, \ldots, \mathbf{v}_n) \\
&\qquad = \frac{1}{n!} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \left(\boldsymbol{f}(\mathbf{v}_{\sigma(1)})\right)(\mathbf{w}^1) \cdots \left(\boldsymbol{f}(\mathbf{v}_{\sigma(n)})\right)(\mathbf{w}^n) \\
&\qquad = \frac{1}{n!} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \left(\boldsymbol{f}(\mathbf{v}_{\sigma(1)}) \otimes \cdots \otimes \boldsymbol{f}(\mathbf{v}_{\sigma(n)})\right)(\mathbf{w}^1, \ldots, \mathbf{w}^n) \\
&\qquad = \big(\boldsymbol{f}(\mathbf{v}_1) \wedge \cdots \wedge \boldsymbol{f}(\mathbf{v}_n)\big)(\mathbf{w}^1, \ldots, \mathbf{w}^n) ,
\end{aligned}$$

where, recalling the map (B.64), the summation provides the alternation of the tensor $\boldsymbol{f}(\mathbf{v}_1) \otimes \cdots \otimes \boldsymbol{f}(\mathbf{v}_n)$, and so the exterior product of $\boldsymbol{f}(\mathbf{v}_1), \ldots, \boldsymbol{f}(\mathbf{v}_n)$.

In summary, the following identity holds:

$$\big(\boldsymbol{f}^{\mathsf{T}}(\mathbf{w}^1) \wedge \cdots \wedge \boldsymbol{f}^{\mathsf{T}}(\mathbf{w}^n)\big)(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \big(\boldsymbol{f}(\mathbf{v}_1) \wedge \cdots \wedge \boldsymbol{f}(\mathbf{v}_n)\big)(\mathbf{w}^1, \ldots, \mathbf{w}^n)\,,$$

whence, by applying the property (B.78) to both sides, one also has

$$\det(\boldsymbol{f}^{\mathsf{T}})\,(\mathbf{w}^1 \wedge \cdots \wedge \mathbf{w}^n)(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \det(\boldsymbol{f})\,(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n)(\mathbf{w}^1, \ldots, \mathbf{w}^n)\,.$$

Hence, exploiting Proposition B.40 and for the arbitrariness of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathcal{V}$ and the covectors $\mathbf{w}^1, \ldots, \mathbf{w}^n \in \mathcal{V}^*$, one finally finds

$$\det(\boldsymbol{f}^{\mathsf{T}}) = \det(\boldsymbol{f})\,.$$ □

B.3 Scalar Product Spaces

Definition B.47. Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. A *scalar product* on $\mathcal{V}$ is a symmetric bilinear map

$$\langle \cdot, \cdot \rangle_{\mathcal{V}} \colon \mathcal{V} \times \mathcal{V} \to \mathbb{F}\,.$$

Explicitly, the scalar product satisfies the following properties:

- $\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{V}} = \langle \mathbf{v}, \mathbf{u} \rangle_{\mathcal{V}}\,,\ \ \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{V}$;
- $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle_{\mathcal{V}} = \langle \mathbf{u}, \mathbf{w} \rangle_{\mathcal{V}} + \langle \mathbf{v}, \mathbf{w} \rangle_{\mathcal{V}}\,,\ \ \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{V}$;
- $\langle a\mathbf{u}, \mathbf{v} \rangle_{\mathcal{V}} = a\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{V}}\,,\ \ \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{V}\,,\ a \in \mathbb{F}$.

By virtue of the symmetry, it is clear that the linearity of $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ also applies to the second argument.

Moreover, when the vector space is uniquely specified, the subscript can be omitted and the notation $\langle \cdot, \cdot \rangle$ is adopted.

The standard scalar product for the vector space $\mathbb{F}^n$, also called the *dot product*, is defined by directly operating on the entries of the involved

n-tuples:

$$\begin{aligned} \langle \cdot, \cdot \rangle_{\mathbb{F}^n} : \; & \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F} \\ & (\bar{\mathbf{u}}, \bar{\mathbf{v}}) \mapsto \bar{\mathbf{u}} \cdot \bar{\mathbf{v}} = u^1 v^1 + \ldots + u^n v^n \,. \end{aligned} \tag{B.84}$$

Since the elements of the field $\mathbb{F}$ do commute with respect to both addition and multiplication, one can easily verify that the dot product is actually symmetric and bilinear, satisfying the requirements of Definition B.47.

B.3.1 Orthogonality

Definition B.48. Two elements $\mathbf{u}$ and $\mathbf{v}$ of vector space $\mathcal{V}$, endowed with a scalar product, are *orthogonal* or *perpendicular* if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$:

$$\mathbf{u} \perp \mathbf{v} \quad \Leftrightarrow \quad \langle \mathbf{u}, \mathbf{v} \rangle = 0 \,.$$

Please notice that the null vector $\mathbf{o} \in \mathcal{V}$ is orthogonal to any other vector $\mathbf{v}$:

$$\langle \mathbf{v}, \mathbf{o} \rangle = 0 \,, \;\; \forall \, \mathbf{v} \in \mathcal{V} \,.$$

Such a result comes from the property $\mathbf{o} = 0\mathbf{v}$ (cf. Proposition A.26) and the third property of Definition B.47, that is

$$\langle \mathbf{v}, 0\mathbf{v} \rangle = 0 \langle \mathbf{v}, \mathbf{v} \rangle = 0 \,, \;\; \forall \, \mathbf{v} \in \mathcal{V} \,.$$

Specifically, the scalar product of the vector space $\mathcal{V}$ is *non-degenerate* if the null vector is the only element of $\mathcal{V}$ to be orthogonal to any other vector:

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0 \,, \;\; \forall \, \mathbf{v} \in \mathcal{V} \quad \Leftrightarrow \quad \mathbf{u} = \mathbf{o} \,. \tag{B.85}$$

Finally, considering a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of the vector space $\mathcal{V}$, it is called an *orthogonal basis* if all pairs of distinct vectors of $\mathcal{B}$ are orthogonal:

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0 \,, \;\; \forall \, i \neq j \,. \tag{B.86}$$

B.3.2 Duality in Scalar Product Spaces

Let us consider a vector space $\mathcal{V}$ with a non-degenerate scalar product. Since the scalar product is a bilinear form, fixing one of the two entries induces a linear map with respect to the other one. Such a map acts as a linear form on $\mathcal{V}$ and then is an element of a dual space $\mathcal{V}^*$.

Explicitly, let us introduce the following map:

$$\begin{aligned} \psi :\ & \mathcal{V} \to \mathcal{V}^* \\ & \mathbf{u} \mapsto \mathbf{u}^* = \psi(\mathbf{u})\,, \end{aligned} \tag{B.87}$$

where the form $\mathbf{u}^* : \mathcal{V} \to \mathbb{F}$ satisfies

$$\mathbf{u}^*(\mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle\,, \ \ \forall\, \mathbf{v} \in \mathcal{V}\,. \tag{B.88}$$

It is easy to verify that the map ψ defined by (B.87) is a vector space isomorphism.

Actually, the linearity of ψ readily comes from the bilinearity of the scalar product in (B.88), that is $(\mathbf{u} + \mathbf{u}')^* = \mathbf{u}^* + \mathbf{u}'^*$, as well as $(c\mathbf{u})^* = c\mathbf{u}^*$.

Moreover, since the scalar product in (B.88) is non-degenerate, by (B.85) the image $\mathbf{u}^*$ is the null functional $\mathbf{o}^* \in \mathcal{V}^*$ only if $\mathbf{u} = \mathbf{o}$, which means $\mathrm{Ker}(\psi) = \{\mathbf{o}\}$. Then, recalling that $\mathcal{V}$ and $\mathcal{V}^*$ have the same dimension, by Corollary B.14 the linear transformation ψ is bijective.

The inverse of ψ is the linear map defined as follows:

$$\begin{aligned} \psi^{-1} :\ & \mathcal{V}^* \to \mathcal{V} \\ & \mathbf{u}^* \mapsto \mathbf{u} = \psi^{-1}(\mathbf{u}^*)\,, \end{aligned} \tag{B.89}$$

where, consistently with (B.88), the image $\mathbf{u}$ is the only element of $\mathcal{V}$ whose scalar product with any vector $\mathbf{v}$ provides the same value as applying the functional $\mathbf{u}^*$.

It is worth noting that, by virtue of (B.88), the relation between the vector space $\mathcal{V}$ and the dual space $\mathcal{V}^*$ does not depend on the choice of a basis.

Then, any vector space $\mathcal{V}$ endowed with a non-degenerate scalar product is canonically isomorphic with its dual space, allowing one to identify the dual space $\mathcal{V}^*$ with $\mathcal{V}$ itself:

$$\mathcal{V}^* \cong \mathcal{V}\,. \tag{B.90}$$

In this sense, $\mathcal{V}$ is said *self-dual* (see, e.g., Dimitrienko (2002)).

Transpose in Scalar Product Spaces

Let us now consider two vector spaces $\mathcal{V}$ and $\mathcal{W}$ over the same field $\mathbb{F}$, each with a non-degenerate scalar product, and let $\boldsymbol{f}\colon \mathcal{V} \to \mathcal{W}$ be a linear map.

