Tutorial

Manifold Calculus in System Theory and Control—Fundamentals and First-Order Systems

Simone Fiori

Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Via Brecce Bianche,
I-60131 Ancona, Italy; s.fiori@univpm.it

Abstract: The aim of the present tutorial paper is to recall notions from manifold calculus and to illustrate how these tools prove useful in describing system-theoretic properties. Special emphasis is put on embedded manifold calculus (which is coordinate-free and relies on the embedding of a manifold into a larger ambient space). In addition, we also consider the control of non-linear systems whose states belong to curved manifolds. As a case study, synchronization of non-linear systems by feedback control on smooth manifolds (including Lie groups) is surveyed. Special emphasis is also put on numerical methods to simulate non-linear control systems on curved manifolds. The present tutorial is meant to cover a portion of the mentioned topics, such as first-order systems, but it does not cover topics such as covariant derivation and second-order dynamical systems, which will be covered in a subsequent tutorial paper.

1. Introduction

The theory of dynamical systems whose state spaces possess the structures of curved manifolds has been applied primarily in physics (especially to mathematically describe the theory of general relativity at the beginning of the 20th century). More recently, dynamical systems on manifolds have proven their relevance in a number of subjects in engineering and applied sciences such as in robotics and in biomedical engineering [1–5] to name a few. The observation at the core of such applications is that those dynamical systems whose descriptive variables are bound to one another by non-linear holonomic constraints may be studied by means of the rich variety of mathematical tools provided by manifold calculus (a shortened terminology to denote 'calculus on manifold') and may be framed in the class of *dynamical systems on manifold*.

The structures of the state manifolds of such dynamical systems depend on the application at hand. As a concrete example, whenever the dynamics of a rigid body is concerned, the state manifold of choice is the special orthogonal group $\mathbb{SO}(3)$, because it encodes the attitude of a flying drone or a submarine robot. Most applications of interest (such as the computation of the dynamics of flying bodies) concern well-studied and well-understood curved manifolds, such as the special orthogonal group, the unit hypersphere and the group of symmetric, positive-definite matrices.

From the perspective of performing numerical simulations of dynamical systems on a computing platform, it is necessary to design adequate numerical methods to compute approximate solutions which still meet the structures of the state manifolds. Classical numerical methods, such as those in the Euler class or in the Runge–Kutta class, will fail when applied directly to such dynamical systems because they were designed to work on flat spaces and cannot cope with non-flat manifolds [6].

As a specific applied field, the time synchronization of first-order dynamical systems on curved state manifolds by non-linear control will be surveyed. The time synchronization of non-autonomous dynamical systems has been applied in physiology [7,8], ecology [9],

atmosphere physics [10], neurology [11] and many more applied fields [12–14]. Synchronization theory is a topic that appears to be very interesting from an abstract point of view and, at the same time, very useful in a number of applications. Indeed, synchronization theory appears to be an exciting and interdisciplinary research topic that combines key ideas from system theory, control theory and manifold calculus. In particular, non-linear control theory on manifolds may supply tools to design control fields to force a pair of dynamical systems into synchronizing their dynamics over time [15,16].

The present tutorial paper is devoted to recalling fundamental notions from manifold calculus and to explaining how these concepts apply to system theory and to non-linear control theory, taking time synchronization as a representative case of study in non-linear control. The present contribution is oriented to readers who either possess a good command of calculus but not of manifold calculus, or to readers who claim an understanding of theoretical manifold calculus, but do not have enough insight in the application/computation aspects of this field. In addition, a basic understanding of system theory and control theory will be assumed. The content of the present tutorial paper may be summarized as follows:

- It provides a clear and well-motivated introduction to manifold calculus, the basis of system and control theories on manifolds, with special emphasis on computational and applicational aspects. The present contribution provides practical formulas to deal with those *real-valued* manifolds that, in the author's experience, are the most accessed in engineering and applied science problems. As a matter of fact, *complex-valued* manifolds are not treated at all.
- It clearly states and illustrates the idea that, when one wishes to perform a simulation, by a computing platform, of dynamical systems on manifolds described in terms of differential equations, it is necessary to time-discretize such differential equations in a suitable way. In order to achieve such a discretization, it is not safe to invoke standard discretization methods (such as the ones based on Euler forward–backward discretization), which do not work as they stand on curved manifolds. One should therefore resort to more sophisticated numerical integration techniques.
- By the author's choice, the present tutorial paper does not carry any graphical illustrations nor any numerical simulation results. Readers who are interested in deepening their understanding of this topic are invited to to sketch graphs autonomously and to code examples in their favorite programming language.

The present paper is organized as follows. Section 2 lays out some introductory material, such as the motivation behind manifold calculus and a list of manifolds mostly accessed in applications. Section 3 introduces the notion of curves and tangent bundles and allied topics, such as normal spaces. Section 4 provides a first introduction, with examples, to first-order dynamical systems on manifolds. Section 5 presents a special kind of derivative of a function having a manifold as a domain and a manifold as a co-domain, termed a pushforward map. Section 6 introduces a class of manifolds that have peculiar features: the Lie groups. Section 7, in detail, treats the fundamental concept of metrization of a curved space through the more familiar notion of metrization of vector spaces. Section 8 introduces notions such as geodesic lines, Riemannian distance and exponential maps. Section 9 surveys the notion of Riemannian gradient and illustrates such a concept via examples. Section 10 introduces and exemplifies the notion of parallel transport along a curve, which is of paramount importance in manifold calculus and its computer-based implementations. Section 11 outlines the concept of manifold retraction and vector transport, which are computationally convenient approximations of exponential maps and parallel transport, respectively. Section 12 illustrates a feedback control theory suitable for application to first-order systems insisting on state manifolds. Section 13 introduces the notion of Riemannian Hessian, which stems from a second-order approximation of a manifold-to-scalar function, and recalls optimization algorithms that extend the Newton method to look for a zero of a vector field. Section 14 concludes this tutorial paper.

As a distinctive aspect of the present tutorial paper, the main flow of discussion is based on *coordinate-free* (or component-free) expressions, which facilitates the imple-

mentation of the main equations by a matrix-friendly computational language (such as MATLAB©). Coordinate-free manifold calculus is introduced by the technical tool of embedded calculus, which stems from embedding a manifold into a larger ambient space where the calculation rules are simpler and more familiar to readers. The starred subsections, namely subsections marked by an asterisk (*), address some specific arguments, related to *coordinate-prone* manifold calculus, which may be skipped by the uninterested readers without detriment to the comprehension of the main flow of this presentation.

The present tutorial paper does not cover a number of subjects, such as the covariant derivation of vector fields, continuous-time second-order dynamical systems arising from a Lagrangian framework nor higher-order discrete-time dynamical systems, nor the key topics related to manifold curvature. These topics will be the subject of a forthcoming tutorial paper.

2. Coordinate-Free Embedded Manifold Calculus

In this section, we recall the relevant notation and fundamental properties of matrix calculus, as well as several examples of manifolds of interest in engineering and applied sciences.

In manifold calculus, both embedded (or extrinsic) and intrinsic coordinates may be accessed. With the aim of elucidating the difference between extrinsic and intrinsic coordinates for manifold elements, let us mention the case of the space $\mathbb{SO}(2)$ of planar rotations. The space $\mathbb{SO}(2)$ is a mono-dimensional manifold; therefore, any element $X \in \mathbb{SO}(2)$ may be pointed to by one *intrinsic* coordinate: a matrix $X \in \mathbb{SO}(2)$ may be represented as

$$X(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}, \tag{1}$$

with $\theta \in [0, 2\pi)$. On the other hand, by *embedding* the space $\mathbb{SO}(2)$ into the space $\mathbb{R}^{2\times2}$ of 2×2 real-valued matrices, it turns out that any element of $\mathbb{SO}(2)$ may be regarded as a two-by-two real-valued matrix

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \tag{2}$$

whose entries must satisfy the four non-linear constrains:

$$\begin{cases} X_{11}^2 + X_{21}^2 = 1, \\ X_{22}^2 + X_{12}^2 = 1, \\ X_{11}X_{12} + X_{21}X_{22} = 0, \\ X_{11}X_{22} - X_{12}X_{21} = 1. \end{cases} \tag{3}$$

The four parameters $X_{11}, X_{12}, X_{21}, X_{22}$ denote embedded or extrinsic coordinates. The orthogonality and normality constraints of the columns of the matrix X may be represented through two coordinate-free constraints in a compact way as

$$X^\top X = I_2 \text{ and } \det(X) = 1, \tag{4}$$

where a superscript $^\top$ denotes matrix transposition and I_2 denotes a 2×2 identity matrix.

Another illustrative example is the ordinary sphere $\mathbb{S}^2$ that may be parametrized as follows:

$$\begin{cases} x = \cos\theta\sin\varphi \\ y = \sin\theta\sin\varphi \\ z = \cos\varphi \end{cases} \tag{5}$$

with parameters $0 \leq \theta < 2\pi$ and $0 \leq \varphi \leq \pi$ (in cartography, the coordinate θ is termed 'latitude', while the coordinate φ is termed 'longitude'.) In general, the dimension of a manifold denotes the minimal number of parameters that are required to individuate

a point on a manifold. While in common speech the ordinary sphere is termed 'three-dimensional', it is indeed a bi-dimensional manifold embedded in a three-dimensional ambient space. The dimension of the manifold is 2; the dimension of the embedding space, instead, is 3.

In general, in engineering and applied mathematics, using intrinsic coordinates is deprecated. As we have just seen, representing an ordinary sphere through intrinsic coordinates is fairly easy, but what about, for example, a hypersphere $\mathbb{S}^9$? Any intrinsic representation would require nine angles and a series of complicated trigonometric expressions. Assuming that a manifold of interest may be embedded into a larger ambient space $\mathbb{A}$, it is way easier to represent one such manifold as a subset of points of an ambient space that meet certain constraints. For example, the sphere $\mathbb{S}^9$ may be embedded in $\mathbb{R}^{10}$ and every point of such a sphere may be represented by an array $x \in \mathbb{R}^{10}$ such that

$$\sum_{i=1}^{10} x_i^2 = x^\top x = 1. \tag{6}$$

Clearly, the ten coordinates x_i, termed *embedded coordinates*, turn out to be unnecessary (nine coordinates would be enough). The existence of an embedding may be evaluated through classical results of manifold calculus, which also state the best dimension of the ambient space $\mathbb{A}$ on the basis of the structural properties of the embedding. Two such results are the *strong Whitney embedding theorem* and the *Nash isometric embedding theorem* for Riemannian manifolds. In the present tutorial, *we are concerned exclusively with manifolds that may be embedded into a linear ambient space*.

The number of extrinsic coordinates is, in general, far larger than the number of intrinsic coordinates; hence, a coordinate-free embedded representation is, in general, redundant. However, as long as computer implementation is concerned, a coordinate-free representation is advantageous because computing languages, such as MATLAB$^©$ and Python, generally deal seamlessly with bi-dimensional arrays. With reference to the sphere, let us examine a simple example that tells us why manifolds need their own calculus.

Example 1. *Take a point $x \in \mathbb{S}^2$ and a direction $v \in \mathbb{R}^3 - \{0\}$, such that $x^\top v = 0$, and let us write the parametric equation of a line departing from x in the direction v as $\ell(t) := x + t\,v$, with $0 \leq t \leq 1$. It is important to point out that none of the points over such a line belong to the sphere, except for the point x. In fact, we have*

$$\ell^\top(t)\ell(t) = (x + t\,v)^\top(x + t\,v) = 1 + t^2\|v\|^2, \tag{7}$$

since v is orthogonal to x. It turns out that $\ell^\top(t)\ell(t) \neq 1$ if $t \neq 0$, hence the assertion. ∎

This simple example tells us that standard constructs, such as straight lines, do not work on curved manifolds, hence the need of a specific calculus to be consistently developed.

2.1. General Notation and Properties

In general, column arrays are denoted by a lower-case letter, while matrices are denoted by an upper-case letter. Manifold elements, for a generic manifold, are denoted as lower-case letters.

A number of matrix functions and factorizations will be invoked in this paper, namely:

*   **Matrix trace**: The trace of a square matrix $M \in \mathbb{R}^{p \times p}$ (namely the sum of its principal-diagonal entries) is denoted by $\mathrm{tr}(M)$. Matrix trace has a cyclic permutation invariance property. For example, given three conformable (i.e., mutually multipliable) matrices A, B, C, it holds that:
    $$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB). \tag{8}$$

- **Matrix square root**: Given a matrix $M \in \mathbb{R}^{p \times p}$, its square root R is the unique matrix such that $R^2 = P$. Not every matrix admits a square root. Special square roots (such as a symmetric square root) will be defined later.
- **Spectral factorization**: Given a matrix $M \in \mathbb{R}^{p \times p}$, let us assume that there exists an orthogonal matrix X (i.e., one such that $X^\top X = I_p$) and a real diagonal matrix D such that $M = XDX^\top$. The expression on the right-hand side denotes the spectral factorization of the matrix M. Such a factorization turns out to be very useful in evaluating matrix polynomials. For example, it holds that

$$
\begin{aligned}
M^4 = (XDX^\top)^4 = (XDX^\top)(XDX^\top)(XDX^\top)(XDX^\top) = \\
XD(X^\top X)D(X^\top X)D(X^\top X)DX^\top = XDDDDX^\top = XD^4X^\top.
\end{aligned}
\tag{9}
$$

  Exponentiating full matrices is cumbersome, while exponentiating diagonal matrices is amiable.
- **Thin QR factorization**: Any matrix $M \in \mathbb{R}^{p \times q}$, with $p \geq q$, may be factored as the product of a $p \times p$ orthogonal matrix Q and a $p \times q$ upper triangular matrix R, namely $M = QR$. In general, a QR factorization is not unique [17]. To remove such kind of indeterminacy, the R-factor may be chosen with strictly positive entries on its main diagonal, so that the factorization is unique.
- **Compact singular value factorization (SVD)**: Compact SVD of a matrix $M \in \mathbb{R}^{p \times q}$ is a matrix factorization of the type $M = ADB^\top$ in which D is square diagonal of size $r \times r$, where $r \leq \min\{p, q\}$ is the rank of M, and has only nonzero singular values. In this variant, A denotes a $p \times r$ matrix and B denotes a $q \times r$ matrix, such that $A^\top A = B^\top B = I_r$ [18].
- **Polar factorization**: Given a real-valued $p \times n$ matrix M, its polar factorization is written as $M = XS$, where X denotes a $p \times n$ matrix such that $X^\top X = I_n$, termed *polar factor*, and S denotes a symmetric positive semidefinite $n \times n$ matrix [19]. The polar factorization of a matrix always exists and, if the matrix is full rank, its polar factor is unique.
- **Matrix exponential**: Given a matrix $M \in \mathbb{R}^{p \times p}$, its matrix exponential is denoted as $\mathrm{Exp}\,(M)$. Matrix exponential is defined via a series as

$$
\mathrm{Exp}\,(M) := \sum_{k=0}^{\infty} \frac{M^k}{k!}.
\tag{10}
$$

  There exist special formulas to compute the matrix exponential via a finite number of operations for special matrices (see, for example, [20] and references therein). For example, for a symmetric matrix $M \in \mathbb{R}^{p \times p}$, that admits a spectral factorization $M = XDX^\top$, it holds that

$$
\mathrm{Exp}\,(M) = X \,\mathrm{Exp}\,(D) X^\top,
\tag{11}
$$

  where $\mathrm{Exp}\,(\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p)) = \mathrm{diag}(e^{\lambda_1}, e^{\lambda_2}, e^{\lambda_3}, \dots, e^{\lambda_p})$.
- **Principal matrix logarithm**: Given a matrix $M \in \mathbb{R}^{p \times p}$, its principal matrix logarithm is denoted as $\mathrm{Log}\,(M)$. Matrix logarithm is defined via a series as

$$
\mathrm{Log}\,(M) := -\sum_{k=1}^{\infty} \frac{(I_p - M)^k}{k}, \text{ defined for } \|I_p - M\| < 1.
\tag{12}
$$

  In the specialized literaure it is possible to find special formulas to compute the principal matrix logarithm via a finite number of operations for special matrices. For example, for a positive-definite matrix $M \in \mathbb{R}^{p \times p}$, which admits a spectral factorization $M = XDX^\top$, it holds that

$$
\mathrm{Log}\,(M) = X \,\mathrm{Log}\,(D) X^\top,
\tag{13}
$$

where $\mathrm{Log}\,(\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p)) = \mathrm{diag}(\log \lambda_1, \log \lambda_2, \log \lambda_3, \ldots, \log \lambda_p)$. Recall that any symmetric, positive-definite matrix has only positive eigenvalues; hence, $\mathrm{Log}\,(D)$ is a real-valued matrix.