Then, since both $\mathcal{V}$ and $\mathcal{W}$ are self-dual, the transpose of $\boldsymbol{f}$ is the following linear map:

$$\begin{aligned} \boldsymbol{f}^{\mathsf{T}} : \mathcal{W} &\to \mathcal{V} \\ \mathbf{w} &\mapsto \boldsymbol{f}^{\mathsf{T}}(\mathbf{w})\,, \end{aligned} \tag{B.91}$$

where the image $\boldsymbol{f}^{\mathsf{T}}(\mathbf{w})$ satisfies

$$\langle \boldsymbol{f}^{\mathsf{T}}(\mathbf{w}), \mathbf{v}\rangle_{\mathcal{V}} = \langle \mathbf{w}, \boldsymbol{f}(\mathbf{v})\rangle_{\mathcal{W}}\,, \;\; \forall\, \mathbf{v} \in \mathcal{V}\,,\; \mathbf{w} \in \mathcal{W}\,. \tag{B.92}$$

Please notice that (B.91) readily comes from Definition B.26 with the identifications $\mathcal{V}^* \cong \mathcal{V}$ and $\mathcal{W}^* \cong \mathcal{W}$. In addition, since the vector $\boldsymbol{f}^{\mathsf{T}}(\mathbf{w})$ is a functional acting on $\mathcal{V}$, just as $\boldsymbol{f}(\mathbf{v})$ is a form on $\mathcal{W}$, the identity (B.92) is a consequence of the property (B.88).

B.3.3 Real Vector Spaces

Let $\mathcal{V}$ a vector space defined over the real field $\mathbb{R}$ and supposed it is endowed with a scalar product. Since the set of real numbers is an ordered field, it makes sense to introduce an additional requirement for the scalar product.

Definition B.49. A scalar product on an $\mathbb{R}$-vector space $\mathcal{V}$ is said *positive definite* if the following conditions are satisfied:

- $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0\,, \;\; \forall\, \mathbf{v} \in \mathcal{V}(\mathbb{R})$;
- $\langle \mathbf{v}, \mathbf{v}\rangle = 0 \;\;\Leftrightarrow\;\; \mathbf{v} = \mathbf{o}$.

One can easily observe that, since the second property here introduced is consistent with the condition (B.85), a positive scalar product is also non-degenerate.

Moreover, a consequence of the positiveness is the so-called *Cauchy-Schwartz' inequality*:

$$\langle \mathbf{v}, \mathbf{w} \rangle^2 \leq \langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle\,, \quad \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{V}\,. \tag{B.93}$$

The validity of Cauchy-Schwartz' inequality is trivial when $\mathbf{v} = \mathbf{o}$. Conversely, if $\mathbf{v} \neq \mathbf{o}$ one can set $c = \langle \mathbf{v}, \mathbf{w} \rangle / \langle \mathbf{v}, \mathbf{v} \rangle$, so that, by the positiveness of the scalar product and recalling the symmetry of the scalar product, one finds

$$\begin{aligned} 0 \leq \langle \mathbf{w} - c\mathbf{v}, \mathbf{w} - c\mathbf{v} \rangle &= \langle \mathbf{w}, \mathbf{w} \rangle - 2c \langle \mathbf{v}, \mathbf{w} \rangle + c^2 \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \langle \mathbf{w}, \mathbf{w} \rangle - 2 \frac{\langle \mathbf{v}, \mathbf{w} \rangle^2}{\langle \mathbf{v}, \mathbf{v} \rangle} + \frac{\langle \mathbf{v}, \mathbf{w} \rangle^2}{\langle \mathbf{v}, \mathbf{v} \rangle^2} \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \langle \mathbf{w}, \mathbf{w} \rangle - \frac{\langle \mathbf{v}, \mathbf{w} \rangle^2}{\langle \mathbf{v}, \mathbf{v} \rangle}\,, \end{aligned}$$

which implies the inequality (B.93).

When a vector space $\mathcal{V}$ is defined over the real field, or a general ordered field, the norm of a vector is also introduced (see, e.g., Lang (2002), Lee (2012)).

Definition B.50. Let $\mathcal{V}$ be a vector space over the field $\mathbb{R}$. A *norm* on $\mathcal{V}$ is a real-valued function

$$\|\cdot\| : \mathcal{V} \to \mathbb{R}\,, \tag{B.94}$$

such that the following properties are satisfied:

- positivity:

$$\|\mathbf{v}\| \geq 0\,, \quad \forall\, \mathbf{v} \in \mathcal{V}\,, \quad \text{with } \|\mathbf{v}\| = 0 \text{ if, and only if, } \mathbf{v} = \mathbf{o}\,;$$

- homogeneity:

$$\|c\mathbf{v}\| = |c| \|\mathbf{v}\|\,, \quad \forall\, \mathbf{v} \in \mathcal{V}\,,\ c \in \mathbb{R}\,;$$

- triangle inequality:

$$\|\mathbf{v} + \mathbf{v}'\| \leq \|\mathbf{v}\| + \|\mathbf{v}'\|\,, \;\; \forall\, \mathbf{v}, \mathbf{v}' \in \mathcal{V}\,.$$

When the vector space $\mathcal{V}$ is endowed with a specific choice of a norm is called a *normed vector space*.

Orientation of a Real Vector Space

Please recall that the space of the alternating covariant tensors on a vector space $\mathcal{V}$, is itself a linear space over the same field of $\mathcal{V}$.

Specifically, if $\mathcal{V}$ is a real n-dimensional vector space, by means of (B.73), the space $\Lambda^n(\mathcal{V}^*)$ is a one-dimensional real space and hence it is isomorphic with $\mathbb{R}$ itself. Consequently, $\Lambda^n(\mathcal{V}^*)\backslash\{\boldsymbol{o}^*\}$ has exactly two connected components, i.e. the subsets related to $\mathbb{R}^+$ and $\mathbb{R}^-$, respectively.

With this specifications, the following definition can be introduced (see, e.g., Berger (1987), Lee (2012)).

Definition B.51. Let $\mathcal{V}$ be an n-dimensional vector space over $\mathbb{R}$ and let $\Lambda^n(\mathcal{V}^*)$ be the space of the alternating covariant n-tensors. An *orientation* $\mathcal{O}$ for $\mathcal{V}$ is the choice of one of the two connected components of $\Lambda^n(\mathcal{V}^*)\backslash\{\boldsymbol{o}^*\}$.

Any form ω^* of $\Lambda^n(\mathcal{V}^*)$ is called *positive* if it belongs to the fixed orientation $\mathcal{O}$, and a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of $\mathcal{V}$ is said to be *positively oriented* if $\omega^*(\mathbf{b}_1, \ldots, \mathbf{b}_n) > 0$ for any ω^* in $\mathcal{O}$.

Please observe that a multicovector ω'^* is in the same orientation $\mathcal{O}$ of ω^* if there exists a scalar $c > 0$ such that $\omega'^* = c\omega^*$.

As a matter of fact, ω^* can be chosen as a basis tensor for $\Lambda^n(\mathcal{V}^*)$, so that the coordinate of ω^* is trivially 1, while the coordinate of ω'^* is exactly the coefficient c. Then, since 1 and c are both in $\mathbb{R}^+$, the forms ω'^* and ω^* are in the same connected component of $\Lambda^n(\mathcal{V}^*)$.

It is worth pointing out that fixing a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ for $\mathcal{V}$ naturally induces an orientation $\mathcal{O}$, which is the one identified by the exterior product of the covectors $\mathbf{b}^1, \ldots, \mathbf{b}^n$ of the dual basis $\mathcal{B}^*$.

Actually, setting $\omega^* = \mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n$ and recalling that exterior product relies on the alternation map (cf. Definition B.38), by means of the property (B.54) one obtains

$$
\begin{aligned}
(\mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n)(\mathbf{b}_1, \ldots, \mathbf{b}_n) &= \operatorname{Alt}(\mathbf{b}^1 \otimes \cdots \otimes \mathbf{b}^n)(\mathbf{b}_1, \ldots, \mathbf{b}_n) \\
&= \frac{1}{n!} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)\, (\mathbf{b}^1 \otimes \cdots \otimes \mathbf{b}^n)(\mathbf{b}_{\sigma(1)}, \ldots, \mathbf{b}_{\sigma(n)}) \\
&= \frac{1}{n!} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)\, \mathbf{b}^1(\mathbf{b}_{\sigma(1)}) \cdots \mathbf{b}^n(\mathbf{b}_{\sigma(n)})\,,
\end{aligned}
$$

where, using the property (B.8), the scalars $\mathbf{b}^i(\mathbf{b}_{\sigma(i)})$, with $i = 1, \ldots, n$, are other than 0 only when $\sigma(i) = i$.

Consequently, the only permutation σ providing a non-vanishing contribution is the identity map $\mathbf{id}_{S_n}$ of the symmetric group S_n, and the above relation provides

$$
(\mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n)(\mathbf{b}_1, \ldots, \mathbf{b}_n) = \frac{1}{n!} > 0\,. \tag{B.95}
$$

Then, any other basis $\mathcal{B}' = \{\mathbf{b}'_1, \ldots, \mathbf{b}'_n\}$ of $\mathcal{V}$ satisfying the condition $(\mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n)(\mathbf{b}'_1, \ldots, \mathbf{b}'_n) > 0$, is positively oriented.

Since the bases $\mathcal{B}$ and $\mathcal{B}'$ identify the same orientation, we also say that they are *consistently oriented*.

Proposition B.52. *Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{B}' = \{\mathbf{b}'_1, \ldots, \mathbf{b}'_n\}$ be two bases for the real vector space $\mathcal{V}$. Then, they are consistently oriented if, and only if, the determinant of the change of basis map is positive:*

$$
\det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}\right) > 0\,. \tag{B.96}
$$

Proof. Consider the orientation $\mathcal{O}$ of $\mathcal{V}$ defined by the basis $\mathcal{B}$, that is the one associated with the multicovector $\mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n$.

By Proposition B.40, applying $\mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n$ to the vectors of $\mathcal{B}'$ reads

$$
(\mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n)(\mathbf{b}'_1, \ldots, \mathbf{b}'_n) = (\mathbf{b}'_1 \wedge \cdots \wedge \mathbf{b}'_n)(\mathbf{b}^1, \ldots, \mathbf{b}^n)\,.
$$

Moreover, recalling Proposition B.21, the change of basis map from $\mathcal{B}$

to $\mathcal{B}'$ satisfies the property $\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}(\mathbf{b}_i) = \mathbf{b}'_i$, obtaining

$$\begin{aligned}(\mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n)(\mathbf{b}'_1, \ldots, \mathbf{b}'_n) &= \left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}(\mathbf{b}_1) \wedge \cdots \wedge \boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}(\mathbf{b}_n)\right)(\mathbf{b}^1, \ldots, \mathbf{b}^n) \\ &= \det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}\right)(\mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_n)(\mathbf{b}^1, \ldots, \mathbf{b}^n)\,,\end{aligned}$$

where the defining property (B.78) of the determinant has been applied.

Then, using again the result of Proposition B.21, and also exploiting (B.95), one finally finds the following identity:

$$(\mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n)(\mathbf{b}'_1, \ldots, \mathbf{b}'_n) = \frac{\det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}\right)}{n!}\,.$$

It is now clear how the sign of $(\mathbf{b}^1 \wedge \cdots \wedge \mathbf{b}^n)(\mathbf{b}'_1, \ldots, \mathbf{b}'_n)$ is the same of the determinant of $\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}$. Hence, the positiveness of $\det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}\right)$ implies the consistency of the orientations associated with $\mathcal{B}$ and $\mathcal{B}'$, and vice-versa. □

Please notice that the consistency of orientation between the bases of a vector space $\mathcal{V}$ represents an equivalence relation.

In addition, exploiting the results of Proposition B.52, such an equivalence is the same as the relation between the automorphisms of $\mathcal{V}$ with positive determinant:

$$\mathcal{B} \sim_O \mathcal{B}' \quad \Leftrightarrow \quad \det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}\right) > 0\,. \tag{B.97}$$

In order to prove that the relation here above is actually an equivalence relation, let us show that it is reflective, symmetric and transitive.

Actually, since $\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}} = \mathrm{id}_{\mathcal{V}}$, the reflexivity of (B.97) readily results from (B.79):

$$\det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}}\right) = \det(\mathrm{id}_{\mathcal{V}}) = 1 > 0\,.$$

Moreover, recalling that $\boldsymbol{h}_{\mathcal{B}'}^{\mathcal{B}} = \boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'^{-1}}$, by (B.81) one has

$$\det\left(\boldsymbol{h}_{\mathcal{B}'}^{\mathcal{B}}\right) = \det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}\right)^{-1} > 0 \quad \Leftrightarrow \quad \det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}\right) > 0\,,$$

whence the symmetry of the binary relation (B.97).

Finally, if $\mathcal{B} \sim_O \mathcal{B}'$ and $\mathcal{B}' \sim_O \mathcal{B}''$, composing the relevant change of basis maps gives the linear transformation $\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}''} = \boldsymbol{h}_{\mathcal{B}'}^{\mathcal{B}''}\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}$. Since $\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}$ and $\boldsymbol{h}_{\mathcal{B}'}^{\mathcal{B}''}$ both have a positive determinant, the resulting map, by means of (B.80), satisfies

$$\det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}''}\right) = \det\left(\boldsymbol{h}_{\mathcal{B}'}^{\mathcal{B}''}\right)\det\left(\boldsymbol{h}_{\mathcal{B}}^{\mathcal{B}'}\right) > 0\,,$$

which proves the transitivity of the equivalence relation (B.97).

Remark. In the light of the equivalence relation (B.97), the orientation $\mathcal{O}$ of the vector space $\mathcal{V}$ can be alternatively defined as the equivalence class of the bases related by the automorphisms of $\mathcal{V}$ with positive determinant (see, e.g., Lee (2012), Mac Lane and Birkhoff (1999)).

B.4 Linear Maps and Tensors

Let us consider the $\mathbb{F}$-vector spaces $\mathcal{V}$ and $\mathcal{W}$. For any $\mathbf{w} \in \mathcal{W}$ and $\mathbf{v}^* \in \mathcal{V}^*$, the tensor product $\mathbf{w} \otimes \mathbf{v}^*$ is an element of the tensor space $\mathcal{W} \otimes \mathcal{V}^*$ and induces a linear map between $\mathcal{V}$ and $\mathcal{W}$ by means of the following identification:

$$\begin{aligned} \boldsymbol{f}_{\mathbf{w}\otimes\mathbf{v}^*} : \mathcal{V} &\to \mathcal{W} \\ \mathbf{x} &\mapsto \boldsymbol{f}_{\mathbf{w}\otimes\mathbf{v}^*}(\mathbf{x}) = \mathbf{v}^*(\mathbf{x})\,\mathbf{w}\,, \end{aligned} \tag{B.98}$$

where $\mathbf{v}^*(\mathbf{x})$ is the scalar of $\mathbb{F}$ resulting from applying the functional $\mathbf{v}^* \in \mathcal{V}^*$ at $\mathbf{x} \in \mathcal{V}$, so that the linearity of the functional $\mathbf{v}^*$ implies the one of the map $\boldsymbol{f}_{\mathbf{w}\otimes\mathbf{v}^*}$.

At the same time, if $\boldsymbol{g} : \mathcal{V} \to \mathcal{W}$ is a linear map, let us consider some $\mathbf{w} \in \mathcal{W}$ and $\mathbf{v}^* \in \mathcal{V}^*$ such that the tensor $(\mathbf{w} \otimes \mathbf{v}^*)_g$ is defined as

$$\begin{aligned} (\mathbf{w} \otimes \mathbf{v}^*)_g : \mathcal{W}^* \times \mathcal{V} &\to \mathbb{F} \\ (\mathbf{y}^*, \mathbf{x}) &\mapsto (\mathbf{w} \otimes \mathbf{v}^*)_g(\mathbf{y}^*, \mathbf{x}) = \mathbf{w}(\mathbf{y}^*)\mathbf{v}^*(\mathbf{x})\,, \end{aligned} \tag{B.99}$$

where, by means of (B.17), the usual identification $\mathcal{W}^{**} \cong \mathcal{W}$ has been exploited.

Moreover, applying an arbitrary $\mathbf{y}^* \in \mathcal{W}^*$ to the vector $\mathbf{v}^*(\mathbf{x})\,\mathbf{w} \in \mathcal{W}$

provided by the map (B.98), one has

$$\mathbf{y}^*\big(f_{\mathbf{w}\otimes\mathbf{v}^*}(\mathbf{x})\big) = \mathbf{v}^*(\mathbf{x})\mathbf{y}^*(\mathbf{w}) = \mathbf{w}(\mathbf{y}^*)\mathbf{v}^*(\mathbf{x}) = (\mathbf{w}\otimes\mathbf{v}^*)_g(\mathbf{y}^*,\mathbf{x})\,, \quad \forall\,\mathbf{x}\in\mathcal{V}\,,\ \mathbf{y}^*\in\mathcal{W}^*\,, \tag{B.100}$$

whence, for the arbitrariness of $\mathbf{y}\in\mathcal{W}^*$, as well as $\mathbf{x}\in\mathcal{V}$, it is clear that the linear map $f_{\mathbf{w}\otimes\mathbf{v}^*}$ induced by $\mathbf{w}\otimes\mathbf{v}^*$ is exactly the assigned map g inducing the tensor $(\mathbf{w}\otimes\mathbf{v}^*)_g$.

It is worth noting that the relation (B.100) implies the identification $f_{\mathbf{w}\otimes\mathbf{v}^*} = g$ on the condition that the linear map $g:\mathcal{V}\to\mathcal{W}$ induces a multilinear functional, defined on $\mathcal{W}^*\times\mathcal{V}$, to be expressed in the form of a tensor product $(\mathbf{w}\otimes\mathbf{v}^*)_g$.

However, such an identification can be easily extended to any tensor in $\mathcal{W}\otimes\mathcal{V}^*$, as stated in the following proposition.

Proposition B.53. *Let $\mathcal{V}$ and $\mathcal{W}$ be $\mathbb{F}$-vector spaces. Then, there exists a vector space isomorphism*

$$\begin{aligned}\Theta:\ \mathcal{L}(\mathcal{V},\mathcal{W}) &\to \mathcal{W}\otimes\mathcal{V}^*\\ f &\mapsto \mathbf{A} = \Theta(f)\,,\end{aligned} \tag{B.101}$$

such that

$$\mathbf{A}(\mathbf{y}^*,\mathbf{x}) = \mathbf{y}^*\big(f(\mathbf{x})\big)\,,\ \ \forall\,\mathbf{x}\in\mathcal{V}\,,\ \mathbf{y}^*\in\mathcal{W}^*\,. \tag{B.102}$$

Proof. To prove that Θ is linear, consider $f,g\in\mathcal{L}(\mathcal{V},\mathcal{W})$ and $\mathbf{A},\mathbf{B}\in\mathcal{W}\otimes\mathcal{V}^*$ such that $\mathbf{A}=\Theta(f)$ and $\mathbf{B}=\Theta(g)$.