The other notation used within this paper is defined at its earliest occurrence. We just recall that the symbol := denotes a definition and that the derivative of a function $F(t)$ with respect to its scalar parameter t is denoted as $\dot{F}$.

2.2. Manifolds and Embedded Manifolds (or Submanifolds)

The formal definition of a smooth manifold is quite convoluted, as it requires notions from mathematical topology [21]. More practically, a manifold may be essentially regarded as a generalization of surfaces in higher dimensions that is endowed with the noticeable property of being locally similar to a flat space. In addition, a manifold may generally be regarded as an abstract mathematical object, not necessarily ready for computation, whereas, in practice, manifolds of interest in applied science and engineering are essentially matrix manifolds.

Let us consider a smooth manifold $\mathbb{M}$ and a point x on it. From an abstract point of view, x is an element of a set $\mathbb{M}$ and does not necessarily carry any intrinsic numerical features. In order to be able to develop calculus on manifolds, to, for instance, compute the directional derivative of a function $f : \mathbb{M} \to \mathbb{R}$, it is convenient to 'coordinatize' a manifold. To this aim, let us take a neighborhood (open set) $U \subset \mathbb{M}$ that contains the point x and a coordinate map $\psi : U \to \mathbb{R}^p$. The size p denotes the minimal number of coordinates that is necessary to specify the location of the point x unequivocally and is taken as the dimension of the manifold. The map ψ needs to be a one-to-one map. In this way, we attach a set of coordinates $\psi(x)$ to the point x. Such a construction establishes a smooth one-to-one correspondence between a point on a manifold and its coordinate point; therefore, the two concepts may be confused and we may safely speak of a point $x \in \mathbb{M}$ when actually speaking of its coordinates.

The above theoretical construction carries a number of practical drawbacks. For instance, notice that in general a manifold cannot be covered by a single coordinate map. Indeed, in general a manifold needs to be covered by a number of neighborhoods U_k, each of which is equipped with a coordinate map $\psi_k : U_k \to \mathbb{R}^p$. The set $\{U_k\}$ thus forms a *basis* for the manifold. Such a basis need not be finite, although it needs to be countable and is termed 'atlas'. In general, the basis neighborhoods U_k may happen to be overlapping one another; hence, the coordinate maps ψ_k need to satisfy the compatibility conditions. Such conditions formalize the natural requirement that there need to be a one-to-one smooth correspondence between any two different coordinate systems insisting on regions of the manifolds belonging to more than one neighborhood. In formal terms, if $U_k \cap U_h \neq \varnothing$, then the 'transition functions' $\psi_k^{-1} \circ \psi_h$ and $\psi_h^{-1} \circ \psi_k$ should possess the structure of diffeomorphisms, namely C^∞ functions endowed with C^∞ inverses.

As mentioned in the introduction, in the present tutorial we are neglecting coordinates in favor of embeddings, except in starred sections that cover coordinate-prone calculations and that may be skipped by uninterested readers. A smooth manifold is by nature a continuous object. Manifolds of interest in applications, described in embedded terms, may be summarized as follows:

- **Hypercube**: The simplest manifold of interest is perhaps the hypercube $\mathbb{R}^p$, which is essentially the set spanned by p real-valued variables (or p-tuples).
- **Hypersphere, oblique manifold, hyperellipsoid**: A hypersphere is represented as $\mathbb{S}^{p-1} := \{x \in \mathbb{R}^p \mid x^\top x = 1\}$ and is the subset of points of the hypercube with unit Euclidean distance from point 0. This is a smooth manifold of dimension $p - 1$ embedded in the hypercube $\mathbb{R}^p$; in fact, with only $p - 1$ coordinates, we can identify unequivocally any point on a sphere. In [21], it is shown how to 'coordinatize' such a manifold through, e.g., the stereographic projection, which requires two coordinate maps applied to two convenient neighborhoods (each including only one 'pole') on the sphere. The special cases are $\mathbb{S}^1$, the unit circle, and $\mathbb{S}^2$, the ordinary sphere. There

exist a number of applications insisting on the hyperspheres $\mathbb{S}^{p-1}$ such as, for instance, blind deconvolution [22,23], data classification [24], adaptive pattern recognition [25] and motion planning, optimization, and verification in robotics and in computational biology [26]. A smooth manifold closely related to the unit hypersphere is the oblique manifold [27], defined as:

$$\mathrm{OB}(p) := \{X \in \mathbb{R}^{p \times p} \mid \mathrm{diag}(X^\top X) = I_p\}, \tag{14}$$

where the operator $\mathrm{diag}(\cdot)$ returns the zero matrix except for the main diagonal that is copied from the main diagonal of its argument. The structure of the oblique manifold may be easily studied on the basis of the unit hypersphere; in fact, the following identification holds true:

$$\mathrm{OB}(p) \cong \underbrace{\mathbb{S}^{p-1} \times \mathbb{S}^{p-1} \times \cdots \times \mathbb{S}^{p-1}}_{p \text{ times}}, \tag{15}$$

hence, each column of a $\mathrm{OB}(p)$-matrix may be treated as a $\mathbb{S}^{p-1}$-column array. The hyperellipsoid L^{p-1} is defined as:

$$\mathrm{L}^{p-1} := \{x \in \mathbb{R}^p \mid \|DRx\|^2 = 1\}, \tag{16}$$

with $D \in \mathbb{R}^{p \times p}$ being diagonal and positive-definite and $R \in \mathbb{R}^{p \times p}$ being a hyper-rotation such that $R^\top R = I_p$ and $\det(R) = 1$. The mathematical structure of the hyperellipsoid L^{p-1} may be studied on the basis of the manifold structure of the hypersphere. The hyperellipsoid L^{p-1} is used, e.g., in the calibration of magnetometers [28].

- **General linear group and special linear group**: The general linear group is defined as $\mathbb{GL}(p) := \{X \in \mathbb{R}^{p \times p} \mid \det(X) \neq 0\}$. This is the subset of the space of $p \times p$ matrices $\mathbb{R}^{p \times p}$ which are invertible. The special linear group is defined as $\mathrm{Sl}(p) := \{X \in \mathbb{R}^{p \times p} \mid \det(X) = 1\}$. This is the subset of the general linear group made by all matrices with a unitary determinant.

- **Orthogonal group, special orthogonal group, special Euclidean group**: An orthogonal group of size p is defined by $\mathbb{O}(p) := \{X \in \mathbb{R}^{p \times p} \mid X^\top X = I_p\}$. The manifold $\mathbb{O}(p)$ has dimension $2^{-1}p(p-1)$. In fact, every matrix in $\mathbb{O}(p)$ possesses p^2 entries which are constrained by $2^{-1}p(p+1)$ orthogonality/normality restrictions. The manifold of special orthogonal matrices is defined as $\mathbb{SO}(p) := \{X \in \mathbb{R}^{p \times p} \mid X^\top X = I_p, \det(X) = 1\}$. A smooth manifold closely related to the special orthogonal group is the special Euclidean group, denoted as $\mathbb{SE}(p)$, that finds applications in robotics (see, e.g., [29]). The special Euclidean group is a set of $(p+1) \times (p+1)$ matrices defined as:

$$\mathbb{SE}(p) := \left\{ \left[\begin{array}{cc} X & \delta \\ 0 & 1 \end{array} \right] \middle| X \in \mathbb{SO}(p), \, \delta \in \mathbb{R}^p \right\}. \tag{17}$$

- **Stiefel manifold**: The (compact) Stiefel manifold is defined as:

$$\mathrm{St}(n, p) := \{X \in \mathbb{R}^{n \times p} \mid X^\top X = I_p\}, \tag{18}$$

where $p \leq n$. Avery Stiefel matrix has np entries, but its elements are constrained by $2^{-1}p(p-1)$ non-linear constraints, and hence the dimension of a Stiefel manifold $\mathrm{St}(n, p)$ is $np - 2^{-1}p(p-1)$. Exemplary applications are blind source separation [30], non-negative matrix factorization [31], best basis search/selection [32], electronic structures computation [33] and factor analysis in psychometrics [34]. A generalization of the Stiefel manifold, to be applied to principal subspace tracking, was studied in the contribution [35]. Such a generalized Stiefel manifold is defined as:

$$\mathrm{St}_B(n, p) = \{X \in \mathbb{R}^{n \times p} \mid X^\top B X = I_p\}, \tag{19}$$

where B denotes any symmetric, positive-definite $n \times n$ matrix. The contribution [35] studied the structure of the tangent bundle of the generalized Stiefel manifold and suggested a computationally convenient calculus for this manifold.

- **Real symplectic group**: The real symplectic group is defined as

$$\mathrm{Sp}(2n) := \{Q \in \mathbb{R}^{2n \times 2n} | Q^\top J Q = J\}, \ J := \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix}, \tag{20}$$

where the symbol I_n denotes again a $n \times n$ identity matrix and the symbol 0_n denotes a whole-zero $n \times n$ matrix. The skew-symmetric matrix J enjoys the following properties: $J^2 = -I_{2n}$, $J^{-1} = J^\top = -J$.

- **Manifold of symmetric, positive-definite (SPD) matrices**: The main features of the space

$$\mathbb{S}^+(p) := \{P \in \mathbb{R}^{p \times p} | P = P^\top, P > 0\} \tag{21}$$

of symmetric, positive-definite matrices was surveyed, e.g., in [36,37]. We recall that a matrix $P \in \mathbb{R}^{p \times p}$ is termed positive definite if for every column array $z \in \mathbb{R}^p - \{0\}$ it holds that $z^\top P z > 0$. A useful property is that the (real) eigenvalues of any SPD matrix are all strictly positive. A related manifold is the space $\mathbb{S}^+(n, p)$ of symmetric *fixed-rank* positive-semidefinite matrices, that may be defined as:

$$\mathbb{S}^+(n, p) := \{X X^\top \mid X \in \mathbb{R}^{n \times p}, \mathrm{rank}(X) = p\}. \tag{22}$$

The growing use of low-rank matrix approximations to retain tractability in large-scale applications boosted extensions of the calculus of positive-definite matrices to their low-rank counterparts [38].

- **Grassmann manifold**: A Grassmann manifold $\mathrm{Gr}(n, p)$ is a set of subspaces of $\mathbb{R}^n$ spanned by p-independent vectors, namely

$$\mathrm{Gr}(n, p) = \{\mathrm{span}(w_1, w_2, w_3, \dots, w_p)\}, \tag{23}$$

with $w_1, w_2, w_3, \dots, w_p \in \mathbb{R}^n$ being a p-tuple of arbitrary and linearly independent n-dimensional arrays. Grassmann manifolds are compact, smooth manifolds and are special cases of more general objects termed *flag manifolds* [39]. A representation of any of such subspaces may be assumed as the equivalence class $[X] = \{XR \mid X \in \mathrm{St}(n, p), \ R \in \mathbb{O}(p)\}$ [33]. In practice, an element $[X]$ of the Grassmann manifold $\mathrm{Gr}(n, p)$ is represented by a matrix in $\mathrm{St}(n, p)$, whose columns span the subspace $[X]$.

It is easy to see that manifolds of interest in applications may exhibit large dimensions. Coordinatizing these manifolds may be inconvenient for practical purposes when their sizes are larger than some units. For this reason, matrix manifolds are treated as submanifolds of $\mathbb{R}^p$ or $\mathbb{R}^{p \times n}$ and their elements are represented by solid arrays.

In general, we shall denote as $\mathbb{A}$ an ambient space that a manifold is embedded into. We shall assume that any ambient space $\mathbb{A}$ of interest in the present tutorial paper is a *Euclidean vector space*, namely, a finite-dimensional vector space over the real numbers endowed with an inner product, that we shall denote as $\langle \cdot, \cdot \rangle^{\mathbb{A}}$.

Example 2. *An elementary example of manifold calculus involving the circle $\mathbb{S}^1$ is the Borsuk–Ulam theorem.*

Theorem 1. *Let $f : \mathbb{S}^1 \to \mathbb{R}$ denote a continuous function. There exist two antipodal points $x, -x \in \mathbb{S}^1$ such that $f(x) = f(-x)$.*

Proof. Let $g(y) := f(y) - f(-y)$ with $y \in \mathbb{S}^1$. Choose an arbitrary point $x_0 \in \mathbb{S}^1$:

- If $g(x_0) = 0$, then $x = x_0$;

- If $g(x_0) \neq 0$, notice that $g(-x_0) = f(-x_0) - f(-(-x_0)) = -(f(x_0) - f(-x_0)) = -g(x_0)$. Since g is continuous and takes a different sign in two different points of its domain, there must exist at least a point $x \in \mathbb{S}^1$ such that $g(x) = 0$.

  The above cases prove the assertion by exhaustion. $\square$

The Borsuk–Ulam theorem has an interesting consequence for the temperature of the Earth. In fact, let us identify the equator with $\mathbb{S}^1$ and the temperature over the equator with the function f, then the Borsuk–Ulam theorem implies that there exist two antipodal points over the equator where the temperature is the same. Indeed, the Borsuk–Ulam theorem holds in every dimension, which implies, for example, that if $\mathbb{S}^2$ represents the Earth's surface and $f : \mathbb{S}^2 \to \mathbb{R}^2$ returns the atmospheric temperature and the Earth's pressure in a point of the Earth surface, then there exist two antipodal points on Earth that exhibit the same temperature-pressure value pair! ∎

Matrix factorizations are characterized by a paramount importance in applied calculations, as illustrated in the following.

Example 3. *Let us recall that every symmetric, positive-definite matrix may be factorized as:*

$$P = RDR^\top, \tag{24}$$

where $P \in \mathbb{S}^+(n)$, $R \in \mathbb{SO}(n)$ and D denotes a diagonal $n \times n$ matrix. In fact, D is the matrix whose in-diagonal entries coincide with the eigenvalues and R is the matrix whose columns coincide with the eigenvectors of P. The factorization (24) simplifies some calculations, for example, it holds that:

$$
\begin{aligned}
P^3 &= PPP = (RDR^\top)(RDR^\top)(RDR^\top) = RD(R^\top R)D(R^\top R)DR^\top \\
&= RDDDR^\top = RD^3 R^\top, \\
\sqrt{P} &= R\sqrt{D}R^\top, \\
\mathrm{Exp}\,(P) &= \sum_{i=0}^{\infty} \frac{P^i}{i!} = \sum_{i=0}^{\infty} \frac{(RDR^\top)^i}{i!} = \sum_{i=0}^{\infty} \frac{RD^i R^\top}{i!} = R\,\mathrm{Exp}\,(D)R^\top, \\
P^{-1} &= RD^{-1}R^\top.
\end{aligned}
\tag{25}
$$

The net result is that convoluted matrix operations may be turned into fairly simple computations that require a finite number of elementary operations. ∎

Manifolds may be classified as *compact* or *non-compact*:

- The manifolds $\mathbb{S}^{n-1} \subset \mathbb{A}$, with $\mathbb{A} := \mathbb{R}^n$, and $\mathbb{SO}(n) \subset \mathbb{A}$, with $\mathbb{A} := \mathbb{R}^{n \times n}$, are compact since there exists a ball $\mathbb{B}(0, r) \subset \mathbb{A}$, with $r < \infty$, that contains them.
- The manifold $\mathbb{S}^+(n)$ is non-compact since no t finite-radius ball exists that contains it.

In practical terms, the elements of a compact manifold are necessarily limited in value, while the elements of a non-compact manifold may take arbitrarily large values.

From a computational point of view, dynamical systems on compact manifolds, even if they turn out to be unstable, do not pose serious implementation problems, while dynamical systems on non-compact manifolds may result in implementational difficulties (and easily cause runtime errors).

3. Smooth Curves, Tangent Vector Fields, Tangent Spaces and Bundle, Normal Spaces

An interesting object we may think of on a smooth manifold $\mathbb{M}$ is a smooth curve $\gamma : [-\epsilon, \epsilon] \to \mathbb{M}$, with $\epsilon > 0$. It is worth remarking that a curve may cross different coordinate charts (U_k, ψ_k); therefore, it is generally necessary to split a curve in as many branches (or segments) as coordinate charts it crosses. The function γ describes a curve on the manifold $\mathbb{M}$ delimited by the endpoints $\gamma(-\epsilon)$ and $\gamma(\epsilon)$.

3.1. Curves and Bundles for Embedded Manifolds

Let us assume that a manifold $\mathbb{M}$ is embedded into an ambient space $\mathbb{A}$ of suitable dimensions (for instance, the sphere $\mathbb{S}^{p-1}$ is embedded in the ambient space $\mathbb{A} := \mathbb{R}^p$). In the following, a number of examples, and a counterexample, are discussed to clarify the important notion of a smooth curve on a manifold.

Example 4. *On a hypersphere $\mathbb{S}^{n-1}$, consider the following function:*

$$\gamma(t) := \frac{x + t\,v}{\sqrt{(x + t\,v)^\top (x + t\,v)}} \tag{26}$$

with $t \in [-\epsilon,\ \epsilon]$ and $x, v \in \mathbb{R}^n$ being arbitrary (but let us avoid the singularity $x + t\,v = 0$). Notice that $\gamma(0) = x/\|x\|$. To verify that such a function γ traces indeed a curve on the hyersphere it suffices to show that, for every $t \in [-\epsilon,\ \epsilon]$ and $x, v \in \mathbb{R}^n$, it holds that $\gamma(t) \in \mathbb{S}^{n-1}$. Indeed,

$$\gamma^\top(t)\gamma(t) = \frac{(x + t\,v)^\top (x + t\,v)}{\left[\sqrt{(x + t\,v)^\top (x + t\,v)}\right]^2} = 1. \tag{27}$$

In addition, consider the following function

$$\gamma(t) := x \cos(\sqrt{v^\top v}\,t) + \frac{v}{\sqrt{v^\top v}} \sin(\sqrt{v^\top v}\,t), \tag{28}$$

with $t \in [-\epsilon,\ \epsilon]$ and with $x \in \mathbb{S}^{n-1}$ and $v \in \mathbb{R}^n - \{0\}$ such that $v^\top x = 0$. (In this case, it turns out that $\gamma(0) = x$.) Similarly to the previous example, in order to prove that γ is a valid curve on the hypersphere, it suffices to show that $\gamma(t) \in \mathbb{S}^{n-1}$ for every $t \in [-\epsilon,\ \epsilon]$:

$$\begin{aligned}
\gamma^\top(t)\gamma(t) &= [x \cos(\|v\|\,t) + \frac{v}{\|v\|} \sin(\|v\|\,t)]^\top [x \cos(\|v\|\,t) + \frac{v}{\|v\|} \sin(\|v\|\,t)] \\
&= (x^\top x)\,\cos^2(\|v\|\,t) + 2\frac{x^\top v}{\|v\|} \sin(\|v\|\,t) \cos(\|v\|\,t) + \frac{v^\top v}{\|v\|^2} \sin^2(\|v\|\,t) \\
&= (1)\cos^2(\|v\|\,t) + 2\frac{(0)}{\|v\|} \sin(\|v\|\,t) \cos(\|v\|\,t) + (1) \sin^2(\|v\|\,t) = 1.
\end{aligned} \tag{29}$$

It is not difficult to evaluate $\gamma'(t)$ and, in particular, to show that $\gamma'(0) = v$.

Let us now consider an example of a curve over the manifold $\mathbb{SO}(n)$ of the hyper-rotations. In particular, let us examine the following curve on the manifold $\mathbb{SO}(2)$:

$$\gamma(t) := \begin{bmatrix} \cos(b\,t) & \sin(b\,t) \\ -\sin(b\,t) & \cos(b\,t) \end{bmatrix} \tag{30}$$

with $t \in [-\epsilon,\ \epsilon]$ and $b \in \mathbb{R}$ being arbitrarily chosen. (Notice that $\gamma(0) = I_2$.) In order to show that function γ represents a curve over the space of planar rotations, it suffices to compute the product $\gamma^\top(t)\gamma(t)$ and to show that it keeps equal to the identity I_2 and that $\det(\gamma(t)) = 1$ for every value of t in its range:

$$\begin{aligned}
\gamma^\top(t)\gamma(t) &= \begin{bmatrix} \cos(b\,t) & \sin(b\,t) \\ -\sin(b\,t) & \cos(b\,t) \end{bmatrix}^\top \begin{bmatrix} \cos(b\,t) & \sin(b\,t) \\ -\sin(b\,t) & \cos(b\,t) \end{bmatrix} \\
&= \begin{bmatrix} \cos(b\,t) & -\sin(b\,t) \\ \sin(b\,t) & \cos(b\,t) \end{bmatrix} \begin{bmatrix} \cos(b\,t) & \sin(b\,t) \\ -\sin(b\,t) & \cos(b\,t) \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\
\det(\gamma(t)) &= \cos^2(b\,t) - (-\sin^2(b\,t)) = 1.
\end{aligned} \tag{31}$$

Further to this, let us prove that the following function represents a curve over the manifold $\mathbb{SO}(n)$:

$$\gamma(t) := X(I_n + t\,H)(I_n - t\,H)^{-1} \tag{32}$$

with $t \in [-\epsilon,\,\epsilon]$, $X \in \mathbb{SO}(n)$ *and* $H^\top = -H \in \mathbb{R}^{n \times n}$. *(Even in this case, it holds that* $\gamma(0) = X$.) *In order to prove such a statement, it is necessary to compute* $\gamma^\top(t)\gamma(t)$ *and to show that it equals* I_n *and that* $\det(\gamma(t)) = 1$ *at any time:*

$$
\begin{aligned}
\gamma^\top(t)\gamma(t) &= [X(I_n + t\,H)(I_n - t\,H)^{-1}]^\top X(I_n + t\,H)(I_n - t\,H)^{-1} \\
&= (I_n - t\,H)^{-\top}(I_n + t\,H)^\top (X^\top X)(I_n + t\,H)(I_n - t\,H)^{-1} \\
&= (I_n + t\,H)^{-1}(I_n + t\,H)^\top (I_n + t\,H)(I_n - t\,H)^{-1} \\
&= (I_n + t\,H)^\top [(I_n + t\,H)^{-1}(I_n + t\,H)](I_n - t\,H)^{-1} \\
&= (I_n + t\,H)^\top (I_n - t\,H)^{-1} \\
&= (I_n - t\,H)(I_n - t\,H)^{-1} = I_n, \\
\det(\gamma(t)) &= \det(X)\det(I_n + t\,H)\det^{-1}(I_n - t\,H) \\
&= \det(X)\det(I_n + t\,H)\det^{-1}((I_n + t\,H)^\top) = 1,
\end{aligned}
\tag{33}
$$

where the superscript $^{-\top}$ *denotes the inverse of the transposed matrix.*

Notice that, in the proof, we have made use of some matrix identities, including the commutativity property $(I_n + t\,H)(I_n - t\,H)^{-1} = (I_n - t\,H)^{-1}(I_n + t\,H)$ *and* $\det(A) = \det(A^\top)$.

As a further example, let us consider the following function, which we shall prove to represent a curve over the manifold of the 2×2 *symmetric, positive-definite matrices:*

$$\gamma(t) := \begin{bmatrix} a\,(t+1)^2 & b\,(t+1)^2 \\ b\,(t+1)^2 & c\,(t+1)^2 \end{bmatrix} \tag{34}$$

with $t \in [-\epsilon,\,\epsilon]$, $0 < \epsilon < 1$, $a > 0$ *and* $ac - b^2 > 0$. *(Notice that* $\gamma(0) = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$.) *To prove the assertion, it suffices to show that* $\gamma^\top(t) = \gamma(t)$ *and that the eigenvalues of* $\gamma(t)$ *are strictly positive for any t in its range. Symmetry is immediately verified. The eigenvalues are* $a\,(t+1)^2 > 0$ *and* $(ac - b^2)(t+1)^4 > 0$.

Let us now consider a counterexample. Define the following function:

$$\gamma(t) := P + t\,Q, \tag{35}$$

with $t \in [-\epsilon,\,\epsilon]$ *and* $P, Q \in \mathbb{S}^+(n)$ *arbitrary. In general, the above function does not represent a curve in* $\mathbb{S}^+(n)$. *To prove such an assertion, it suffices to recall that, in general, a linear combination of two positive-definite matrices does not result in a positive-definite matrix. As a numerical example, consider the following:* $P := \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$ *and* $Q := \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$. *Then,* $\gamma(t) = \begin{bmatrix} 4 + 2t & 1 - t \\ 1 - t & 1 + t \end{bmatrix}$. *It is readily verified that if* $t \in [-4, 4]$, *then* $\gamma(t)$ *is not necessarily positive-definite for every value of t in the range (for example,* $\gamma(-3)$ *is not positive-definite). From this counterexample, we can verify that the manifold* $\mathbb{S}^+(n)$, *as all manifolds just exemplified, is not a linear space with respect to standard matrix operations.* ■

From a system-theoretic perspective, a curve may be thought of as a trajectory generated by a dynamical system whose state space is a smooth manifold. Take, for instance, the case of a little magnetized ball rolling over a metallic spherical surface: no matter how the little ball moves over the surface of the sphere, during its motion the contact point will describe a curve $\gamma(t)$ over the larger sphere from the initial time $t = -\epsilon$ to $t = +\epsilon$. The parameter t may be thought of as a time index, whose progression corresponds to the flowing of time. Thinking of physical objects moving over curved surfaces helps in understanding the main concepts in manifold calculus, even though a certain degree of

abstraction is still necessary, not only because the dimensions of manifolds of interest may be far larger than 3, but also because most manifolds are difficult to visualize in practice.

A noteworthy property of curves is that they are equivalent up to a regular reparameterization.

Comparing the position over a smooth curve in two close-by instants provides the notion of *speed*, namely, how quickly an object is moving along a trajectory and in which direction. In particular, denoting by $\gamma_x : [-\epsilon, \epsilon] \to \mathbb{M}$ a curve such that $\gamma_x(0) = x$, the quantity

$$v_x := \lim_{h \to 0} \frac{\gamma_x(h) - \gamma_x(0)}{h} = \left. \frac{\mathrm{d}\gamma_x(t)}{\mathrm{d}t} \right|_{t=0} = \dot{\gamma}_x(0) \in \mathbb{A}, \tag{36}$$

denotes such a speed, which is represented as a *tangent vector* v_x at the point x on the manifold. Clearly, the vector v_x does not belong to the curved manifold $\mathbb{M}$, whereas it is tangent to it in the point x. (Notice that in this tutorial the word 'vector' is reserved to tangent vectors, while a column-type array—which may represent a location—is termed 'array'.)

Consider every possible smooth curve on a manifold of dimension p passing through the point x and compute the tangent vectors to all these curves in the point x at once. The collection of these tangent vectors span a linear space of dimension p, which is referred to as *tangent space* (or 'tangent plane') to the manifold $\mathbb{M}$ at the point x, and is denoted with

$$T_x\mathbb{M} := \{\dot{\gamma}_x(0) \mid \forall \, \gamma_x\} \subset \mathbb{A}. \tag{37}$$

Since there exist infinitely many curves passing through a given point with the same velocity, one should define a tangent vector as an *equivalence class* of curves passing through such a given point, while being tangent to each other at that point. A tangent plane is hence a collection of such vectors. In addition, we should underline that each tangent space is a subset of the ambient space due to the special nature of $\mathbb{A}$, being both a point set and a vector space.

Taking a point $x \in \mathbb{M}$ and a tangent vector $v_x \in \mathbb{M}$, any pair (x, v_x) is thought of as belonging to an abstract set termed *tangent bundle*, defined as

$$T\mathbb{M} := \{(x, v) \in \mathbb{A}^2 \mid x \in \mathbb{M}, v \in T_x\mathbb{M}\}. \tag{38}$$

In order to facilitate computation, it is useful to introduce the concept of the *normal space* of an embedded manifold in a given point under a chosen metric:

$$N_x\mathbb{M} := \{z \in \mathbb{A} \mid \langle z, v \rangle^{\mathbb{A}} = 0, \ \forall v \in T_x\mathbb{M}\}. \tag{39}$$

The normal space $N_x\mathbb{M}$ represents the orthogonal complement of the tangent space with respect to a Euclidean ambient space $\mathbb{A}$ that the manifold $\mathbb{M}$ is embedded into (in fact, some authors denote normal spaces as $T_x^\perp\mathbb{M}$).

Let us examine the structure of the tangent/normal spaces of a number of smooth manifolds:

- **Hypercube**: Since the space $\mathbb{R}^p$ is linear, when it is embedded into itself, each tangent space coincides to the whole space, namely, for every, $x \in \mathbb{R}^p$ it holds that $T_x\mathbb{R}^p \equiv \mathbb{R}^p$. Since a normal space is the orthogonal complement to a tangent space, we must conclude that $N_x\mathbb{R}^p = \{0\}$.
- **Hypersphere**: At every point $x \in \mathbb{S}^{p-1}$, the tangent space has the structure

$$T_x\mathbb{S}^{p-1} := \{v \in \mathbb{R}^p \mid v^\top x = 0\}. \tag{40}$$

The normal space $N_x \mathbb{S}^{p-1}$ at every point of the hypersphere, which is the orthogonal complement of the tangent space with respect to the ambient space $\mathbb{A} := \mathbb{R}^p$ that the manifold $\mathbb{S}^{p-1}$ is embedded in, has the structure

$$N_x \mathbb{S}^{p-1} := \{\lambda x \mid \lambda \in \mathbb{R}\} \tag{41}$$

in the case that the ambient space is equipped with the Euclidean inner product $\langle v, w \rangle^{\mathbb{A}} := v^\top w$.

- **Special orthogonal group**: The tangent space of the manifold $\mathbb{SO}(p)$ has the structure

$$T_X \mathbb{SO}(p) = \{V \in \mathbb{R}^{p \times p} \mid V^\top X + X^\top V = 0_p\}. \tag{42}$$

This may be proven by differentiating a generic curve $\gamma(t) \in \mathbb{SO}(p)$ passing through X at $t = 0$. Every such curve satisfies the orthogonal-group characteristic equation $\gamma^\top(t)\gamma(t) = I_p$; therefore, after differentiation, one obtains

$$\dot{\gamma}^\top(0)\gamma(0) + \gamma^\top(0)\dot{\gamma}(0) = 0_p. \tag{43}$$

By recalling that the tangent space is formed by velocity vectors $\dot{\gamma}(0)$, the above-mentioned result is readily achieved. Provided the ambient space $\mathbb{A} := \mathbb{R}^{p \times p}$ is endowed with the canonical Euclidean metric $\langle V, W \rangle^{\mathbb{A}} := \text{tr}(V^\top W)$, the normal space at a point X may be defined as

$$N_X \mathbb{SO}(p) = \{N \in \mathbb{R}^{p \times p} \mid \text{tr}(N^\top V) = 0, \ \forall V \in T_X \mathbb{SO}(p)\}. \tag{44}$$

It is easy to convince oneself that every tangent vector $V \in T_X \mathbb{SO}(p)$ may be written as $V = XH$, with H skew-symmetric (i.e., such that $H^\top = -H$); then, any element $N \in N_X \mathbb{SO}(p)$ may be written as $N = XS$, with $S = S^\top$. In fact, the normality condition implies $0 = \text{tr}(V^\top(XS)) = \text{tr}(SV^\top X) = \text{tr}(X^\top V S^\top)$, which is equivalent to $-\text{tr}((V^\top X)S^\top)$; therefore, the normality condition may be recast as $\text{tr}((V^\top X)(S - S^\top)) = 0$. It is hence necessary and sufficient that $S = S^\top$; therefore,

$$N_X \mathbb{SO}(p) = \{XS \mid S^\top = S \in \mathbb{R}^{p \times p}\}. \tag{45}$$

- **Stiefel manifold**: Given a trajectory $[-\epsilon, \ \epsilon] \ni t \mapsto X(t) \in \text{St}(n, p)$, derivation with respect to the parameter t yields $\dot{X}^\top X + X^\top \dot{X} = 0$, which means that the tangent space to the manifold $\text{St}(n, p)$ in a point $X \in \text{St}(n, p)$ has the structure:

$$T_X \text{St}(n, p) = \{V \in \mathbb{R}^{n \times p} \mid V^\top X + X^\top V = 0\}. \tag{46}$$

The normal space has the structure:

$$N_X \text{St}(n, p) = \{XS \mid S \in \mathbb{R}^{p \times p}, \ S^\top - S = 0\}. \tag{47}$$

- **Real symplectic group**: The tangent space associated with the real symplectic group has the structure:

$$T_Q \text{Sp}(2n) = \{V \in \mathbb{R}^{2n \times 2n} \mid V^\top JQ + Q^\top JV = 0_{2n}\}. \tag{48}$$

The tangent spaces and the normal space associated with the real symplectic group may be characterized as follows:

$$\begin{cases} T_Q \text{Sp}(2n) = \{V = QJS \mid S \in \mathbb{R}^{2n \times 2n}, \ S^\top = S\}, \\ N_Q \text{Sp}(2n) = \{N = JQH \mid H \in \mathbb{R}^{2n \times 2n}, \ H^\top = -H\}. \end{cases} \tag{49}$$

- **Space of symmetric, positive-definite matrices**: Given a point $P \in \mathbb{S}^+(p)$, its tangent bundle may be characterized simply by observing that every curve $\gamma : [-\epsilon, \epsilon] \to$

$\mathbb{S}^+(p)$ satisfies $\gamma(t) = \gamma(t)^\top$ and $\gamma(t) > 0$. Only the equality constraint influences the structure of the tangent space; therefore,

$$T_P\mathbb{S}^+(p) = \{S \in \mathbb{R}^{p \times p} \mid S^\top = S\}. \tag{50}$$

Notice that every tangent space is identical to each other as it does not depend on the base point P.