By the linearity of $\mathcal{L}(\mathcal{V},\mathcal{W})$ stated in Proposition B.17, applying the defining property (B.102) to a linear combination of f and g, one has

$$\mathbf{y}^*\big((af+bg)(\mathbf{x})\big) = \mathbf{y}^*\big(af(\mathbf{x})+bg(\mathbf{x})\big) = a\mathbf{y}^*\big(f(\mathbf{x})\big)+b\mathbf{y}^*\big(g(\mathbf{x})\big)\,,$$

where the linearity of the functional $\mathbf{y}^*\in\mathcal{W}^*$ has been also exploited.

Hence, since (B.102) applies to both $\mathbf{A}=\Theta(f)$ and $\mathbf{B}=\Theta(g)$, the sum and the multiplication by a scalar of multilinear forms, defined by (B.18) and (B.19), respectively, provide

$$\mathbf{y}^*\big((a\boldsymbol{f} + b\boldsymbol{g})(\mathbf{x})\big) = a\mathbf{A}(\mathbf{y}^*, \mathbf{x}) + b\mathbf{B}(\mathbf{y}^*, \mathbf{x}) = (a\mathbf{A} + b\mathbf{B})(\mathbf{y}^*, \mathbf{x}) \,,$$

whence

$$a\mathbf{A} + b\mathbf{B} = \Theta(a\boldsymbol{f} + b\boldsymbol{g}) \,, \ \ \forall\, \boldsymbol{f}, \boldsymbol{g} \in \mathcal{L}(\mathcal{V}, \mathcal{W}) \,, \ a, b \in \mathbb{F} \,.$$

which assures the linearity of Θ.

The injectivity of Θ can be proved considering the condition $\Theta(\boldsymbol{f}) = \mathbf{0}$, where $\mathbf{0}$ is the null tensor in $\mathcal{W} \otimes \mathcal{V}^*$ and satisfies

$$\mathbf{0}(\mathbf{y}^*, \mathbf{x}) = 0 \,, \ \ \forall\, \mathbf{x} \in \mathcal{V} \,, \ \mathbf{y}^* \in \mathcal{W}^* \,,$$

which means

$$\mathbf{y}^*\big(\boldsymbol{f}(\mathbf{x})\big) = 0 \,, \ \ \forall\, \mathbf{x} \in \mathcal{V} \,, \ \mathbf{y}^* \in \mathcal{W}^* \,.$$

By Part 1 of Proposition B.25, the arbitrariness of $\mathbf{y}^* \in \mathcal{W}^*$ implies the vanishing of $\boldsymbol{f}(\mathbf{x}) \in \mathcal{W}$. Then, the condition

$$\boldsymbol{f}(\mathbf{x}) = \mathbf{o}_{\mathcal{W}} \,, \ \ \forall\, \mathbf{x} \in \mathcal{V}$$

is satisfied if, and only if, $\boldsymbol{f}$ is the null map $\boldsymbol{o} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, that is $\mathrm{Ker}(\Theta) = \{\boldsymbol{o}\}$ and Θ is injective.

In order to verify the surjectivity of Θ, let us consider the bases $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ for $\mathcal{V}$ and $\mathcal{W}$, respectively, with the relevant dual bases $\mathcal{A}^* = \{\mathbf{a}^1, \ldots, \mathbf{a}^n\}$ and $\mathcal{B}^* = \{\mathbf{b}^1, \ldots, \mathbf{b}^m\}$.

Consequently, specializing (B.30), a basis for $\mathcal{W} \otimes \mathcal{V}^*$ is given in the form $\mathcal{B}_\otimes = \{\, \mathbf{b}_i \otimes \mathbf{a}^j \mid i = 1, \ldots, m \,,\ j = 1, \ldots, n \,\}$ and any tensor $\mathbf{A} \in \mathcal{W} \otimes \mathcal{V}^*$ can be expressed as

$$\mathbf{A} = A^i{}_j \mathbf{b}_i \otimes \mathbf{a}^j \,,$$

where the coefficients $A^i{}_j = \mathbf{A}(\mathbf{b}^i, \mathbf{a}_j)$ are values of $\mathbf{A}$ relevant to the pairs $(\mathbf{b}^i, \mathbf{a}_j)$, with $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

Evaluating $\mathbf{A}$ at an arbitrary pair of vectors $(\mathbf{y}^*, \mathbf{x})$, one has

$$\mathbf{A}(\mathbf{y}^*, \mathbf{x}) = A^i{}_j \mathbf{b}_i \otimes \mathbf{a}^j(\mathbf{y}^*, \mathbf{x}) = A^i{}_j \mathbf{b}_i(\mathbf{y}^*)\mathbf{a}^j(\mathbf{x}) = A^i{}_j \mathbf{y}^*(\mathbf{b}_i)\mathbf{a}^j(\mathbf{x})$$

whence, for the linearity of $\mathbf{y}^* \in \mathcal{W}^*$, the following relation holds true:

$$\mathbf{A}(\mathbf{y}^*, \mathbf{x}) = \mathbf{y}^*\left(A^i{}_j \mathbf{a}^j(\mathbf{x})\, \mathbf{b}_i\right),$$

Since each $\mathbf{a}^j \in \mathcal{A}^*$ is a functional on $\mathcal{V}$, the linear composition $A^i{}_j \mathbf{a}^j(\mathbf{x})\, \mathbf{b}_i$ is a vector of $\mathcal{W}$ which depends linearly on $\mathbf{x} \in \mathcal{V}$, and can be considered as the image of the following linear map

$$\begin{aligned} \boldsymbol{f} : \mathcal{V} &\to \mathcal{W} \\ \mathbf{x} &\mapsto \boldsymbol{f}(\mathbf{x}) = A^i{}_j \mathbf{a}^j(\mathbf{x})\, \mathbf{b}_i\,. \end{aligned} \tag{B.103}$$

Since the relation (B.102) is fulfilled by construction, i.e.

$$\mathbf{y}^*\left(\boldsymbol{f}(\mathbf{x})\right) = \mathbf{y}^*\left(A^i{}_j \mathbf{a}^j(\mathbf{x})\, \mathbf{b}_i\right) = \mathbf{A}(\mathbf{y}^*, \mathbf{x}), \quad \forall\, \mathbf{x} \in \mathcal{V}, \ \mathbf{y}^* \in \mathcal{W}^*,$$

the map $\boldsymbol{f}$ is such that $\mathbf{A} = \Theta(\boldsymbol{f})$, ensuring the surjectivity of Θ. □

Please observe that the surjectivity of Θ has been proved introducing the bases $\mathcal{A}$ and $\mathcal{B}$. However, if different bases are considered, say $\mathcal{A}'$ and $\mathcal{B}'$, respectively, the map in (B.103) is the same, since the scalars $A^i{}_j$ are consistently modified:

$$\begin{aligned} \boldsymbol{f}(\mathbf{x}) &= A^i{}_j \mathbf{a}^j(\mathbf{x})\, \mathbf{b}_i = \mathbf{A}(\mathbf{b}^i, \mathbf{a}_j)\mathbf{a}^j(\mathbf{x})\, \mathbf{b}_i = \mathbf{A}(\mathbf{b}'^i, \mathbf{a}'_j)\mathbf{a}'^j(\mathbf{x})\, \mathbf{b}'_i \\ &= A'^i{}_j \mathbf{a}'^j(\mathbf{x})\, \mathbf{b}'_i\,. \end{aligned}$$

Then, the isomorphism (B.101) is canonical and the spaces $\mathcal{L}(\mathcal{V}, \mathcal{W})$ and $\mathcal{W} \otimes \mathcal{V}^*$ can be unambiguously identified:

$$\mathcal{L}(\mathcal{V}, \mathcal{W}) \cong \mathcal{W} \otimes \mathcal{V}^*\,. \tag{B.104}$$

Please notice that, using the identity (B.12), the relation between the linear map $\boldsymbol{f}$ and the tensor $\mathbf{A}$ also reads:

$$\mathbf{A}(\mathbf{y}^*, \mathbf{x}) = \mathbf{y}^*\left(\boldsymbol{f}(\mathbf{x})\right) = \left(\boldsymbol{f}^\mathsf{T}(\mathbf{y}^*)\right)(\mathbf{x}) = \mathbf{x}\left(\boldsymbol{f}^\mathsf{T}(\mathbf{y}^*)\right) = \mathbf{A}^\mathsf{T}(\mathbf{x}, \mathbf{y}^*)\,,$$

where the identification $\mathcal{V} \cong \mathcal{V}^{**}$ has been exploited and the notation $\mathbf{A}^\mathsf{T}$ has been introduced for the tensor associated with the transpose map $\boldsymbol{f}^\mathsf{T}$.

Consequently, the following definition can be introduced.