- **Grassmann manifold**: For every element $[X] \in \mathrm{Gr}(n, p)$, the tangent space may be represented as:

$$T_{[X]}\mathrm{Gr}(n, p) = \{V \in \mathbb{R}^{n \times p} | X^\top V = 0\}. \tag{51}$$

A tangent space $T_{[X]}\mathrm{Gr}(n, p)$ may be decomposed as the direct sum of a *horizontal space* and of a *vertical space* at $[X] \in \mathrm{Gr}(n, p)$ [40]. Starting from a point $[X] \in \mathrm{Gr}(n, p)$, moving along a horizontal direction causes a change in subspace, while moving along a vertical direction does not change the subspace $[X]$.

3.2. Vector Fields

Let us now consider a map $v_x : \mathbb{M} \to T\mathbb{M}$ that associates a tangent vector v_x with every point x of a manifold. One such map is termed *tangent vector field* (or simply vector field, for short). The set of all tangent vector fields of a manifold $\mathbb{M}$ is denoted as $\Gamma(\mathbb{M})$. Vector fields are objects of prime importance in manifold calculus. In physical system dynamics, vector fields are associated with, e.g., speed and acceleration.

Example 5. *A vector field on a manifold $\mathbb{M}$ is a function $f : \mathbb{M} \to T\mathbb{M}$ that assigns to every point $x \in \mathbb{M}$ a tangent vector $f(x) \in T_x\mathbb{M}$. A vector field may even depend on the time parameter, in which case it will be denoted as $f(t, x)$ with $f : \mathbb{R} \times \mathbb{M} \to T\mathbb{M}$.*

Let us now consider the hypersphere $\mathbb{S}^7$ embedded in $\mathbb{A} := \mathbb{R}^8$ and the function $f : \mathbb{S}^7 \to \mathbb{R}^8$ defined as:

$$f(x) := (I_8 - xx^\top)Ax, \tag{52}$$

where x indicates a column array of 8 elements and $A \in \mathbb{R}^{8 \times 8}$ is an arbitrary constant matrix. Such a function describes a vector field in $\Gamma(\mathbb{S}^7)$ because it holds that $f(x) \in T_x\mathbb{S}^7$ for every x. In fact,

$$x^\top f(x) = x^\top(Ax - xx^\top Ax) = x^\top Ax - (x^\top x)(x^\top Ax) = 0. \tag{53}$$

Notice that the tangent vector $f(x)$ is inextricably tied to the point x on the manifold. In fact, in general, taking two distinct points $x, y \in \mathbb{M}$, it turns out that $f(x) \notin T_y\mathbb{M}$. It is then sensible to describe a vector field as a set of pairs:

$$(x, f(x)) \in T\mathbb{M}. \tag{54}$$

As a further example, let us consider the manifold $\mathbb{SO}(8)$ embedded in $\mathbb{A} := \mathbb{R}^{8 \times 8}$ and a function $f : \mathbb{R} \times \mathbb{SO}(8) \to \mathbb{R}^{8 \times 8}$:

$$f(t, R) := RAt + R^2BR^\top, \tag{55}$$

where R is a 8×8 orthogonal matrix variable, while A and B are 8×8 skew-symmetric constant matrices (namely, $A^\top = -A$ and $B^\top = -B$). Such a function represents a (time-varying) vector field on the manifold of hyper-rotations in that, for every $R \in \mathbb{SO}(8)$ and $t \in \mathbb{R}$, it holds that $f(t, R) \in T_R\mathbb{SO}(8)$. Such an assertion may be proven as follows

$$\begin{aligned} R^\top f(t, R) + f^\top(t, R)R &= R^\top(RAt + R^2BR^\top) + (RAt + R^2BR^\top)^\top R \\ &= (R^\top R)At + (R^\top R)RBR^\top + A^\top(R^\top R)t + RB^\top R^\top(R^\top R) \\ &= At + RBR^\top - At - RBR^\top \\ &= 0_n. \end{aligned} \tag{56}$$

As a last example, let us take the manifold $\mathbb{S}^+(5)$ embedded in $\mathbb{A} := \mathbb{R}^{5\times 5}$ and a function $f : \mathbb{S}^+(5) \to \mathbb{R}^{5\times 5}$ defined as

$$f(P) := P + P^2 + 3\,P^{-1}, \tag{57}$$

where P is a 5×5 symmetric, positive-definite matrix variable (hence, even $P^2 = PP$ and P^{-1} are symmetric, positive-definite). It turns out that the function f represents a vector field in $\Gamma(\mathbb{S}^+(5))$; in fact, for every $P \in \mathbb{S}^+(5)$, it holds that $f(P) \in T_P \mathbb{S}^+(5)$:

$$\begin{aligned}
f^\top(P) - f(P) &= (P + P^2 + 3\,P^{-1})^\top - (P + P^2 + 3\,P^{-1}) \\
&= P^\top + P^{2\top} + 3\,P^{-\top} - P - P^2 - 3\,P^{-1} \\
&= P + P^2 + 3\,P^{-1} - P - P^2 - 3\,P^{-1} \\
&= 0_n.
\end{aligned} \tag{58}$$

(Notice that, in the proof, only the symmetry of the matrix P and of its powers P^2 and P^{-1} has been made use of.) ∎

As a matter of fact, manifold calculus *primarily deals with two kinds of objects*:

- **Points on a manifold**, denoted as $x \in \mathbb{M} \subset \mathbb{A}$;
- **Tangent vectors**, denoted as $v \in T_x\mathbb{M} \subset \mathbb{A}$.

and occasionally on normal arrays, which are instrumental in calculations. In the case of matrix manifolds, which are of prime importance in applications, the ambient space $\mathbb{A} := \mathbb{R}^{p\times q}$; hence, points x and tangent vectors v are essentially arrays or matrices (either rectangular or square).

3.3. Canonical Curves, Canonical Basis of a Tangent Space*

In intrinsic manifold calculus, the way to regard, e.g., tangent spaces and vector fields is based on differential operators [21]. Formally, if $\mathcal{F}$ denotes a smooth function space, a tangent vector $v \in T_x\mathbb{M}$ is defined in such a way that $v : \mathcal{F} \to \mathbb{R}$; namely, $v(f)$ denotes the directional derivative of the function $f \in \mathcal{F}$ along the direction v. Recall that a differential operator in $T_x\mathbb{M}$ may be written as a linear combination of elementary differential operators, namely the derivatives along the basis axes, through some coefficients.

Let $\mathbb{M}$ denote a smooth manifold of dimension p and let $\gamma_x : [-\epsilon, \epsilon] \to \mathbb{M}$ denote any smooth curve such that $\gamma_x(0) = x$. Let us assume, for simplicity, that the image of γ_x is entirely contained in a chart (U, ψ). The map $\psi \circ \gamma_x$ traces out a smooth curve in the parameter space $\mathbb{R}^p$, which is differentiable in the usual multivariable calculus sense. Let us denote the intrinsic coordinates of the points over the curve as

$$(\theta^1(t), \theta^2(t), \theta^3(t), \ldots, \theta^p(t)) = \psi(\gamma_x(t)). \tag{59}$$

The superscript notation θ^i to denote the ith coordinate is standard in manifold calculus. This notation helps when checking expressions written in intrinsic coordinates. Since the chart ψ is invertible by definition, we may represent the curve γ in local coordinates by

$$\gamma_x(t) = \psi^{-1}(\theta^1(t), \theta^2(t), \theta^3(t), \ldots, \theta^p(t)). \tag{60}$$

In particular, it holds that $x = \psi^{-1}(\theta^1(0), \theta^2(0), \theta^3(0), \ldots, \theta^p(0))$. On the basis of the representation (60), we may define as many as p *canonical curves* as $\gamma_x^i(t) := \psi^{-1}(\theta^1(0), \theta^2(0), \ldots, \theta^i(0) + t, \ldots, \theta^p(0))$ for $t \in [-\epsilon, \epsilon]$ and $i = 1, 2, 3, \ldots, p$. Namely, a canonical curve around a point is traced on a manifold by letting only one of the p coordinates vary at a time. Let us now consider the tangent vectors

$$\partial_i^x := \dot\gamma_x^i(0) = \left.\frac{\mathrm{d}\gamma_x^i(t)}{\mathrm{d}t}\right|_{t=0} \in T_x\mathbb{M}. \tag{61}$$

The set $\{\partial_1^x, \partial_2^x, \ldots, \partial_p^x\}$ is termed *canonical basis* of the vector space $T_x\mathbb{M}$. Its property of being a basis is implied by the fact that such canonical vectors are linearly independent from one another. Therefore, any tangent vector may be written as a linear combination of canonical vectors, that is

$$v_x = v_x^1 \partial_1^x + v_x^2 \partial_2^x + \cdots + v_x^p \partial_p^x, \tag{62}$$

where the quantities v_x^1 represent the coefficients of the linear combination. To shorten the equations, the above expression may be written simply as $v = v^i \partial_i$, where the information about what happens at the point x has been suppressed and the summation is implied by the presence of a repeated index (i), one in upper position (^{i}) and one in lower position ($_i$), which is commonly referred to as *Einstein summation convention*.

Example 6. *Let us consider the case of a sphere $\mathbb{S}^2$ embedded into $\mathbb{R}^3$. A parametrization of $\mathbb{S}^2$ (excluding the South pole) is*

$$x = \psi^{-1}(\theta^1, \theta^2) := \begin{bmatrix} \sin\theta^1 \cos\theta^2 \\ \sin\theta^1 \sin\theta^2 \\ \cos\theta^1 \end{bmatrix} \tag{63}$$

for $0 \le \theta^1 \le \pi$ (where θ^1 is commonly referred to as 'latitude') and $0 \le \theta^2 < 2\pi$ (where θ^2 is commonly referred to as 'longitude'). For example, the point $\psi^{-1}(0,0)$ coincides with the North pole, while the point $\psi^{-1}(\frac{\pi}{2}, 0)$ lays on the equator of the sphere. Given a point $x = \psi^{-1}(\theta_0^1, \theta_0^2)$, the curve

$$\gamma_x^1(t) = \begin{bmatrix} \sin(\theta_0^1 + t) \cos\theta_0^2 \\ \sin(\theta_0^1 + t) \sin\theta_0^2 \\ \cos(\theta_0^1 + t) \end{bmatrix} \tag{64}$$

traces a 'meridian' through x, while the curve

$$\gamma_x^2(t) = \begin{bmatrix} \sin\theta_0^1 \cos(\theta_0^2 + t) \\ \sin\theta_0^1 \sin(\theta_0^2 + t) \\ \cos\theta_0^1 \end{bmatrix} \tag{65}$$

traces a 'parallel' through x. Hence, by definition, the tangent space $T_x\mathbb{S}^2$ is spanned by the canonical basis vectors

$$\partial_1^x = \dot{\gamma}_x^1(0) = \left.\frac{d\gamma_x^1(t)}{dt}\right|_{t=0} = \begin{bmatrix} \cos\theta_0^1 \cos\theta_0^2 \\ \cos\theta_0^1 \sin\theta_0^2 \\ -\sin\theta_0^1 \end{bmatrix} \tag{66}$$

and

$$\partial_2^x = \dot{\gamma}_x^2(0) = \left.\frac{d\gamma_x^2(t)}{dt}\right|_{t=0} = \begin{bmatrix} -\sin\theta_0^1 \sin\theta_0^2 \\ \sin\theta_0^1 \cos\theta_0^2 \\ 0 \end{bmatrix}. \tag{67}$$

For example, it is readily seen that

$$\partial_1^{(\frac{\pi}{2}, 0)} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \ \partial_2^{(\frac{\pi}{2}, 0)} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \tag{68}$$

represents tangent vectors at $[1\ 0\ 0]^\top$ (i.e., the North pole). ∎

The dual space to a tangent space $T_x\mathbb{M}$, denoted as $T_x^\star\mathbb{M}$, is termed *cotangent space*. Elements of a cotangent space are termed *covectors* $\omega \in T_x^*\mathbb{M}$. Covectors combine with vectors by 'annihilation' to a scalar. Given a canonical basis $\{\partial_1^x, \partial_2^x, \ldots, \partial_p^x\}$ of $T_x\mathbb{M}$, it is possible to define a canonical basis $\{dx^1, dx^2, \ldots, dx^p\}$ of $T_x^*\mathbb{M}$ by the conditions $dx^i(\partial_j) =$

δ^i_j. The union of all cotangent spaces form a *cotangent bundle* $T^*\mathbb{M} := \{(x, \omega) \mid x \in \mathbb{M}, \omega \in T^*_x\mathbb{M}\}$.

4. First-Order Dynamical Systems on Manifolds

A number of physical phenomena are described in system theory as systems of coupled differential equations in several variables. A fairly general representation of such systems is

$$\dot{x}(t) = f(t, x(t)), \; x(0) = x_0 \in \mathbb{R}^p, \; t \in [0, t_{\mathrm{f}}], \qquad (69)$$

where $x \in \mathbb{R}^p$ denotes a state variable array and represents a set of descriptive variables, and $f : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^p$ represents a state-transition function. The multi-variable array-type function f denotes a (possibly time-varying) vector field that represents the set of velocities of change of the state of the system. The solutions of the differential Equation (69) represent integral curves of the vector field f. In this case, the space $\mathbb{R}^p$ represents the *flat* state space and the velocity space of the model (69) at once.

In order to take into account further constraints on the state variables, termed *invariants*, it is convenient to introduce the notion of *curved* state space in the form of a smooth manifold. This logical process leads to a type of first-order dynamical system described by

$$\dot{x}(t) = f(t, x(t)), \; x(0) = x_0 \in \mathbb{M}, \; t \in [0, t_{\mathrm{f}}], \qquad (70)$$

where $f : \mathbb{R} \times \mathbb{M} \to T\mathbb{M}$ denotes the state-transition function of the mathematical model and $x \in \mathbb{M}$ denotes the state of the system at any given time t. Even in this case, the function f denotes the velocity field; however, in this case, the state space is the manifold $\mathbb{M}$, while the velocity space is the tangent bundle $T\mathbb{M}$.

Let us underline two aspects of the above equation:

- **Meaning of $\dot{x}(t)$ in the expression** (70): Since we assumed the manifold $\mathbb{M}$ to be embedded in an ambient space $\mathbb{A}$ of type $\mathbb{R}^{p \times q}$, the quantity $\dot{x}(t)$ is an array or a matrix made of the derivatives of the entries of $x(t)$ with respect to the time parameter t. Let us recall, however, that even if such a specification helps one understand the subject, *we shall never write any relation involving single components of such matrix-type objects as we shall treat any state variable x as a whole*, (except in a low-dimensional example). The solution $x(t)$ of the differential Equation (70) is the trajectory of the system and is represented by a curve on the manifold $\mathbb{M}$; hence, $\dot{x}(t)$ denotes the speed along the trajectory since, for every $t \in [0, t_{\mathrm{f}}]$, it holds that $(x(t), \dot{x}(t)) \in T\mathbb{M}$, namely $\dot{x} \in \Gamma(\mathbb{M})$.

- **Separation (or non-equivalency) of first-order and second-order systems**: When dealing with dynamical systems in $\mathbb{R}^{p \times q}$, the system (69) is actually fairly general since, upon introducing additional variables, it is possible to turn an n-th order system into a first-order system. We consider now how such a property stems from the legitimate 'confusion' between the state space and velocity space. In contrast to this, when dealing with dynamical systems on manifolds, such confusion is not legitimate, since $x \in \mathbb{M}$ while $\dot{x} \in T_x\mathbb{M}$; namely, the system state and the system velocity belong to very different spaces. It suffices to recall that $\mathbb{M}$ is generally a curved space (i.e., non linear) while each $T_x\mathbb{M}$ is a vector space (flat, linear). For this reason, second-order systems are not assimilable to first-order systems.

Example 7. *Let us reconsider the function* (52)*. On the basis of such a vector field, we may define the following first-order system on the hypersphere* $\mathbb{S}^{n-1}$:

$$\dot{x}(t) = (I_n - x(t)\, x^\top(t))\, A\, x(t), \; x(0) = x_0 \in \mathbb{S}^{n-1}. \qquad (71)$$

Such a system may be rewritten equivalently as:

$$\dot{x} = A\, x - (x^\top A\, x)x. \qquad (72)$$

Let us examine the equilibrium points of such dynamical system. Any equilibrium point $x_\star$ must satisfy the equation

$$A x_\star = (x_\star^\top A x_\star) x_\star. \tag{73}$$

Whenever A is symmetric and positive-definite, Equation (73) establishes that $x_\star$ is an eigenvector of matrix A. In effect, the state of the system (72) evolves towards an eigenvector of the matrix A. For this reason, the dynamical system (73) represents a prototypical example of a continuous-time calculation system.

The differential Equation (72) is commonly referred to as the Oja equation and was studied by Prof. Erkki Oja from the Helsinki University of Technology (currently Aalto University). It was studied in several contexts, including automation and control [41], and constitutes an exemplification of the fact that even a fairly simple computing element might be able to perform a complex calculations, namely, extracting an eigenvector $x_\star$ (and the corresponding eigenvalue $\lambda_\star := x_\star^\top A x_\star$) from an arbitrarily-sized matrix. ∎

Let us now consider an example, drawn from circuit theory, that introduces the notion of 'invariants' for a dynamical system, which indeed motivates the study of dynamical systems on manifolds.

Example 8. *In order to emphasize the notion of* invariants *in connection to dynamical systems, let us consider the simple model of an ideal DC-to-DC converter studied in [42]. The mathematical model of such converter reads:*

$$\begin{cases} C_1 \frac{dv_1}{dt} = (1-u)i_3, \\ C_2 \frac{dv_2}{dt} = u\,i_3, \\ L_3 \frac{di_3}{dt} = -(1-u)v_1 - u\,v_2, \end{cases} \tag{74}$$

where v_1, v_2 are state voltages (across capacitors), i_3 is a state current (across an inductor) and $u \in \{0,1\}$ denotes the control input (a switch). The following function is an invariant for such an electrical circuit

$$\mathscr{I} := \tfrac{1}{2}C_1 v_1^2 + \tfrac{1}{2}C_2 v_2^2 + \tfrac{1}{2}L_3 i_3^2. \tag{75}$$

Indeed, such a quantity represents the total energy across the electrical network. By invariant it is meant that $\dot{\mathscr{I}} = 0$, namely that the time-function $\mathscr{I}(t)$ keeps the same value for every t. Notice, in fact, that

$$\begin{aligned} \frac{d\mathscr{I}}{dt} &= v_1 C_1 \frac{dv_1}{dt} + v_2 C_2 \frac{dv_2}{dt} + i_3 L_3 \frac{di_3}{dt} \\ &= v_1(1-u)i_3 + v_2\,u\,i_3 + i_3(-(1-u)v_1 - u\,v_2) \\ &= 0. \end{aligned} \tag{76}$$

In order to simplify the analysis of such a dynamical system, let us define the following abstract state variables:

$$x_1 := \sqrt{\frac{C_1}{2}}v_1, \ x_2 := \sqrt{\frac{C_2}{2}}v_2, \ x_3 := \sqrt{\frac{L_3}{2}}i_3. \tag{77}$$

Let us arrange the above state variables into the state array $x := \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. The system (74) may thus be rewritten as $\dot{x} = f(t,x)$, with

$$f\left(t, \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) := \underbrace{\begin{bmatrix} 0 & 0 & \sqrt{\frac{1}{C_1 L_3}}(1-u(t)) \\ 0 & 0 & \sqrt{\frac{1}{C_2 L_3}}u(t) \\ -\sqrt{\frac{1}{C_1 L_3}}(1-u(t)) & -\sqrt{\frac{1}{C_2 L_3}}u(t) & 0 \end{bmatrix}}_{=:H(t)} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{78}$$

In short, $f(t, x) = H(t)x$.

In terms of the new state variables, the invariant may be rewritten as:

$$\mathscr{I}(x) := x_1^2 + x_2^2 + x_3^2. \tag{79}$$

Assuming, for instance, that $\mathscr{I}(x_0) = 1$, Equation (79) establishes that the state x belongs to the sphere $\mathbb{S}^2$. Hence, we may observe that the state space of the DC-to-DC converter is the unit sphere $\mathbb{S}^2$ embedded in $\mathbb{R}^3$.

As a last verification step, let us prove that $f(t, x) = H(t)x \in T_x\mathbb{S}^2$ for every t. It suffices to prove that $(Hx)^\top x = 0$, namely that $x^\top H^\top x = 0$. By definition of H, it is readily found to be a skew-symmetric matrix, namely that for every t, it holds that $H^\top(t) = -H(t)$. By transposing the product $x^\top H^\top x$, we find the sought result; in fact,

$$(x^\top H^\top x)^\top = x^\top H x = -(x^\top H^\top x), \tag{80}$$

and since $x^\top H^\top x$ is a scalar, it must hold that $(x^\top H^\top x)^\top = x^\top H^\top x$, therefore it must be equal to zero. ∎

First-order dynamical systems are characterized by a (possibly varying) vector field that determines its dynamics, as illustrated by the following example.

Example 9. *Let us consider the following first-order dynamical system on the manifold $\mathbb{SO}(3)$ of three-dimensional rotations in space:*

$$\dot{R}(t) = A - R(t)A^\top R(t), \ R(0) = I_3. \tag{81}$$

In this system, the matrix $R \in \mathbb{SO}(3)$ represents the orientation of a moving orthogonal frame attached to a rigid body with respect to an inertial reference frame, while the arbitrary constant matrix $A \in \mathbb{R}^{3 \times 3}$ determines its rotational speed (namely, the orientation of the axis of rotation and the rotational velocity).

The orientation matrix is often termed attitude *for rigid bodies such as drones and satellites:*

- *The instance $R = I_3$ indicates that the object is horizontal with respect to the reference frame;*
- *The instance $R \neq I_3$ indicates that it is necessary to rotate the body-fixed axes to align them to the inertial axes.*

The dynamical system (81) is of the type $\dot{R} = f(R)$. Let us verify that $f(R) := A - RA^\top R$ is a vector field of $T\mathbb{SO}(3)$. To show such a property, it suffices to prove that $f^\top(R)R + R^\top f(R) = 0_3$. This is in fact true:

$$\begin{aligned} (A - RA^\top R)^\top R + R^\top(A - RA^\top R) &= A^\top R - R^\top A(R^\top R) + R^\top A - (R^\top R)A^\top R \\ &= A^\top R - R^\top A + R^\top A - A^\top R \\ &= 0_3. \end{aligned} \tag{82}$$

It is worth noticing that the property $f \in \Gamma(\mathbb{SO}(3))$ holds true even in the case that A is a time-varying matrix field, namely $A = A(t)$. ∎

5. Tangent Maps: Pushforward and Pullback

Let us consider two Riemannian manifolds $\mathbb{M}$ and $\mathbb{N}$. Any smooth function $f : \mathbb{M} \to \mathbb{N}$ transforms a curve in $\mathbb{M}$ into a curve in $\mathbb{N}$. Since both curves are associated with their velocity vector fields, which define their tangent spaces, one might wonder how the function f maps tangent spaces in $\mathbb{M}$ into tangent spaces in $\mathbb{N}$. The answer is materialized as a *pushforward map*.

Let $x \in \mathbb{M}$, then a pushforward map $\mathrm{d}_x f : T_x \mathbb{M} \to T_{f(x)} \mathbb{N}$ is defined such that for every smooth curve $\gamma_x : [-\epsilon, \epsilon] \to \mathbb{M}$, $\gamma_x(0) = x$ it holds that:

$$\left. \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma_x(t)) \right|_{t=0} =: \mathrm{d}_x f(\dot{\gamma}_x(0)). \tag{83}$$

In general, therefore, given a function $f : \mathbb{M} \to \mathbb{N}$ that maps a point x from the manifold $\mathbb{M}$ to the point $f(x)$ on the manifold $\mathbb{N}$, the map $\mathrm{d}_x f$ associates to any tangent vector v belonging to the tangent space $T_x \mathbb{M}$ a tangent vector $\mathrm{d}_x f(v)$ belonging to the tangent space $T_{f(x)} \mathbb{N}$ by 'pushing' such vector to such space.

A pushforward map is indeed a linear approximation of a smooth map on tangent spaces. Any tangent map $\mathrm{d}_x f(v)$ is linear in the argument v. The map $\mathrm{d}_x f$ at a point x represents, in practical terms, the best linear approximation of the function f near x. A pushforward map may be regarded as a generalization of the total derivative of ordinary calculus. (Some authors would denote a pushforward map by an asterisk as in $f_\star$; however, this notation takes up the space of the reference point x hence hindering it.)

Let us consider two special cases of interest.

1. *Pushforward of a manifold-to-scalar function*: The special case that $\mathbb{N} := \mathbb{R}$, namely that f is a manifold-to-scalar function, is particularly important in applications. Such a special case will be covered later since it involves the notion of *Riemannian gradient*.

2. *Pushforward of a matrix-to-matrix function*: This is the case that the smooth manifolds $\mathbb{M}$ and $\mathbb{N}$ are real matrix manifolds embedded in $\mathbb{R}^{p \times p}$. Any smooth function between any such pair of manifolds is of matrix-to-matrix type. Let us assume that the function f is analytic about a point $X_0 \in \mathbb{M}$, namely, that it may be expressed as a polynomial series:

$$f(X) = \sum_{k=0}^{\infty} a_k (X - X_0)^k, \; a_k \in \mathbb{R}. \tag{84}$$

Then, the pushforward map $\mathrm{d}_X f(V)$ in a point $X \in \mathbb{M}$ applied to the tangent direction $V \in T_X \mathbb{M}$ may be expressed as:

$$\mathrm{d}_X f(V) = \sum_{k=1}^{\infty} a_k \sum_{r=0}^{k-1} (X - X_0)^r V (X - X_0)^{k-r-1}. \tag{85}$$

It is easily recognized that the tangent map $\mathrm{d}f_X(V)$ is linear in the argument V. As a reference for the readers, we recall the analytic expansion of three matrix-to-matrix functions, that may be used to compute the corresponding pushforward maps:

- **Matrix exponential**: For the map $f(X) = \mathrm{Exp}(X)$, it holds that $X_0 = 0$, $a_k = (k!)^{-1}$ for $k \geq 0$;
- **Principal matrix logarithm**: For the map $f(X) = \mathrm{Log}(X)$, it holds that $X_0 = I_p$, $a_0 = 0$, $a_k = (-1)^{k+1} k^{-1}$ for $k \geq 1$;
- **Matrix inversion**: For the map $f(X) = X^{-1}$, it holds that $X_0 = I_p$, $a_k = (-1)^k$ for $k \geq 0$.

The inverse of a pushforward map $\mathrm{d}_x f$ is termed *pullback map* and is denoted as $(\mathrm{d}_x f)^{-1}$. A map $(\mathrm{d}_x f)^{-1}(w)$ 'pulls' any tangent vector w from the tangent space $T_{f(x)} \mathbb{N}$ back to the tangent space $T_x \mathbb{M}$.

6. Lie Groups, Lie Algebras, Lie Brackets

Lie groups are hybrid mathematical constructions sharing properties that characterize smooth manifolds and algebraic groups. Let us recall that an *algebraic group* is an algebraic structure $(\mathbb{G}, m, i, e)$ made of a set $\mathbb{G}$, either discrete or continuous, endowed with an internal operation denoted as $m : \mathbb{G} \times \mathbb{G} \to \mathbb{G}$, usually referred to as *group multiplication* (not necessarily a multiplication is the standard sense, though), an inversion operation denoted as $i : \mathbb{G} \to \mathbb{G}$, and an identity element with respect to group multiplication

denoted as e. The multiplication, inversion and identity are related in such a way that, for every triple $x, y, z \in \mathbb{G}$, it holds that:

$$
\begin{aligned}
m(x, e) &= m(e, x) = x \\
m(x, i(x)) &= m(i(x), x) = e, \\
m(x, m(y, z)) &= m(m(x, y), z).
\end{aligned}
\tag{86}
$$

Note that, in general, group multiplication is not commutative, i.e. $m(x, y) \neq m(y, x)$.

Two instances of algebraic groups are $(\mathbb{Z}, +, -, 0)$, a discrete group, and $(\mathbb{GL}(p), \cdot, ^{-1}, I_p)$, a continuous group. The structure $(\mathbb{Z}, +, -, 0)$ represents the set of integer numbers (either positive or negative, also including 0) with the standard addition as a group multiplication, which implies that the inverse is the standard subtraction while the identity is the 0 element. The second example, namely the structure $(\mathbb{GL}(p), \cdot, ^{-1}, I_p)$, represents the subset of non-singular matrices of size $p \times p$, endowed with the standard matrix-to-matrix multiplication '·' as group multiplication. In such a case, the inverse operation coincides with the standard matrix inversion while the group identity coincides with the identity matrix I_p. It is straightforward to show that such group operations/identities satisfy the recalled group axioms. A counterexample of a structure that is not an algebraic group is given by the set of non-negative integer numbers $\mathbb{Z}_0^+$, which does not form a group under standard addition/subtraction (in fact, the subtraction of two positive integers does not necessarily return a positive integer).

With the notions of algebraic groups and smooth manifolds, we may now define a well-known object of manifold calculus, namely a *Lie group*. A Lie group conjugates the properties of an algebraic group and of a smooth manifold, as it is a set endowed with group properties and a manifold structure. Paraphrasing Wirth:

$$\text{Lie group} := \text{Manifold} + \text{Algebraic group.}$$

Let us denote by $T_x\mathbb{G}$ the tangent space of a Lie group $\mathbb{G}$ at a point $x \in \mathbb{G}$. The tangent space at the identity, namely $T_e\mathbb{G}$, represents a special instance of tangent spaces. In fact, such a tangent space, upon being endowed with a binary operator termed *Lie brackets*, possesses the structure of a so-called *Lie algebra* and will be denoted as $\mathfrak{g}$.

Let us examine the structure of the Lie algebra of the following Lie groups:

- **Hypercube**: The hypercube, also known as a translation group, is a Lie group under standard matrix sum (matrix subtraction and zero matrix complete the group structure). The Lie algebra of $\mathbb{R}^p$ coincides with itself.

- **General linear group and the special linear group**: Both the general linear group $\mathbb{GL}(p)$ and the special linear group $\mathrm{Sl}(p)$ are Lie groups under standard matrix multiplication and inversion. The Lie algebra of the general linear group, namely $\mathfrak{gl}(p)$, coincides with $\mathbb{R}^{p \times p}$. The linear algebra associated with the special linear group is more interesting and its determination involves some clever matrix computations. Let us consider $\mathrm{Sl}(p)$ endowed with standard matrix multiplication, inversion and I_p as the group identity and a curve $\gamma : [-\epsilon, \epsilon] \to \mathrm{Sl}(p)$ defined by $\gamma(t) := \mathrm{Exp}(t A)$, where Exp denotes the matrix exponential and $A \in \mathbb{R}^{p \times p}$. Clearly, $\gamma(0) = I_p$; hence, $\dot{\gamma}(0)$ represents any element of the Lie algebra $\mathfrak{sl}(p)$. It is not hard to prove that

$$
\dot{\gamma}(t) = A \, \mathrm{Exp}(t A),
\tag{87}
$$

hence, $\dot{\gamma}(0) = A \in \mathfrak{sl}(p)$. Now, let us recall a result from matrix calculus [43]:

$$
\frac{\mathrm{d}}{\mathrm{d}t} \det(\gamma(t)) \Big|_{t=0} = \mathrm{tr}(\dot{\gamma}(0)).
\tag{88}
$$

In the present case, since $\det(\gamma(t)) \equiv 1$, it follows from the above considerations that $\operatorname{tr}(A) = 0$. In conclusion, we found that

$$\mathfrak{sl}(p) = \{A \in \mathbb{R}^{p \times p} \mid \operatorname{tr}(A) = 0\}, \tag{89}$$

the space of traceless matrices. The algebra $\mathfrak{sl}(p)$ has dimension $p^2 - 1$.