Definition B.54. The *transpose* of a tensor $\mathbf{A} \in \mathcal{W} \otimes \mathcal{V}^*$ is a tensor $\mathbf{A}^\mathsf{T} \in \mathcal{V}^* \otimes \mathcal{W}$ such that

$$\mathbf{A}^\mathsf{T}(\mathbf{x}, \mathbf{y}^*) = \mathbf{A}(\mathbf{y}^*, \mathbf{x}), \ \ \forall \, \mathbf{x} \in \mathcal{V}, \ \mathbf{y}^* \in \mathcal{W}^*. \tag{B.105}$$

Introducing the bases $\mathcal{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ for $\mathcal{V}$ and $\mathcal{W}$, respectively, the defining property of $\mathbf{A}^\mathsf{T}$ reads

$$\begin{aligned}\mathbf{A}^\mathsf{T}(\mathbf{x}, \mathbf{y}^*) &= {A^\mathsf{T}}_i^{\ j}\mathbf{a}^i \otimes \mathbf{b}_j(\mathbf{x}, \mathbf{y}^*) = {A^\mathsf{T}}_i^{\ j}\mathbf{a}^i(\mathbf{x})\mathbf{b}_j(\mathbf{y}^*) = {A^\mathsf{T}}_i^{\ j}\mathbf{b}_j(\mathbf{y}^*)\mathbf{a}^i(\mathbf{x}) \\ &= {A^\mathsf{T}}_i^{\ j}\mathbf{b}_j \otimes \mathbf{a}^i(\mathbf{y}^*, \mathbf{x}),\end{aligned}$$

whence, being also $\mathbf{A}(\mathbf{y}^*, \mathbf{x}) = A^j_{\ i}\mathbf{b}_j \otimes \mathbf{a}^i(\mathbf{y}^*, \mathbf{x})$, for the arbitrariness of $\mathbf{x}$ and $\mathbf{y}^*$, one finds

$${A^\mathsf{T}}_i^{\ j} = A^j_{\ i}. \tag{B.106}$$

B.4.1 Matrix Representation of Linear Maps

The canonical isomorphism B.102 implies that a tensor $\mathbf{A} \in \mathcal{W} \otimes \mathcal{V}^*$ can itself be considered as a linear operator from $\mathcal{V}$ to $\mathcal{W}$.

Moreover, introducing the bases $\mathcal{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ for $\mathcal{V}$ and $\mathcal{W}$, respectively, along with the relevant dual bases $\mathcal{A}^*$ and $\mathcal{B}^*$, the expression of the tensor $\mathbf{A}$ is

$$\mathbf{A} = A^i_{\ j}\mathbf{b}_i \otimes \mathbf{a}^j,$$

where the coefficients $A^i_{\ j}$, with $i = 1, \dots, m$ and $j = 1, \dots, n$ result from the evaluation $A^i_{\ j} = \mathbf{A}(\mathbf{b}^i, \mathbf{a}_j)$.

Each scalar $A^i_{\ j} \in \mathbb{F}$ can be considered as the entry placed at the i-th row and the j-th column of a matrix $[\mathbf{A}] \in M_{m \times n}$. Then, since any tensor $\mathbf{A}$ is uniquely related with both a linear map $\boldsymbol{f}$ and, through its coordinates, to a matrix $[\mathbf{A}]$, it is possible to define the following isomorphism between the space $\mathcal{L}(\mathcal{V},\mathcal{W})$ and the space $M_{m\times n}$:

$$\begin{aligned} \Psi_{(\mathcal{A},\mathcal{B})} :\ \mathcal{L}(\mathcal{V},\mathcal{W}) &\to M_{m\times n} & \qquad \Psi^{-1}_{(\mathcal{A},\mathcal{B})} :\ M_{m\times n} &\to \mathcal{L}(\mathcal{V},\mathcal{W}) \\ \boldsymbol{f} &\mapsto [\mathbf{A}]\,, & [\mathbf{A}] &\mapsto \boldsymbol{f}\,. \end{aligned} \tag{B.107}$$

Please observe that, differently from $\Theta : \mathcal{L}(\mathcal{V},\mathcal{W}) \to \mathcal{W} \otimes \mathcal{V}^*$, the isomorphism $\Psi_{(\mathcal{A},\mathcal{B})}$ is not canonical, but it depends on the bases $\mathcal{A}$ and $\mathcal{B}$ of the vector spaces.

However, by fixing the bases $\mathcal{A}$ and $\mathcal{B}$, the linear map $\boldsymbol{f} \in \mathcal{L}(\mathcal{V},\mathcal{W})$, the tensor $\mathbf{A} \in \mathcal{W} \otimes \mathcal{V}^*$ and the matrix $[\mathbf{A}] \in M_{m\times n}$ can be identified:

$$\boldsymbol{f} \cong \mathbf{A} \cong [\mathbf{A}]\,. \tag{B.108}$$

Moreover, recalling that the linear map $\boldsymbol{f}$ associated with $\mathbf{A}$ is given by (B.103), the image of $\mathbf{x} \in \mathcal{V}$ through $\boldsymbol{f}$ is a vector $\mathbf{y}$ expressed as

$$\mathbf{y} = \boldsymbol{f}(\mathbf{x}) = A^i{}_j \mathbf{a}^j(\mathbf{x})\mathbf{b}_i = A^i{}_j x^j \mathbf{b}_i\,,$$

where the role of $\mathbf{a}^j \in \mathcal{A}^*$ as the j-th coordinated function, defined by (B.6), has been exploited.

Then, expressing $\mathbf{y} \in \mathcal{W}$ with respect to the basis $\mathcal{B}$ as $\mathbf{y} = y^i \mathbf{b}_i$, one has

$$y^i = A^i{}_j x^j\,.$$

which also represents as the i-th entry of a column matrix $[\mathbf{y}] \in M_{m\times 1}$ resulting from the matrix multiplication of $[\mathbf{A}] \in M_{m\times n}$ and $[\mathbf{x}] \in M_{n\times 1}$:

$$[\mathbf{y}] = [\mathbf{A}][\mathbf{x}]\,. \tag{B.109}$$

Since the entries of $[\mathbf{x}]$ are exactly the coordinates of $\mathbf{x} \in \mathcal{V}$ with respect to the basis $\mathcal{A}$, the column matrix $[\mathbf{x}] \in M_{n\times 1}$ is isomorphic with the coordinated vector $\bar{\mathbf{x}} \in \mathbb{F}^n$, provided by (B.1) as $\bar{\mathbf{x}} = \varphi_{\mathcal{A}}(\mathbf{x})$.

More generally, if $\mathcal{V}$ is an n-dimensional vector space with a basis $\mathcal{B}$, the following map is a vector space isomorphism:

$$\begin{aligned} \theta_{\mathcal{B}} : \mathcal{V} &\to M_{n\times 1} & \theta_{\mathcal{B}}^{-1} : M_{n\times 1} &\to \mathcal{V} \\ \mathbf{v} &\mapsto [\mathbf{v}]\,, & [\mathbf{v}] &\mapsto \mathbf{v}\,, \end{aligned} \tag{B.110}$$

so that, once a basis $\mathcal{B}$ has been fixed, the vectors $\mathbf{v} \in \mathcal{V}$, $\bar{\mathbf{v}} \in \mathbb{F}^n$ and $[\mathbf{v}] \in M_{n\times 1}$ can be uniquely identified:

$$\mathbf{v} \cong \bar{\mathbf{v}} \cong [\mathbf{v}]\,. \tag{B.111}$$

B.4.2 Endomorphisms of a Vector Space

When the map f is an endomorphism of the vector space $\mathcal{V}$, the codomain of the map defined by (B.101) specializes to the tensor space $\mathcal{V} \otimes \mathcal{V}^*$, which is the space of the mixed tensors on $\mathcal{V}$ of rank $(1, 1)$.

Then, the identification (B.104) reads

$$\mathsf{End}(\mathcal{V}) \cong \mathcal{T}^{(1,1)}(\mathcal{V})\,, \tag{B.112}$$

and, introducing a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ for $\mathcal{V}$, a tensor $\mathbf{A} \in \mathcal{T}^{(1,1)}(\mathcal{V})$ results

$$\mathbf{A} = A^i{}_j \mathbf{b}_i \otimes \mathbf{b}^j\,,$$

with the associated matrix $[\mathbf{A}]$ in the space of the square matrices of order n, denoted as M_n.

Since by Definition B.18 the product of linear maps is the same as the composition, the product of endomorphisms of $\mathcal{V}$ is itself an endomorphism of $\mathcal{V}$, i.e.

$$fg \in \mathsf{End}(\mathcal{V})\,, \;\; \forall\, f, g \in \mathsf{End}(\mathcal{V})\,.$$

Exploiting the product of endomorphisms, the space $\mathsf{End}(\mathcal{V})$ can be endowed with the Lie bracket operation, which provides a Lie algebra denoted as $\mathfrak{gl}(\mathcal{V})$.

Definition B.55. Given an $\mathbb{F}$-vector space $\mathcal{V}$, the *general linear Lie algebra* of $\mathcal{V}$, denoted as $\mathfrak{gl}(\mathcal{V})$, is the Lie algebra of the endomorphisms of $\mathcal{V}$ obtained by endowing the space $\mathsf{End}(\mathcal{V})$ with the following operation:

$$\begin{aligned} [\cdot,\cdot]:\ & \mathfrak{gl}(\mathcal{V}) \times \mathfrak{gl}(\mathcal{V}) \to \mathfrak{gl}(\mathcal{V}) \\ & (f,g) \mapsto [f,g] = fg - gf\,. \end{aligned} \tag{B.113}$$

It is easy to verify that the operation defined by (B.113) is a Lie bracket satisfying the properties introduced by Definition A.27, and $\mathfrak{gl}(\mathcal{V})$ is actually a Lie algebra.