- **Special orthogonal group**: The $\mathbb{SO}(n)$ manifold is a Lie group under standard matrix multiplication. The Lie algebra associated with the special orthogonal group is the set of skew-symmetric matrices $\mathfrak{so}(p) := \{H \in \mathbb{R}^{p \times p} \mid H + H^\top = 0_p\}$. In fact, at the identity it holds that $T_{I_p}\mathbb{SO}(p) = \mathfrak{so}(p)$. The Lie algebra $\mathfrak{so}(p)$ is a vector space of dimension $2^{-1}p(p-1)$.

- **Real symplectic group**: The Lie algebra associated with the real symplectic group may be characterized as follows:

$$\mathfrak{sp}(2n) = \{H = JS \mid S \in \mathbb{R}^{2n \times 2n}, \ S^\top = S\}, \tag{90}$$

where the quantity J denotes again the fundamental skew-symmetric matrix.

- **Manifold of symmetric, positive-definite matrices**: The Lie algebra associated with the Lie group $\mathbb{S}^+(p)$ is the set of $p \times p$ symmetric matrices, namely $\mathfrak{s}^+(p) := \{S \in \mathbb{R}^{p \times p} \mid S^\top = S\}$. *The space of symmetric, positive-definite matrices is not a group under standard matrix multiplication*. We recall from [44] the following group structure $(\mathbb{S}^+(p), m, i, e)$:

  - **Multiplication**: $m(P, Q) := \operatorname{Exp}(\operatorname{Log}P + \operatorname{Log}Q)$, (logarithmic multiplication), with $P, Q \in \mathbb{S}^+(p)$, where 'Log' denotes the principal matrix logarithm;
  - **Identity element**: $e = I_p$ (notice that $\operatorname{Log}I_p = 0_p$);
  - **Inverse**: $i(P) = \operatorname{Exp}(-\operatorname{Log}P)$ (matrix inversion), with $P \in \mathbb{S}^+(p)$ (any symmetric, positive-definite matrix is non-singular).

  It is easy to verify that the proposed instances of m, i, e satisfy the algebraic-group axioms in $\mathbb{S}^+(p)$. Additionally, the logarithmic multiplication on $\mathbb{S}^+(p)$ is compatible with its smooth manifold structure, as the map $(P, Q) \mapsto m(P, i(Q)) = \operatorname{Exp}(\operatorname{Log}P - \operatorname{Log}Q)$ is smooth.

An essential peculiarity of any Lie groups $(\mathbb{G}, m, i, e)$ is that the whole group may always be brought back to a convenient neighborhood of the identity e and the same holds true for every tangent space $T_x\mathbb{G}$, $\forall x \in \mathbb{G}$, which may be brought back to the algebra $\mathfrak{g}$. Let us consider, for instance, a curve $\gamma_x : [-\epsilon, \epsilon] \to \mathbb{G}$ such that $\gamma_x(0) = x$. We may define a new curve

$$\tilde{\gamma}_x(t) := m(\gamma_x(t), i(x)) \tag{91}$$

that has the property $\tilde{\gamma}_x(0) = e$. Conversely, $\gamma_x(t) = m(\tilde{\gamma}_x(t), x)$. This operation closely resembles a translation of a curve into a neighborhood of the group identity.

The above observation leads to defining two Lie-group functions:

- **Right translation**: Defined as a function $R_x : \mathbb{G} \to \mathbb{G}$ by $R_x(y) := m(y, x)$ for every pair $(x, y) \in \mathbb{G}$;
- **Left-translation**: Defined as a function $L_x : \mathbb{G} \to \mathbb{G}$ by $L_x(y) := m(i(x), y)$ for every pair $(x, y) \in \mathbb{G}$;

the distinction between the left translation and right translation is somewhat arbitrary: we chose a rather non-standard one.

Notice that R_x and L_x commute; in fact,

$$R_x(L_x(y)) = m(m(i(x), y), x) = m(i(x), m(y, x)) = L_x(R_x(y)). \tag{92}$$

Let us consider a simple example to get acquainted with the notion of left/right translation.

Example 10. *Let us particularize the notion of left/right translation to the familiar hypercube. In $\mathbb{R}^p$, considered as a group under standard array addition, these functions may be defined as $R_x(y) := y + x$ and $L_x(y) = -x + y$. The composition $R_x(L_x(y))$ returns $R_x(L_x(y)) = L_x(y) + x = -x + y + x$, namely $R_x(L_x(y)) = y$; hence, $R_x \circ L_x = \mathrm{id}_x$. Since R_x and L_x commute, we may say that, in this specific case, the functions R_x and L_x are the inverse of one another. In general, this is not true, though, and the failure of the composition $R_x \circ L_x$ to equal an identity map is caused by the lack of commutativity of a group.* ∎

Now, let us consider the curve $L_x(\gamma_x)$: it crosses the identity e at $t = 0$. Consequently, the pushforward $\mathrm{d}_x L_x$ maps every vector in $T_x \mathbb{G}$ to a vector in $T_e \mathbb{G}$. Namely,

$$\mathrm{d}_x L_x : T_x \mathbb{G} \to \mathfrak{g}. \tag{93}$$

Likewise, taking an arbitrary smooth curve γ_e passing by the identity $e \in \mathbb{G}$ at $t = 0$ and considering the curve $R_x(\gamma_e)$, it is readily seen that the latter crosses the point x at $t = 0$. Consequently, the pushforward $\mathrm{d}_e R_x$ maps every vector in $T_e \mathbb{G}$ to a vector in $T_x \mathbb{G}$. Namely,

$$\mathrm{d}_e R_x : \mathfrak{g} \to T_x \mathbb{G}. \tag{94}$$

Recall that $T_x \mathbb{G}$ and $\mathfrak{g}$ are of equal dimension and that the pushforward map is linear; hence, the pushforward map $\mathrm{d}_x L_x$ allows us to translate a vector belonging to a tangent space of a group to a vector belonging to its algebra (and vice versa through its pullback). This is the reason for which the Lie algebra of a Lie group is sometimes termed the 'generator' of a group.

It is easy to see that, if the structure of $\mathfrak{g}$ is known for a group $\mathbb{G}$, it might be convenient to coordinatize a neighborhood of the identity of $\mathbb{G}$ through elements of the associated algebra with the help of a conveniently selected homeomorphism (namely, a continuous function between topological spaces that has a continuous inverse function). Such a homeomorphism is known in the literature as an *exponential map* and is denoted as exp: $\mathfrak{g} \to \mathbb{G}$. We shall discuss the notion of exponential map for general manifolds in a later section.

As mentioned, a Lie algebra is endowed with a binary operator termed *Lie bracket*, denoted as $[\cdot, \cdot]$. Let us survey its derivation. Let us define the function $\Psi_x : \mathbb{G} \to \mathbb{G}$ termed *inner isomorphism* as

$$\Psi_x(y) := L_x(R_x(y)). \tag{95}$$

The inner isomorphism provides a measure of non-commutativity of a Lie group. It is interesting to notice that the function Ψ_x preserves the inner product of $\mathbb{G}$; in fact, taken two elements $y, z \in \mathbb{G}$, it holds that

$$\begin{aligned}
\Psi_x(m(y,z)) &= m(i(x), m(m(y,z), x)) \\
&= m(i(x), m(m(m(y,x), m(i(x),z)), x)) \\
&= m(m(i(x), m(y,x)), m(i(x), m(z,x))) \\
&= m(\Psi_x(y), \Psi_x(z));
\end{aligned} \tag{96}$$

hence, it represents a Lie-group automorphism.

Now, let us take the differential of $\Psi_x(y)$ with respect to the variable y at $y = e$. To this aim, it suffices to consider a curve $y : [-\epsilon, \epsilon] \to \mathbb{G}$ such that $y(0) = e$ and an element $x \in \mathbb{G}$, and to compute the derivative of $\Psi_x(y(t))$ with respect to t at $t = 0$. The calculation of the derivative gives

$$\frac{\mathrm{d}}{\mathrm{d}t} \Psi_x(y(t)) = (\mathrm{d}_{y(t)} \Psi_x)(\dot{y}(t)) = (\mathrm{d}_{R_x(y(t))} L_x) \circ (\mathrm{d}_{y(t)} R_x)(\dot{y}(t)). \tag{97}$$

Setting $y = e$ and letting $v := \dot{y}(0) \in \mathfrak{g}$ and $\mathrm{Ad}_x := \mathrm{d}_e \Psi_x$ yields

$$\mathrm{Ad}_x(v) := (\mathrm{d}_x L_x) \circ (\mathrm{d}_e R_x)(v). \tag{98}$$

By virtue of the properties (93) and (94), Ad_x is a linear map from $\mathfrak{g}$ to itself, namely an endomorphism in $\mathrm{End}(\mathfrak{g})$. The map Ad_x is termed *adjoint representation of the Lie algebra* $\mathfrak{g}$.

In addition, let us now consider a smooth curve $x : [-\epsilon, \epsilon] \to \mathbb{G}$ such that $x(0) = e$ and $\dot{x}(0) = u \in \mathfrak{g}$ and let us evaluate the map

$$\mathrm{ad}_u := \left. \frac{\mathrm{d}}{\mathrm{d}t} (\mathrm{d}_x L_x) \circ (\mathrm{d}_e R_x) \right|_{t=0}. \tag{99}$$

By the chain rule, one obtains

$$\mathrm{ad}_u := \left. \frac{\mathrm{d}}{\mathrm{d}t} (\mathrm{d}_{x(t)} L_{x(t)}) \right|_{t=0} \circ (\mathrm{d}_e R_e) + (\mathrm{d}_e L_e) \circ \left. \frac{\mathrm{d}}{\mathrm{d}t} (\mathrm{d}_e R_{x(t)}) \right|_{t=0}. \tag{100}$$

This is termed *adjoint operator* and coincides with the *Lie bracket* of two tangent vectors up to sign, namely

$$[u, v] := -\mathrm{ad}_u(v). \tag{101}$$

The Lie bracket provides a measure of non-commutativity of the algebra $\mathfrak{g}$.

Example 11. *Let us consider the Lie group $\mathbb{GL}(n)$ endowed with standard matrix multiplication and let us write down explicitly the inner isomorphism, the adjoint representation and the Lie bracket.*

Since $L_X(Y) = X^{-1}Y$ and $R_X(Y) = YX$, the inner isomorphism reads $\Psi_X(Y) = X^{-1}YX$. Taking a curve $Y(t) \in \mathbb{GL}(n)$ such that $Y(0) = I_n$ and $\dot{Y}(0) = V \in \mathbb{R}^{n \times n}$ yields

$$\mathrm{Ad}_X(V) = \left. \frac{\mathrm{d}}{\mathrm{d}t}(X^{-1}Y(t)X) \right|_{t=0} = X^{-1}VX. \tag{102}$$

To end with, taking a curve $X(t) \in \mathbb{GL}(n)$ such that $X(0) = I_n$ and $\dot{X}(0) = U \in \mathbb{R}^{n \times n}$ yields

$$\mathrm{ad}_U(V) = \left. \frac{\mathrm{d}}{\mathrm{d}t}(X^{-1}(t)VX(t)) \right|_{t=0} = (-UV) + VU = -[U, V]. \tag{103}$$

In the above calculation, we have used a known matrix identity, namely

$$\frac{\mathrm{d}}{\mathrm{d}t}X^{-1}(t) = -X^{-1}(t)\dot{X}(t)X^{-1}(t). \tag{104}$$

In this case, $[A, B] := AB - BA$ denotes the matrix commutator. ∎

7. Metrization, Riemannian Manifolds

A Riemannian manifold is a smooth manifold $\mathbb{M}$ whose tangent bundle is equipped with a smooth family of positive-definite inner products $\mathbb{M} \ni x \mapsto \langle \cdot, \cdot \rangle_x \in \mathbb{R}$. An inner product is locally defined at every point $x \in \mathbb{M}$ of a manifold as a bilinear function from $T_x\mathbb{M} \times T_x\mathbb{M}$ to $\mathbb{R}$. It is important to remark that the inner product acts on two elements of a tangent space to the manifold at a given point x; it therefore depends (smoothly) on the point x. Hence, such an inner product gives rise to a *local* metric. Whenever an inner product $\langle \cdot, \cdot \rangle_x$ does not depend explicitly on the point x, it is termed *uniform*. Given two tangent vectors $v, w \in T_x\mathbb{M}$, their inner products are denoted as $\langle v, w \rangle_x$.

7.1. Coordinate-Free Metrization by Inner Products and Metric Kernels

Let us recall some of the general properties of any inner product. In general, any vector space $\mathbb{V}$ may be endowed with an inner product (also termed scalar product) denoted by $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$. Any inner product has a set of properties:

- For every $u, v \in \mathbb{V}$, it holds that $\langle u, v \rangle = \langle v, u \rangle$,
- For every $u, v, w \in \mathbb{V}$, it holds that $\langle u + w, v \rangle = \langle u, v \rangle + \langle w, v \rangle$,
- For every $u, v \in \mathbb{V}$ and $c \in \mathbb{R}$, it holds that $\langle c \, u, v \rangle = c \langle u, v \rangle$,
- The norm of a vector $v \in \mathbb{V}$ is defined as $\|v\| := \sqrt{\langle v, v \rangle}$,
- An inner product is non-degenerate if and only if $\langle u, v \rangle = 0$ for every $v \in \mathbb{V}$ implies $u = 0$.

Let us consider a simple example about the above notions.

Example 12. *Let us consider the special case $\mathbb{V} := \mathbb{R}^p$ and the inner product:*

$$\langle u, v \rangle = u^\top S \, v, \tag{105}$$

with S being a symmetric, $p \times p$ matrix. If S, in addition to being symmetric, is also positive-definite, then such an inner product is non-degenerate. ∎

It pays to recall the notion of adjoint operator with respect to a metric on a vector space. Let us consider again the vector space $\mathbb{V}$ endowed with an inner product $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ and a linear operator $\omega : \mathbb{V} \to \mathbb{V}$. Ad adjoint operator $\omega^\dagger : \mathbb{V} \to \mathbb{V}$ is one such that, for every $v, w \in \mathbb{V}$, it holds that

$$\langle \omega^\dagger(v), w \rangle = \langle v, \omega(w) \rangle. \tag{106}$$

A *self-adjoint* operator is one such that $\omega^\dagger = \omega$, while an *antiadjoint* operator is one such that $\omega^\dagger = -\omega$. The following observation on self-adjoint operators in quadratic forms is in order.

Observation 1. *Every operator may be decomposed as the sum of a self-adjoint and of an antiadjoint operator; in fact, it holds that*

$$\omega = \tfrac{1}{2}(\omega + \omega^\dagger) + \tfrac{1}{2}(\omega - \omega^\dagger). \tag{107}$$

As far as quadratic forms $\langle \omega(v), v \rangle$ are concerned, only the self-adjoint component plays a role. In fact, it is not hard to show that

$$\langle \omega(v), v \rangle = \langle \tfrac{1}{2}(\omega + \omega^\dagger)(v), v \rangle. \tag{108}$$

This is indeed true for any arbitrary inner product of the type $\langle \omega(v), w \rangle$, in fact, some straightforward algebraic work reveals that

$$\langle \omega(v), w \rangle = \tfrac{1}{2}\{\langle \omega(w + v), w + v \rangle - \langle \omega(w - v), w - v \rangle\}. \tag{109}$$

Since the right-hand side is a combination of quadratic forms, the left-hand side depends only on the self-adjoint part of ω.

In the case of a manifold, at every point, $x \in \mathbb{M}$ corresponds a tangent space $T_x\mathbb{M}$ that may be endowed with an inner product. Therefore, to denote the inner product assigned to the vector space $T_x\mathbb{M}$, we use the notation $\langle \cdot, \cdot \rangle_x$.

Example 13. *Given a Riemannian manifold $(\mathbb{M}, \langle \cdot, \cdot \rangle)$ and two smooth vector fields $f, g \in \Gamma(\mathbb{M})$, we may define the function $h : \mathbb{M} \to \mathbb{R}$ as*

$$h(x) := \langle f(x), g(x) \rangle_x. \tag{110}$$

Notice that the inner product varies smoothly *from one tangent space to another; hence, the function h is regular in $\mathbb{M}$. The function $h : \mathbb{M} \to \mathbb{R}$ represents a scalar field on $\mathbb{M}$.*

As a further example, let us consider an arbitrary curve $\gamma : [-\epsilon, \epsilon] \to \mathbb{M}$ over a manifold $\mathbb{M}$. On the basis of such curve, we can define a function $\varphi : [-\epsilon, \epsilon] \to \mathbb{R}$ as

$$\varphi(t) := \langle f(\gamma(t)), g(\gamma(t)) \rangle_{\gamma(t)}. \tag{111}$$

Notice that the inner product 'accompanies' the point $x = \gamma(t)$ that travels along the curve and involves different tangent spaces. Even the function $\varphi : [-\epsilon, \epsilon] \to \mathbb{R}$ represents a scalar field restricted to the curve γ. ∎

On a Riemannian manifold $\mathbb{M}$, the norm of a vector $v \in T_x\mathbb{M}$ is defined as $\|v\|_x := \sqrt{\langle v, v \rangle_x}$. A specific property of Riemannian manifolds is that the norm is positive-definite; namely $\|v\|_x^2 \geq 0$, where $\|v\|_x = 0$ if an only if $v = 0$.

Choosing the 'right' metric for a given manifold in a given application is one of the most challenging aspects of metrization, especially because it is seldom unclear what 'right' practically means. When in doubt, it is possible to resort to *canonical* metrics. A number of canonical metrics for different manifolds of interest are summarized in the following:

- **Hypercube**: For the space $\mathbb{R}^p$ of (column) arrays, the canonical metric is the Euclidean metric $\langle u, v \rangle_x := u^\top v$, for every $u, v \in \mathbb{R}^p$, while for the space $\mathbb{R}^{p \times q}$ of rectangular matrices, the canonical metric is the Euclidean metric $\langle U, V \rangle_X := \mathrm{tr}(U^\top V)$, for every $U, V \in \mathbb{R}^{p \times q}$. Notice that both these metrics do not depend explicitly on the point that they are calculated at, and hence they are uniform.
- **Hypersphere**: The hypersphere $\mathbb{S}^{p-1}$ embedded into the ambient space $\mathbb{R}^p$ inherits its canonical metric; hence, we shall choose $\langle u, v \rangle_x := u^\top v$ for every $u, v \in T_x\mathbb{S}^{p-1}$. Even this metric is uniform.
- **General linear group and the special linear group**: A metric for the general linear group $\mathbb{GL}(p) \ni A$ is:

$$\langle U, V \rangle_A := \mathrm{tr}((A^{-1}U)^\top (A^{-1}V)), \ \forall U, V \in T_A\mathbb{GL}(p). \tag{112}$$

  Such a metric was popularized, for instance, in [45], in the context of machine learning.
- **Special orthogonal group**: The canonical metric in $\mathbb{SO}(p)$ is defined as $\langle U, V \rangle_X := \mathrm{tr}(U^\top V)$, for any $X \in \mathbb{SO}(p)$ and $U, V \in T_X\mathbb{SO}(p)$. Notice that the norm of a tangent vector $V \in T_X\mathbb{SO}(p)$ is $\|V\|_X = \sqrt{\langle V, V \rangle} = \sqrt{\mathrm{tr}(V^\top V)} =: \|V\|_\mathrm{F}$, known as the *Frobenius norm* [18].
- **Stiefel manifold**: There are two well-known metrics for the Stiefel manifold, namely, the Euclidean metric and the canonical metric.
  1. *Euclidean metric*: A possible metric that the Stiefel manifold may be endowed with is the Euclidean metric, inherited from the embedding of $\mathrm{St}(n, p)$ in $\mathbb{R}^{n \times p}$:

  $$\langle U, V \rangle_X := \mathrm{tr}(U^\top V), \ U, V \in T_X\mathrm{St}(n, p). \tag{113}$$

  2. *Canonical metric*: The Stiefel manifold may be endowed with a second kind of metric, termed 'canonical metric'. The associated inner product reads:

  $$\langle U, V \rangle_X := \mathrm{tr}\left(U^\top \left(I_n - \tfrac{1}{2}XX^\top\right)V\right), \ U, V \in T_X\mathrm{St}(n, p), \tag{114}$$

  which, unlike the Euclidean metric (113), is not uniform over the Stiefel manifold.
- **Real symplectic group**: There exist two known metrics in the scientific literature that were applied to the real symplectic group.
  1. *Khvedelidze–Mladenov metric*: A metric for the real symplectic group $\mathrm{Sp}(2n)$ is:

  $$\langle U, V \rangle_Q := \mathrm{tr}(Q^{-1}UQ^{-1}V), \ \forall U, V \in T_Q\mathrm{Sp}(2n). \tag{115}$$

It is referred to as Khvedelidze–Mladenov metric (or KM metric, for short, [46]). This is an indefinite metric; hence, a manifold endowed with this metric is not Riemannian (it is in fact referred to as a *pseudo-Riemannian* manifold).

2. *Euclidean metric*: A further metric for the real symplectic group $\mathrm{Sp}(2n)$ is:

$$\langle U, V \rangle_Q := \mathrm{tr}((Q^{-1}U)^\top(Q^{-1}V)), \ \forall U, V \in T_Q\mathrm{Sp}(2n). \tag{116}$$

Such a metric is inherited from the embedding of a real symplectic group $\mathrm{Sp}(2n)$ into the space of real invertible matrices $\mathbb{GL}(2n)$ that, in turn, is embedded into the real hypercube $\mathbb{R}^{2n \times 2n}$ and hence inherits its canonical metric.

- **Space of symmetric, positive-definite matrices**: The canonical metric in $\mathbb{S}^+(p)$ is defined as $\langle U, V \rangle_P := \mathrm{tr}(UP^{-1}VP^{-1})$, for any $P \in \mathbb{S}^+(p)$ and $U, V \in T_P\mathbb{S}^+(p)$. Clearly, this is not a uniform metric.
- **Grassmann manifold**: The canonical metric on a Grassmann manifold is

$$\langle U, V \rangle_{[X]} = \mathrm{tr}(U^\top V), \ \forall U, V \in T_{[X]}\mathrm{Gr}(n, p), \tag{117}$$

which corresponds to the Euclidean metric in the Stiefel manifold that is used to represent elements of the Grassmann manifold.

Example 14. *The inner product* (115) *gives rise to a metric which is not positive-definite on the space* $\mathrm{Sp}(2n) \subset \mathbb{GL}(2n)$*. To verify this property, it suffices to evaluate the structure of the squared norm* $\|V\|_Q^2 = \mathrm{tr}((Q^{-1}V)^2)$ *with* $Q \in \mathrm{Sp}(2n)$ *and* $V \in T_Q\mathrm{Sp}(2n)$*. By the structure of the tangent space* $T_Q\mathrm{Sp}(2n)$*, it is known that* $Q^{-1}V = JS$ *with* $S \in \mathbb{R}^{2n \times 2n}$ *symmetric. It holds that:*

$$JS = \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix} \begin{bmatrix} S_1 & S_2 \\ S_2^T & S_3 \end{bmatrix} = \begin{bmatrix} S_2^T & S_3 \\ -S_1 & -S_2 \end{bmatrix},$$

with $S_1, S_3 \in \mathbb{R}^{n \times n}$ *being symmetric and* $S_2 \in \mathbb{R}^{n \times n}$ *being arbitrary. Hence,* $\mathrm{tr}((JS)^2) = 2\mathrm{tr}(S_2^2) - 2\mathrm{tr}(S_1S_3)$*, which has an indefinite sign.* ∎

In general, there exists a canonical metric inherited from the ambient space $\mathbb{A}$ that a manifold $\mathbb{M}$ is embedded in. We know that whenever a manifold is embedded in an ambient space $\mathbb{A}$, the inner product between two tangent vectors may be written as

$$\langle u, w \rangle_x = \langle u, G_x(w) \rangle^{\mathbb{A}}, \tag{118}$$

where $\langle \cdot, \cdot \rangle^{\mathbb{A}}$ denotes an inner product in $\mathbb{A}$. The expression (118) is based on a *metric kernel* $G : T\mathbb{M} \to T\mathbb{M}$. Metric kernels play a prominent role in coordinate-free embedded manifold calculus since a kernel and its derivative determine most of the main functions and maps that we shall encounter in the next sections. We assume that the metric kernel has the following properties:

1. **Linearity**: $G_x(v)$ is linear in v, namely $G_x(v + \alpha\,w) = G_x(v) + \alpha\,G_x(w)$, for every $\alpha \in \mathbb{R}$, $x \in \mathbb{M}$ and $v, w \in T_x\mathbb{M}$;
2. **Symmetry**: G_x is a self-adjoint operator, namely $\langle u, G_x(w) \rangle^{\mathbb{A}} = \langle G_x(u), w \rangle^{\mathbb{A}}$;
3. **Closure with respect to** $T\mathbb{M}$: G_x is an endomorphism of $T_x\mathbb{M}$, namely $G_x(v) \in T_x\mathbb{M}$, for every $x \in \mathbb{M}$ and $v \in T_x\mathbb{M}$;
4. **Invertibility**: G_x is invertible, namely, its inverse G_x^{-1} is well-defined for every $x \in \mathbb{M}$.

In general, a metric kernel is not well-defined in the whole ambient space $\mathbb{A}$. We shall invoke the fact that it is well-defined in $T\mathbb{M}$ and that it might be extended to $\mathbb{M} \times \mathbb{A}$, in fact in the expression $G_x(a)$ one must take $x \in \mathbb{M}$ but it is allowable to take $a \in \mathbb{A}$, since the metric kernel is linear in the argument a. Occasionally, we might need to extend the metric kernel G by an operator $\overline{G} : \mathbb{A}^2 \to \mathbb{A}$ that is defined at least in a neighborhood of $\mathbb{M} \times \mathbb{A}$ and such that $\overline{G} \equiv G$ in $T\mathbb{M}$. (In principle, such an 'extendability' requirement

is not necessary, although it facilitates some computations and certainly clarifies some computational developments.)

Example 15. *Let us consider a hypersphere $\mathbb{S}^{n-1} \subset \mathbb{A}$ where $\mathbb{A} := \mathbb{R}^n$ endowed with a metric $\langle u, w \rangle_x = \langle u, w \rangle^{\mathbb{A}} := u^\top w$. It is readily seen that, in this example, $G_x = \mathrm{id}_x$; therefore, the metric kernel is linear, self-adjoint, invertible and an endomorphism of $T_x \mathbb{S}^{n-1}$.*

Let us further consider the manifold $\mathbb{S}^+(n) \subset \mathbb{A}$ endowed with the metric $\langle U, W \rangle_P := \mathrm{tr}(UP^{-1}WP^{-1})$, where $\mathbb{A} := \mathbb{R}^{n \times n}$ is endowed with the metric $\langle U, V \rangle^{\mathbb{A}} := \mathrm{tr}(U^\top V)$. In this case, the metric kernel does not coincide with the identity, rather:

$$G_P(W) := P^{-1}WP^{-1}. \tag{119}$$

Let us verify that such metric kernel has the four mentioned properties:

1. **Linearity**: *$G_P(W)$ appears to be linear in W; in fact, it holds that $G_P(V + \alpha\,W) = P^{-1}(V + \alpha\,W)P^{-1} = P^{-1}VP^{-1} + \alpha\,P^{-1}WP^{-1}$, for every $\alpha \in \mathbb{R}$, $V, W \in T_P \mathbb{A}$,*

2. **Symmetry**: *G_P is self-adjoint; in fact, it holds that $\langle U, G_P(W) \rangle^{\mathbb{A}} = \mathrm{tr}(U(P^{-1}WP^{-1})) = \mathrm{tr}((P^{-1}UP^{-1})W) = \langle G_P(U), W \rangle^{\mathbb{A}}$, by virtue of the cyclic permutation invariance of the trace operator,*

3. **Closure**: *G_P turns out to be an endomorphism of $T_P \mathbb{S}^+(n)$; in fact, for every matrix $U \in T_P \mathbb{S}^+(n)$, it holds that $G_P(U) = P^{-1}UP^{-1}$ is a symmetric matrix. Hence, it belongs to $T_P \mathbb{S}^+(n)$ (notice that $(P^{-1}UP^{-1})^\top = P^{-\top}U^\top P^{-\top} = P^{-1}UP^{-1}$ because both U and P are symmetric),*

4. **Invertibility**: *G_P is invertible; in fact, if $W = G_P(U) = P^{-1}UP^{-1}$, then $U = PWP =: G_P^{-1}(W)$.*

Notice that G_P is not well-defined in $\mathbb{R}^{n \times n}$ (in fact, if P is taken in $\mathbb{R}^{n \times n}$, P^{-1} does not necessarily exist).

As a last example, let us consider the Stiefel manifold $\mathrm{St}(n, p) \subset \mathbb{A}$ is endowed with the metric $\langle U, W \rangle_X := \mathrm{tr}(U^\top(I_n - \frac{1}{2}XX^\top)W)$, where $\mathbb{A} := \mathbb{R}^{n \times p}$ endowed with $\langle U, V \rangle^{\mathbb{A}} := \mathrm{tr}(U^\top V)$. In this case, the metric kernel $G_X : \mathbb{R}^{n \times p} \to \mathbb{R}^{n \times p}$ reads

$$G_X(W) := (I_n - \tfrac{1}{2}XX^\top)W. \tag{120}$$

Linearity and symmetry are readily proven. Closure may be proven as follows. Take $X \in \mathrm{St}(n, p)$ and $W \in T_X \mathrm{St}(n, p)$, namely $W^\top X + X^\top W = 0_p$. Define $U := (I_n - \frac{1}{2}XX^\top)W$. It now suffices to show that $U \in T_X \mathrm{St}(n, p)$, which holds true; in fact,

$$\begin{aligned} U^\top X + X^\top U &= W^\top X - \tfrac{1}{2}W^\top XX^\top X + X^\top W - \tfrac{1}{2}X^\top XX^\top W \\ &= W^\top X + X^\top W - \tfrac{1}{2}\left(W^\top X + X^\top W\right) \\ &= 0_p. \end{aligned} \tag{121}$$

The invertibility of the matrix kernel follows from the invertibility of the matrix $I_n - \frac{1}{2}XX^\top$. From the Sherman–Morrison–Woodbury (matrix inversion) formula [47]

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}, \tag{122}$$

it follows, upon setting $A := I_n$, $C := -\frac{1}{2}I_p$, $U := X$ and $V := X^\top$, that

$$\left(I_n - \tfrac{1}{2}XX^\top\right)^{-1} = I_n - X(-2\,I_p + X^\top X)^{-1}X^\top = I_n + XX^\top, \tag{123}$$

that exists for every $X \in \mathrm{St}(n, p)$. ∎

Further, let us examine separately and concisely the structure of the metric kernel for the symplectic group endowed with its canonical metric.

Example 16. *Let us consider the symplectic group* $\mathrm{Sp}(2n)$ *endowed with the canonical metric* $\langle V, W \rangle_Q := \mathrm{tr}((Q^{-1}V)^\top(Q^{-1}W))$. *Assume the symplectic group* $\mathrm{Sp}(2n)$ *to be embedded in the ambient space* $\mathbb{A} := \mathbb{R}^{2n \times 2n}$ *endowed with the Euclidean metric* $\langle U, V \rangle^{\mathbb{A}} := \mathrm{tr}(U^\top V)$. *A simple expansion gives*

$$\langle V, W \rangle_Q := \mathrm{tr}(V^\top Q^{-\top}Q^{-1}W), \tag{124}$$

hence the metric kernel reads $G_Q(W) := (Q^{-\top}Q^{-1})W$.

7.2. Covariancy, Contravariancy, Tensors*

The convention on index used so far comes from the nature of different objects in relation to their behavior upon coordinate changes.