A similar algebraic structure results noting that the matrix multiplication is an internal operation for M_n:

$$[\mathbf{A}][\mathbf{B}] \in M_{n\times n}\,,\ \ \forall\,[\mathbf{A}],[\mathbf{B}] \in M_{n\times n}\,,$$

so that the commutator bracket is a well-defined binary operation for the space M_n of the n-by-n matrices.

Definition B.56. The *general linear Lie algebra* of order n over the field $\mathbb{F}$, denoted as $\mathfrak{gl}(n,\mathbb{F})$ or simply as $\mathfrak{gl}(n)$, is the Lie algebra resulting from the space of matrices M_n along with the following commutator bracket:

$$\begin{aligned} [\cdot,\cdot]:\ & \mathfrak{gl}(n) \times \mathfrak{gl}(n) \to \mathfrak{gl}(n) \\ & ([\mathbf{A}],[\mathbf{B}]) \mapsto \big[[\mathbf{A}],[\mathbf{B}]\big] = [\mathbf{A}][\mathbf{B}] - [\mathbf{B}][\mathbf{A}]\,. \end{aligned} \tag{B.114}$$

Please notice that, since $\mathsf{End}(\mathcal{V})$ is isomorphic with M_n and the composition of linear maps results in the matrix multiplication of the relevant matrices, $\mathfrak{gl}(\mathcal{V})$ and $\mathfrak{gl}(n)$ are isomorphic as Lie algebras:

$$\mathfrak{gl}(\mathcal{V}) \cong \mathfrak{gl}(n)\,. \tag{B.115}$$

B.4.3 The General Linear Group

Within the endomorphisms on a vector space $\mathcal{V}$, a key role is played by the automorphisms, i.e. the bijective linear maps of $\mathcal{V}$.

Since, by Proposition B.44, the determinant of any bijective endomorphism of $\mathcal{V}$ is non-null, the set of the automorphisms of $\mathcal{V}$ can be characterized as follows:

$$\mathsf{Aut}(\mathcal{V}) = \{ \boldsymbol{f} : \mathcal{V} \to \mathcal{V} \mid \det(\boldsymbol{f}) \neq 0 \} \subseteq \mathsf{End}(\mathcal{V}) . \tag{B.116}$$

Moreover, since composing bijective maps results in a bijection, the product of linear maps, introduced by Definition B.18, is an internal operation for $\mathsf{Aut}(\mathcal{V})$.

Specifically, for any pair of maps $\boldsymbol{f}$ and $\boldsymbol{g}$ in $\mathsf{Aut}(\mathcal{V})$, Property 2 in Proposition B.45 reads

$$\det(\boldsymbol{f}\boldsymbol{g}) = \det(\boldsymbol{f}) \det(\boldsymbol{g}) \neq 0 , \quad \forall \boldsymbol{f}, \boldsymbol{g} \in \mathsf{Aut}(\mathcal{V}) ,$$

where the non-vanishing of both $\det(\boldsymbol{f})$ and $\det(\boldsymbol{g})$ implies that $\det(\boldsymbol{f}\boldsymbol{g})$ is itself non-null. Hence, consistently with (B.116), the product $\boldsymbol{f}\boldsymbol{g}$ is a linear map in $\mathsf{Aut}(\mathcal{V})$.

It is easily to see that the set $\mathsf{Aut}(\mathcal{V})$, along with the product of linear maps, has a group structure (cf. Definition A.8). In fact, the product of maps is an associative operation, whose identity element is the identity map $\mathbf{id}_{\mathcal{V}}$, and such that the inverse element of any $\boldsymbol{f}$, with respect to the product, is given by the inverse map $\boldsymbol{f}^{-1}$.

Definition B.57. The *general linear group* of an $\mathbb{F}$-vector space $\mathcal{V}$ is the set $\mathsf{Aut}(\mathcal{V})$ of the automorphisms of $\mathcal{V}$ along with the product of linear transformations, and is denoted as $GL(\mathcal{V})$.

The characterization of the automorphisms of $\mathcal{V}$ in terms of matrix representation comes from the following proposition.

Proposition B.58. *Let $\boldsymbol{f}$ be an automorphism of an $\mathbb{F}$-vector space $\mathcal{V}$. Then, the relevant matrix $[\mathbf{A}]$ is invertible and the inverse matrix $[\mathbf{A}]^{-1}$ is associated with the inverse map $\boldsymbol{f}^{-1}$.*

Proof. Let us consider a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ for $\mathcal{V}$. Specializing (B.103) to the tensor $\mathbf{A} \in \mathcal{V}^* \otimes \mathcal{V}$, the image of an arbitrary $\mathbf{x} \in \mathcal{V}$ through $\boldsymbol{f}$ can be represented as

$$\boldsymbol{f}(\mathbf{x}) = A^i{}_j \mathbf{b}^j(\mathbf{x}) \, \mathbf{b}_i .$$

At the same time, if $\mathbf{B} \in \mathcal{V}^* \otimes \mathcal{V}$ is the tensor associated with the inverse map $\boldsymbol{f}^{-1}\colon \mathcal{V} \to \mathcal{V}$, one has

$$\begin{aligned}\mathbf{x} &= \boldsymbol{f}^{-1}\big(\boldsymbol{f}(\mathbf{x})\big) = B^h{}_k \mathbf{b}^k\big(\boldsymbol{f}(\mathbf{x})\big)\,\mathbf{b}_h = B^h{}_k \mathbf{b}^k\big(A^i{}_j \mathbf{b}^j(\mathbf{x})\,\mathbf{b}_i\big)\,\mathbf{b}_h \\ &= B^h{}_k A^i{}_j \mathbf{b}^j(\mathbf{x})\,\mathbf{b}^k(\mathbf{b}_i)\,\mathbf{b}_h\,,\end{aligned}$$

where the linearity of the functional $\mathbf{b}^k \in \mathcal{V}^*$ has been exploited.

Moreover, applying Definition B.24 and considering $\mathbf{b}^j$ as the j-th coordinate function defined by (B.6), one also infers:

$$\mathbf{x} = B^h{}_k A^i{}_j x^j \delta^k_i\,\mathbf{b}_h = B^h{}_i A^i{}_j x^j \mathbf{b}_h\,.$$

Then, expressing $\mathbf{x}$ as $x^h \mathbf{b}_h = \delta^h_j x^j \mathbf{b}_h$, the comparison of the coordinates implies $\delta^h_j x^j = B^h{}_i A^i{}_j x^j$, whence, for the arbitrariness of the coordinates x^j representing the vector $\mathbf{x}$, one also finds

$$B^h{}_i A^i{}_j = \delta^h_j\,.$$

Such a relation is the component-wise expression of the matrix product

$$[\mathbf{B}][\mathbf{A}] = [\mathbf{I}]\,,$$

which implies $[\mathbf{B}] = [\mathbf{A}]^{-1}$. □

It is trivial to check that the set of all the invertible matrices in M_n, along with the matrix multiplication, is a group with the identity given by $[\mathbf{I}]$ and the inverse element given by the inverse matrix.

Definition B.59. The *general linear group* of degree n over the field $\mathbb{F}$ is the set of all the invertible n-by-n matrices along with the standard matrix multiplication, and it is denoted as $GL(n, \mathbb{F})$, or simply $GL(n)$ when the field is deducible.

By virtue of Proposition (B.58), the product of linear maps in $GL(\mathcal{V})$ and the matrix multiplication in $GL(n)$ do correspond to each other, and then these two sets are isomorphic also as groups:

$$GL(\mathcal{V}) \cong GL(n)\,. \tag{B.117}$$

B.4.4 Change of Basis Matrix

Consider a vector space $\mathcal{V}$ over the field $\mathbb{F}$ and suppose two bases are introduced, say $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$.

It has been shown in Proposition B.21 that the change of basis from $\mathcal{A}$ to $\mathcal{B}$ is provided by an automorphism $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}$ mapping each vector $\mathbf{a}_i \in \mathcal{A}$ to the relevant vector $\mathbf{b}_i \in \mathcal{B}$, with $i = 1, \ldots, n$.

Let $\mathbf{C}_{\mathcal{A}}^{\mathcal{B}}$ be the tensor in $\mathcal{V} \otimes \mathcal{V}^*$ associated with the automorphism $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}$, which is expressed with respect to the basis $\mathcal{A}$ as

$$\mathbf{C}_{\mathcal{A}}^{\mathcal{B}} = (C_{\mathcal{A}}^{\mathcal{B}})^i{}_j \mathbf{a}_i \otimes \mathbf{a}^j \,,$$

so that, specializing (B.103) to $\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}} \colon \mathcal{V} \to \mathcal{V}$, one has

$$\boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}(\mathbf{x}) = (C_{\mathcal{A}}^{\mathcal{B}})^i{}_j \mathbf{a}^j(\mathbf{x})\, \mathbf{a}_i \,, \quad \forall\, \mathbf{x} \in \mathcal{V} \,,$$

which, using the defining property (B.4), provides

$$\mathbf{b}_k = \boldsymbol{h}_{\mathcal{A}}^{\mathcal{B}}(\mathbf{a}_k) = (C_{\mathcal{A}}^{\mathcal{B}})^i{}_j \mathbf{a}^j(\mathbf{a}_k)\, \mathbf{a}_i = (C_{\mathcal{A}}^{\mathcal{B}})^i{}_j \delta^j_k\, \mathbf{a}_i \,.$$

that is

$$\mathbf{b}_k = (C_{\mathcal{A}}^{\mathcal{B}})^i{}_k\, \mathbf{a}_i \,. \tag{B.118}$$

The matrix $[\mathbf{C}_{\mathcal{A}}^{\mathcal{B}}]$, whose entries are the coefficients $(C_{\mathcal{A}}^{\mathcal{B}})^i{}_k$ of the tensor $\mathbf{C}_{\mathcal{A}}^{\mathcal{B}}$ with respect to the basis $\mathcal{A}$, represents the *active* matrix of change of basis from $\mathcal{A}$ to $\mathcal{B}$. In fact, since the entries of k-th row are the coordinates of $\mathbf{b}_k$ with respect to the basis $\mathcal{A}$, the role of $[\mathbf{C}_{\mathcal{A}}^{\mathcal{B}}]$ is to transform the vectors of $\mathcal{A}$ into the vectors of $\mathcal{B}$.

Consider now an arbitrary vector $\mathbf{v}$ of $\mathcal{V}$, whose coordinate n-tuples with respect to the bases $\mathcal{A}$ and $\mathcal{B}$ are $\bar{\mathbf{v}}_{\mathcal{A}}$ and $\bar{\mathbf{v}}_{\mathcal{B}}$, respectively. Since $\mathbf{v}$ can be expressed as $\mathbf{v} = v_{\mathcal{A}}^i \mathbf{a}_i$, with respect to $\mathcal{A}$, and at the same time as $\mathbf{v} = v_{\mathcal{B}}^k \mathbf{b}_k$, with respect to $\mathcal{B}$, the following identity holds true:

$$v_{\mathcal{A}}^i \mathbf{a}_i = v_{\mathcal{B}}^k \mathbf{b}_k = (C_{\mathcal{A}}^{\mathcal{B}})^i{}_k\, v_{\mathcal{B}}^k \mathbf{a}_i \,,$$

where the transformation (B.118) has been exploited.

By comparison, one easily finds $v^i_{\mathcal{A}} = (C^{\mathcal{B}}_{\mathcal{A}})^i{}_k \, v^k_{\mathcal{B}}$, which in matrix form reads

$$[\mathbf{v}_{\mathcal{A}}] = [\mathbf{C}^{\mathcal{B}}_{\mathcal{A}}][\mathbf{v}_{\mathcal{B}}]\,,$$

or equivalently

$$[\mathbf{v}_{\mathcal{B}}] = [\mathbf{C}^{\mathcal{B}}_{\mathcal{A}}]^{-1}[\mathbf{v}_{\mathcal{A}}] = [\mathbf{C}^{\mathcal{A}}_{\mathcal{B}}][\mathbf{v}_{\mathcal{A}}]\,. \tag{B.119}$$

The matrix $[\mathbf{C}^{\mathcal{A}}_{\mathcal{B}}] = [\mathbf{C}^{\mathcal{B}}_{\mathcal{A}}]^{-1}$ represents the *passive* change of basis from $\mathcal{A}$ to $\mathcal{B}$, with the meaning to transform the coordinates of a fixed vector with respect to $\mathcal{A}$ to the ones with respect to $\mathcal{B}$.

Bibliography

Angeles, J. (1998). The application of dual algebra to kinematic analysis. In *Computational Methods in Mechanical Systems*, Volume 161 of *NATO ASI Series (Series F: Computer and Systems Sciences)*, pp. 953–1173. Berlin: Springer-Verlag.

Antman, S. S. (1974). Kirchhoff's problem for nonlinearly elastic rods. *Quarterly of Applied Mathematics 32*(3), 221–240.

Antman, S. S. (2005). *Nonlinear Problems of Elasticity* (Second ed.). New York: Springer-Verlag.

Arnold, V. I. (1989). *Mathematical Methods of Classical Mechanics* (Second ed.). New York: Springer-Verlag.

Auricchio, F., P. Carotenuto, and A. Reali (2008). On the geometrically exact beam model: A consistent, effective and simple derivation from three-dimensional finite-elasticity. *Int. J. Solids Struct. 45*(17), 4766–4781.

Balaeff, A., L. Mahadevan, and K. Schulten (2006). Modeling DNA loops using the theory of elasticity. *Phys. Rev. E 73*(0319199), 1–23.

Belytschko, T. and L. W. Glaum (1979). Applications of higher order corotational stretch theories to nonlinear finite element analysis. *Comput. Struct. 10*(1), 175–182.

Belytschko, T. and B. J. Hsieh (1973). Non-linear transient finite element analysis with convected co-ordinates. *Int. J. Numer. Methods Eng. 7*(3), 255–271.

Belytschko, T., L. Schwer, and M. J. Klein (1977). Large displacement, transient analysis of space frames. *Int. J. Numer. Methods Eng. 11*(1), 65–84.

Berger, M. (1987). *Geometry I* (First ed.). Berlin: Springer-Verlag.

Bergman, C. H. (2011). *Universal Algebra: Fundamentals and Selected Topics* (First ed.). Chapman and Hall/CRC.

Boothby, W. M. (2002). *An Introduction to Differentiable Manifolds and Riemannian Geometry* (Second ed.). Academic Press.

Borri, M. and C. Bottasso (1994). An intrinsic beam model based on a helicoidal approximation–Part I: Formulation. *Int. J. Numer. Methods Eng. 37*(13), 2267–2289.

Bottema, O. and B. Roth (1990). *Theoretical Kinematics* (First ed.). Dover Publications.

Cardona, A. and M. Geradin (1988). A beam finite element non-linear theory with finite rotations. *Int. J. Numer. Methods Eng. 26*(11), 2403–2438.

Cesnik, C. E. and S. Shin (2001). On the modeling of integrally actuated helicopter blades. *Int. J. Solids Struct. 38*(10), 1765–1789.

Chevallier, D. P. (1996). On the transference principle in kinematics: its various forms and limitations. *Mech. Mach. Theory 31*(1), 57–76.

Clifford, W. K. (1871, 11). Preliminary sketch of biquaternions. *Proc. London Math. Soc. s1-4*(1), 381–395.

Cohn, P. M. (1981). *Universal Algebra* (First ed.). Dordrecht: D. Reidel Publishing.

Coleman, B. and D. Swigon (2000). Theory of supercoiled elastic rings with self-contact and its application to DNA plasmids. *J. Elast. 60*(3), 173–221.

Cosserat, E. and F. Cosserat (1909). Théorie des corps déformables. In *Chwolson's Traité de Physique* (Second ed.)., pp. 953–1173. Paris: Librairie scientifique A. Herman et fils.

Crisfield, M. A. (1990). A consistent co-rotational formulation for non-linear, three-dimensional, beam-elements. *Comput. Methods Appl. Mech. Eng. 81*(2), 131–150.

Crisfield, M. A. and G. Jelenić (1999). Objectivity of strain measures in the geometrically exact three-dimensional beam theory and its finite-element implementation. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 455*(1983), 1125–1147.

Daher, M. (2013). *Dual Numbers and Invariant Theory of the Euclidean Group with Applications to Robotics*. Ph. D. thesis, Victoria University of Wellington.

Dimentberg, F. M. (1968). The screw calculus and its applications in mechanics. Technical report, Foreign Technology Division.

Dimitrienko, Y. I. (2002). *Tensor Analysis and Nonlinear Tensor Functions* (First ed.). Dordrecht: Springer.

do Carmo, M. P. (2016). *Differential Geometry of Curves and Surfaces: Revised & Updated Second Edition* (Second ed.). New York: Dover Publications.

Dummit, D. S. and R. M. Foote (2003). *Abstract Algebra* (Third ed.). Hoboken: Wiley.

Epstein, M. (2010). *The Geometrical Language of Continuum Mechanics* (First ed.). New York: Cambridge University Press.

Epstein, M. and R. Segev (1980). Differentiable manifolds and the principle of virtual work in continuum mechanics. *J. Math. Phys. 21*(5), 1243–1245.

Ericksen, J. and C. Truesdell (1957). Exact theory of stress and strain in rods and shells. *Arch. Ration. Mech. Anal. 1*, 295–323.

Eugster, S. (2015). *Geometric Continuum Mechanics and Induced Beam Theories* (First ed.), Volume 75. Springer International Publishing.

Faccio Júnior, C. J., A. C. Pegoraro Cardozo, V. Monteiro Júnior, and A. Gay Neto (2019). Modeling wind turbine blades by geometrically-exact beam and shell elements: A comparative approach. *Eng. Struct. 180*, 357–378.

Fischer, I. S. (1999). *Dual-Number Methods in Kinematics, Statics and Dynamics* (First ed.). Boca Raton: CRC Press.

Gallier, J. and J. Quaintance (2020). *Differential Geometry and Lie Groups: A Computational Perspective* (First ed.). Cham: Springer International Publishing.

Garcea, G., A. Madeo, and R. Casciaro (2012a). Nonlinear FEM analysis for beams and plate assemblages based on the implicit corotational method. *J. Mech. Mater. Struct. 7*, 539–572.

Garcea, G., A. Madeo, and R. Casciaro (2012b). The implicit corotational method and its use in the derivation of nonlinear structural models for beams and plates. *J. Mech. Mater. Struct. 7*, 509–538.

Goldstein, H., C. Poole, and J. Safko (2001). *Classical Mechanics* (Third ed.). Pearsong.

Grazioso, S., G. Di Gironimo, and B. Siciliano (2019a). A geometrically exact model for soft continuum robots: The finite element deformation space formulation. *Soft Rob. 6*(6), 790–811.

Grazioso, S., G. Di Gironimo, and B. Siciliano (2019b). From differential geometry of curves to helical kinematics of continuum robots using exponential mapping. *Advances in Robot Kinematics 2018 8*, 319–326.

Hall, B. C. (2015). *Lie Groups, Lie Algebras and Representations* (Second ed.). Springer.

Halmos, P. R. (1958). *Finite Dimensional Vector Spaces* (Second ed.). Princeton University Press.

Harsch, J., G. Capobianco, and S. R. Eugster (2021). Finite element formulations for constrained spatial nonlinear beam theories. *Math. Mech. Solids 26*(12), 1838–1863.

Higham, N. J. (2008). *Functions of Matrices: Theory and Computation* (First ed.). Philadelphia: Society for Industrial and Applied Mathematics.

Ibrahimbegović, A. (1995). On finite element implementation of geometrically nonlinear Reissner's beam theory: three-dimensional curved beam elements. *Comput. Methods Appl. Mech. Eng. 122*(1), 11–26.

Ibrahimbegović, A., F. Frey, and I. Kožar (1995). Computational aspects of vector-like parametrization of three-dimensional finite rotations. *Int. J. Numer. Methods Eng. 38*(21), 3653–3673.

Iserles, A., H. Z. Munthe-Kaas, S. P. Nørsett, and A. Zanna (2000). Lie-group methods. *Acta Numer. 9*, 215–365.

Jelenić, G. and M. A. Crisfield (1998). Interpolation of rotational variables in nonlinear dynamics of 3D beams. *Int. J. Numer. Methods Eng. 43*(7), 1193–1222.

Jelenić, G. and M. A. Crisfield (1999). Geometrically exact 3D beam theory: implementation of a strain-invariant finite element for statics and dynamics. *Comput. Methods Appl. Mech. Eng. 171*(1), 141–171.

Lang, S. (1987). *Linear Algebra* (Third ed.). New York: Springer-Verlag.

Lang, S. (2002). *Algebra* (Third ed.). New York: Springer-Verlag.

Lee, J. M. (2012). *Introduction to Smooth Manifolds* (Second ed.). New York: Springer-Verlag.

Love, A. E. H. (1944). *A treatise on the mathematical theory of elasticity* (Fourth ed.). New York: Dover.

Mac Lane, S. and G. Birkhoff (1999). *Algebra* (Third ed.). Providence: American Mathematical Society.

Magisano, D., L. Leonetti, and G. Garcea (2017). How to improve efficiency and robustness of the Newton method in geometrically non-linear structural problem discretized via displacement-based finite elements. *Comput. Methods Appl. Mech. Eng. 313*, 986–1005.

Magisano, D., L. Leonetti, A. Madeo, and G. Garcea (2020). A large rotation finite element analysis of 3D beams by incremental rotation vector and exact strain measure with all the desirable features. *Comput. Methods Appl. Mech. Eng. 361*, 112811.

Martínez, J. M. R. and J. Duffy (1993). The principle of transference: History, statement and proof. *Mech. Mach. Theory 28*, 165–177.

McCarthy, J. M. (1990). *Introduction to Theoretical Kinematics* (First ed.). Cambridge: The MIT Press.

Munthe-Kaas, H. (1998). Runge-Kutta methods on Lie groups. *BIT Numer. Math. 38*, 92–111.

Oran, C. (1973a). Tangent stiffness in plane frames. *ASCE J. Engrg. Mech. Div 99*(6), 973–985.

Oran, C. (1973b). Tangent stiffness in space frames. *ASCE J. Engrg. Mech. Div 99*(6), 987–1001.

Oran, C. and A. Kassimali (1976). Large deformations of framed structures under static and dynamic loads. *Comput. Struct. 6*(6), 539–547.

Paradiso, M., F. Marmo, and L. Rosati (2019). Consistent derivation of a beam model from the Saint Venant's solid model. *Int. J. Solids Struct. 159*, 90–110.

Paradiso, M., S. Sessa, N. Vaiana, F. Marmo, and L. Rosati (2021). Shear properties of isotropic and homogeneous beam-like solids having arbitrary cross sections. *Int. J. Solids Struct. 216*, 231–249.

Paradiso, M., N. Vaiana, S. Sessa, F. Marmo, and L. Rosati (2020). A BEM approach to the evaluation of warping functions in the Saint Venant theory. *Eng. Anal. Boundary Elem. 113*, 359–371.

Park, F. C. (1995). Distance Metrics on the Rigid-Body Motions with Applications to Mechanism Design. *J. Mech. Des. 117*(1), 48–54.

Pennestrì, E. and P. P. Valentini (2009). Linear dual algebra algorithms and their application to kinematics. In *Multibody Dynamics*, pp. 207–229. Springer.

Reissner, E. (1972). On one-dimensional finite-strain beam theory: The plane problem. *J. Appl. Math. Phys. (ZAMP) 23*, 795–804.

Rodrigues, O. (1840). Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendants des causes qui peuvent les produire. *J. Math. Pures Appl. 5*, 380–440.

Rucker, D. C., B. A. Jones, and R. J. Webster III (2010). A geometrically exact model for externally loaded concentric-tube continuum robots. *IEEE Trans. Rob. 26*(5), 769–780.

Rucker, D. C. and R. J. Webster III (2011). Statics and dynamics of continuum robots with general tendon routing and external loading. *IEEE Trans. Rob. 27*(6), 1033–1044.

Rudin, W. (1986). *Real and Complex Analysis* (Third ed.). Singapore: McGraw-Hill.

Sabale, A. and K. N. Gopal (2019). Nonlinear aeroelastic response of wind turbines using Simo-Vu-Quoc rods. *Appl. Math. Modell. 65*, 696–716.

Sadati, S. M. H., S. E. Naghibi, A. Shiva, Y. Noh, A. Gupta, I. D. Walker, K. Althoefer, and T. Nanayakkara (2017). A geometry deformation model for braided continuum manipulators. *Front. Rob. AI 4*, 22.

Santana, M. V. B., C. Sansour, M. Hjiaj, and H. Somja (2021). An equilibrium-based formulation with nonlinear configuration dependent interpolation for geometrically exact 3d beams. *Int. J. Numer. Methods Eng.* (Accepted article), 1–21.

Selig, J. M. (2005). *Geometric Fundamentals of Robotics* (Second ed.). New York: Springer-Verlag.

Simo, J. and D. Fox (1989). On a stress resultant geometrically exact shell model. Part I: Formulation and optimal parametrization. *Computer Methods in Applied Mechanics and Engineering 72*(3), 267–304.

Simo, J. C. (1985). A finite strain beam formulation. The three-dimensional dynamic problem. Part I. *Comput. Methods Appl. Mech. Eng. 49*(1), 55–70.

Simo, J. C. and L. Vu-Quoc (1986). A three-dimensional finite-strain rod model. Part II: Computational aspects. *Comput. Methods Appl. Mech. Eng. 58*(1), 79–116.

Simo, J. C. and L. Vu-Quoc (1991). A Geometrically-exact rod model incorporating shear and torsion-warping deformation. *Int. J. Solids Struct. 27*(3), 371–393.

Sonneville, V., A. Cardona, and O. Brüls (2014). Geometrically exact beam finite element formulated on the Special Euclidean Group SE(3). *Comput. Methods Appl. Mech. Eng. 268*, 451–474.

Spivak, M. (1999). *A Comprehensive Introduction to Differential Geometry* (Third ed.), Volume 1. Houston: Publish or Perish, Inc.

Swigon, D., B. D. Coleman, and I. Tobias (1998). The elastic rod model for DNA and its application to the tertiary structure of DNA minicircles in mononucleosomes. *Biophys. J. 74*(5), 2515–2530.

Tarrida, A. R. (2011). *Affine Maps, Euclidean Motions and Quadrics* (Third ed.). London: Springer-Verlag.

Trivedi, D., A. Lotfi, and C. D. Rahn (2008). Geometrically exact models for soft robotic manipulators. *IEEE Trans. Rob. 24*(4), 773–780.

Westcott, T. P., I. Tobias, and W. K. Olsons (1995). Elasticity theory and numerical analysis of DNA supercoiling: An application to DNA looping. *J. Phys. Chem. 99*(51), 17926–17935.

Winitzki, S. (2020). *Linear Algebra via Exterior Products.*

Yang, A. T. and F. Freudenstein (1964). Application of dual-number quaternion algebra to the analysis of spatial mechanisms. *J. Appl. Mech. 31*(2), 300–308.

Yoshiaki, G., W. Yasuhito, K. Toshihiro, and O. Makoto (1992). Elastic buckling phenomenon applicable to deployable rings. *Int. J. Solids Struct. 29*(7), 893–909.

Žefran, M. and V. Kumar (1998). Interpolation schemes for rigid body motions. *Comput.-Aided Des. 30*(3), 179–189.

Zienkiewicz, O., R. Taylor, and J. Zhu (2013). *The Finite Element Method: its Basis and Fundamentals* (Seventh ed.). Oxford: Butterworth-Heinemann.

Zupan, D. and M. Saje (2003a). Finite-element formulation of geometrically exact three-dimensional beam theories based on interpolation of strain measures. *Comput. Methods Appl. Mech. Eng. 192*(49), 5209–5248.

Zupan, D. and M. Saje (2003b). The three-dimensional beam theory: Finite element formulation based on curvature. *Comput. Struct. 81*(18), 1875–1888.