Let a smooth manifold $\mathbb{M}$ be a submanifold of dimension p of a Euclidean space and let us fix a curve $\gamma : [-\varepsilon, \varepsilon] \to \mathbb{M}$, described by:

$$\gamma_x(t) = (x^1(t), x^2(t), \ldots, x^p(t)). \tag{125}$$

The subscript $_x$ tells that the components of the curve are expressed through coordinates x. The tangent vector $v := \dot{\gamma}_x(0)$ has the components:

$$v_x = (\dot{x}^1(0), \dot{x}^2(0), \ldots, \dot{x}^p(0)). \tag{126}$$

Now, let us introduce a change of variable $x^i := \kappa^i(\xi^1, \xi^2, \ldots, \xi^p)$, for $i = 1, 2, \ldots, p$, such that

$$\gamma_\xi(t) = (\kappa^1(\xi^1(t), \xi^2(t), \ldots, \xi^p(t)), \ldots, \kappa^p(\xi^1(t), \xi^2(t), \ldots, \xi^p(t))). \tag{127}$$

In the new coordinates, the tangent vector at $t = 0$ is expressed as

$$v_\xi = \left(\frac{\partial \kappa^1}{\partial \xi^i} \dot{\xi}^i(0), \frac{\partial \kappa^2}{\partial \xi^i} \dot{\xi}^i(0), \ldots, \frac{\partial \kappa^p}{\partial \xi^i} \dot{\xi}^i(0) \right) = \left(\frac{\partial x^1}{\partial \xi^i} \dot{\xi}^i(0), \frac{\partial x^2}{\partial \xi^i} \dot{\xi}^i(0), \ldots, \frac{\partial x^p}{\partial \xi^i} \dot{\xi}^i(0) \right). \tag{128}$$

At the heart of intrinsic manifold calculus is the requirement that one such tangent vector stays exactly the same, no matter which coordinates are used. Observe that the components $v_x^j := \dot{x}^j(0)$ and the components $v_\xi^j := \dot{\xi}^j(0)$ are related by

$$v_x^j = \frac{\partial x^j}{\partial \xi^i} v_\xi^i. \tag{129}$$

Such a set of equations prescribes how the components of a tangent vector vary upon a coordinate change. Every object that obeys the above law is termed *contravariant*. Contravariant objects are marked by an *upper index*.

The canonical basis vectors ∂_i do *not* obey the law (129). In fact, from the requirement that $v_x = v_\xi$, it follows that

$$v_x^j \partial_j^x = \frac{\partial x^j}{\partial \xi^i} v_\xi^i \partial_j^x = v_\xi^i \partial_i^\xi, \tag{130}$$

which implies that $\frac{\partial x^j}{\partial \xi^i} \partial_j^x = \partial_i^\xi$. Multiplying both sides by $\frac{\partial \xi^i}{\partial x^k}$ gives

$$\frac{\partial \xi^i}{\partial x^k} \frac{\partial x^j}{\partial \xi^i} \partial_j^x = \frac{\partial \xi^i}{\partial x^k} \partial_i^\xi. \tag{131}$$

Since $\frac{\partial \xi^i}{\partial x^k} \frac{\partial x^j}{\partial \xi^i} = \frac{\partial x^j}{\partial x^k} = \delta_k^j$, we obtain the transformation law

$$\partial_k^x = \frac{\partial \xi^i}{\partial x^k} \partial_i^\xi. \tag{132}$$

Every object that obeys the above law is termed *covariant*. Covariant objects are marked by a *lower index*.

Concerning tangent vectors, we may summarize the above results saying that the components of a tangent vector $v \in T_x\mathbb{M}$ are contravariant, while the basis of a tangent space $T_x\mathbb{M}$ is covariant.

Given two tangent vectors $u, v \in T_x\mathbb{M}$, their inner products may be written as

$$\langle u, v \rangle_x = \langle u^i \partial_i^x, v^j \partial_j^x \rangle_x = u^i v^j \langle \partial_i^x, \partial_j^x \rangle_x, \tag{133}$$

due to the bilinearity of the inner product. Now, define $g_{ij}(x) := \langle \partial_i^x, \partial_j^x \rangle_x$. Then, $\langle u, v \rangle_x = g_{ij}(x) u^i v^j$. The functions g_{ij} are clearly symmetric in their indexes (hence, on a smooth manifold of dimension p, these functions are p^2, but only $2^{-1}p(p-1)$ are independent). Moreover, these functions are covariant in both indexes.

Given the canonical basis $\{dx^1, dx^2, \ldots, dx^p\}$ of a cotangent space $T_x^*\mathbb{M}$, every covector may be written as $\omega = \omega_i dx^i$. It is easy to show that the components of a covector are covariant, while the basis elements of a cotangent space are contravariant.

Vectors and covectors are special cases of more general objects termed *tensors*. A vector is a $(1, 0)$-tensor, while a covector is a $(0, 1)$-tensor. In general, it is possible to construct (p, q)-tensors by an operation termed *tensor product* denoted by $\otimes$. Each component of one such tensor has p upper indexes and q lower indexes. For example, one can construct the metric tensor $g = g_{ij} dx^i \otimes dx^j$, which is a $(0, 2)$-tensor. Its inverse is $g^{-1} = g^{ij} \partial_i \otimes \partial_j$, which is a $(2, 0)$-tensor. Additionally, δ_j^i represents the mixed components of the so-termed *fundamental tensor* $\delta = \delta_j^i \partial_i \otimes dx^j$, which is a $(1, 1)$-tensor.

Example 17. *Let us show in what sense functions $g_{ij}(x)$ are covariant components of a tensor. Given a variable change $x = \kappa(\xi)$, from property* (132) *it follows that*

$$\begin{aligned} g_{ij}(\kappa(\xi)) &= \langle \partial_i^x, \partial_j^x \rangle_{\kappa(\xi)} \\ &= \left\langle \frac{\partial \xi^k}{\partial x^i} \partial_k^\xi, \frac{\partial \xi^h}{\partial x^j} \partial_h^\xi \right\rangle_{\kappa(\xi)} \\ &= \frac{\partial \xi^k}{\partial x^i} \frac{\partial \xi^h}{\partial x^j} \langle \partial_k^\xi, \partial_h^\xi \rangle_{\kappa(\xi)} \\ &= \frac{\partial \xi^k}{\partial x^i} \frac{\partial \xi^h}{\partial x^j} g_{kh}(\xi). \end{aligned} \tag{134}$$

This is the transformation law of a $(0, 2)$-tensor. The point is that the components $g_{ij}(x)$ transform into components $g_{kh}(\xi)$ through a linear expression, whose coefficients depend on the derivatives of one set of coordinates with respect to the other set. ∎

8. Geodesic Arc, Riemannian Distance, Exponential and Logarithmic Map

A geodesic on a smooth manifold may be intuitively looked at in different ways:

- On a general manifold, the concept of geodesic extends the concept of a straight line from a flat space to a curved space. In fact, let us consider a curved manifold $\mathbb{M}$ embedded in an ambient space $\mathbb{R}^p$. Such ambient space contains straight lines in the usual meaning, but the manifold, being curved, hardly accommodates any straight lines. Geodesics are curves that resemble straight lines in that they copy some of their distinguishing features.
- On a metrizable manifold, a geodesic connecting two points is locally defined as the *shortest curve* on the manifold connecting these endpoints. Therefore, once a metric is specified, the equation of the geodesic arises from the minimization of a length functional. Such a definition comes from the observation that any straight line in $\mathbb{R}^p$ is indeed the shortest path connecting any two given points.

- A further distinguishing feature of straight lines is that they are self-parallel, namely, sliding a straight lines infinitesimally along itself returns the same exact lines. Such a concept gives rise to a definition of geodesic which requires to specify the mathematical meaning of 'sliding a piece of line infinitesimally along itself'. The technical argument to access such a definition is covariant derivation, which is not covered in the present tutorial (while it will be covered in a forthcoming review paper).
- Another intuitive interpretation is based on the observation that a geodesic emanating from a point on a manifold coincides with the path followed by a particle sliding on the manifold at a *constant speed*. For a manifold embedded in a larger ambient space and *in special circumstances*, this is equivalent to the requirement that the naïve (or embedded) acceleration of the particle is either zero or perpendicular to the tangent space to the manifold at every point of its trajectory. (In the present tutorial paper, we use the term *naïve acceleration* to distinguish it from *covariant acceleration*, which will only be defined in a subsequent tutorial.)

Starting from the above informal description, we are going to examine in detail the notion of embedded geodesy. In particular, we shall treat the problem according to an energy-minimizing principle (which is in fact equivalent to a length-minimizing principle).

8.1. Coordinate-Free Embedded Geodesy

The length of a smooth curve $\gamma : [-\epsilon, \epsilon] \to \mathbb{M}$ may be evaluated though a rectification formula as:

$$\mathrm{L}_\gamma := \int_{-\epsilon}^{\epsilon} \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} \mathrm{d}t. \tag{135}$$

The net result of this argument is that, through the definition of an inner product on the tangent bundle to a Riemannian manifold, we are able to measure the lengths of paths in the manifold itself, which turns the manifold into a *metric space*.

Example 18. *Let us consider the manifold described in intrinsic coordinates:*

$$\mathbb{M} := \{(\theta^1, \theta^2, \sin(\theta^1) + 1) \in \mathbb{R}^3 \mid (\theta^1, \theta^2) \in \mathbb{R}^2\}. \tag{136}$$

Let us consider the curve $(\theta^1, \theta^2) = (t, t)$ obtained by mapping a segment of the plane $\mathbb{R}^2$ to the manifold $\mathbb{M}$, namely:

$$\gamma(t) := (t, t, \sin(t) + 1), \ t \in [a, b]. \tag{137}$$

Let us assume the manifold $\mathbb{M}$ to be endowed with the metric $\langle v, w \rangle_x := v^\top w$. In order to evaluate the length of such a curve, it is necessary to evaluate the velocity field $\dot{\gamma}$:

$$\dot{\gamma}(t) = (1, 1, \cos(t)), \ t \in [a, b]. \tag{138}$$

The length of the curve is now obtained through the rectification formula:

$$\mathrm{L}(\gamma) := \int_a^b \sqrt{1^2 + 1^2 + \cos^2(t)} \mathrm{d}t = \int_a^b \sqrt{2 + \cos^2(t)} \mathrm{d}t. \tag{139}$$

Such an integral may not be expressed in terms of elementary functions. ∎

On the basis of the concepts of 'curve' and of 'length of a curve', it is possible to define the notion of 'distance between two points on a Riemannian manifold' $d(\cdot, \cdot)$.

To define a notion of distance, let us consider what follows:

1. Given a manifold $(\mathbb{M}, \langle \cdot, \cdot \rangle)$, take two arbitrary points $p, q \in \mathbb{M}$;
2. Choose a curve $\gamma : [0, 1] \to \mathbb{M}$ that has p and q as endpoints, namely, such that $\gamma(0) = p$ and $\gamma(1) = q$;
3. Take as as distance $d(p, q)$ the length of such a curve, namely $\mathrm{L}(\gamma)$; one such definition seems a good starting point, but needs to be perfected since there exist infinitely many curves joining two given points.

It is important to underline that *the notion of distance is not univocally defined since it depends on the inner product that a manifold is endowed with*. Since there exist infinitely many curves joining two points on a manifold, a distance between two points p and q is actually defined as the length of the shortest curve connecting such two points, namely

$$d(p,q) := \inf_{\gamma} \int_0^1 \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} \, dt. \tag{140}$$

The problem of selecting the shortest path connecting two points is far from easy to solve. A possible method to look for it is the so-called *variational method* that we are going to introduce in the following.

The key point is to introduce an *energy functional* defined as

$$\mathscr{E}(\gamma) := \int_0^1 \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)} \, dt, \tag{141}$$

whose minimum argument coincides to a geodesic. The variational method consists of looking for a curve that makes the energy functional stationary, namely, for a curve such that $\delta \mathscr{E}(\gamma) = 0$. A concept that facilitates the formulation of the variational method is that of *normal space* to a manifold in a point.

Clearly, the above definition introduces yet another degree of freedom, since the definition of normal space is based on a choice of inner product $\langle \cdot, \cdot \rangle^{\mathbb{A}}$ to evaluate orthogonality. It is instructive to notice that, since a normal space is defined as the orthogonal complement to a tangent space with respect to the ambient space, then

$$T_x \mathbb{M} \oplus N_x \mathbb{M} = \mathbb{A}. \tag{142}$$

In other terms, every element in $\mathbb{A}$ may be decomposed exactly as the sum of a tangent and a normal vector.

Example 19. *Let us evaluate the structure of the normal spaces to the hypersphere $\mathbb{S}^{n-1}$ embedded in $\mathbb{A} := \mathbb{R}^n$. Let us recall that $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n \mid x^\top x = 1\}$ and that $T_x \mathbb{S}^{n-1} = \{v \in \mathbb{R}^n \mid v^\top x = 0\}$. Let us select $\langle y, z \rangle^{\mathbb{A}} := y^\top z$ for every $y, z \in \mathbb{A}$. Since, by definition,*

$$N_x \mathbb{S}^{n-1} := \{z \in \mathbb{R}^n \mid z^\top v = 0, \, \forall v \in \mathbb{R}^n \text{ such that } v^\top x = 0\}, \tag{143}$$

it is easily found that

$$N_x \mathbb{S}^{n-1} = \{\lambda x \mid \lambda \in \mathbb{R}\}. \tag{144}$$

Let us further determine the structure of the normal spaces for the manifold $\mathbb{SO}(n)$. Let us recall that $\mathbb{SO}(n) := \{R \in \mathbb{R}^{n \times n} \mid R^\top R = I_n, \, \det(R) = 1\}$ and that $T_R \mathbb{SO}(n) = \{V \in \mathbb{R}^{n \times n} \mid V^\top R + R^\top V = 0\}$. Let us take $\mathbb{A} := \mathbb{R}^{n \times n}$ and $\langle Y, Z \rangle^{\mathbb{A}} := \mathrm{tr}(Y^\top Z)$ for $Y, Z \in \mathbb{A}$. Let us also recall that

$$V^\top R + R^\top V = (R^\top V)^\top + R^\top V = H^\top + H = 0_n, \tag{145}$$

where $H := R^\top V$. Hence, any tangent vector at R may be written as $V = RH$ with H being skew-symmetric. By definition,

$$N_R \mathbb{SO}(n) := \{Z \in \mathbb{R}^{n \times n} \mid \mathrm{tr}(Z^\top V) = 0, \, \forall V = RH, \, H^\top + H = 0_n\}, \tag{146}$$

and hence it is readily found that

$$N_R \mathbb{SO}(n) = \{RS \mid S \in \mathbb{R}^{n \times n}, \, S^\top = S\}. \tag{147}$$

As a verification step, let us show that, for every $n \times n$ special orthogonal matrix R, skew-symmetric matrix H and symmetric matrix S, it holds that $\langle RH, RS \rangle^{\mathbb{A}} = 0$. To this aim, let us show that:

$$\langle RH, RS \rangle^{\mathbb{A}} = \mathrm{tr}((RH)^\top RS) = \mathrm{tr}(H^\top (R^\top R)S) = -\mathrm{tr}(HS). \tag{148}$$

Now, it is not hard to prove that every skew-symmetric matrix H is 'orthogonal' to every symmetric matrix S; in fact:

$$\underbrace{\mathrm{tr}(HS) = \mathrm{tr}((HS)^\top)}_{Because\ \mathrm{tr}(X)=\mathrm{tr}(X^\top)} = \mathrm{tr}(S^\top H^\top) = \underbrace{\mathrm{tr}(S(-H)) = \mathrm{tr}((-H)S)}_{Because\ \mathrm{tr}(XY)=\mathrm{tr}(YX)} = -\mathrm{tr}(HS), \qquad (149)$$

and therefore $\mathrm{tr}(HS) = 0$. From this result it follows that $\langle RH, RS \rangle^{\mathbb{A}} = 0$.

As a last example, let us determine the structure of the normal space to the manifold $\mathbb{S}^+(n)$. Recall that $\mathbb{S}^+(n) := \{P \in \mathbb{R}^{n \times n} \mid P^\top = P,\ P > 0\}$ and that $T_P\mathbb{S}^+(n) = \{V \in \mathbb{R}^{n \times n} \mid V^\top - V = 0\}$. In this case, choose $\mathbb{A} := \mathbb{R}^{n \times n}$ and $\langle Y, Z \rangle^{\mathbb{A}} := \mathrm{tr}(Y^\top Z)$ for every $Y, Z \in \mathbb{A}$. By definition, it holds that

$$N_P\mathbb{S}^+(n) := \{Z \in \mathbb{R}^{n \times n} \mid \mathrm{tr}(Z^\top V) = 0,\ \forall V \text{ such that } V^\top - V = 0\}, \qquad (150)$$

from which it turns out that

$$N_P\mathbb{S}^+(n) = \{Z \in \mathbb{R}^{n \times n} \mid Z^\top + Z = 0\}. \qquad (151)$$

In fact, $N_P\mathbb{S}^+(n)$ represents the set of $n \times n$ matrices that are orthogonal to all $n \times n$ symmetric matrices, which coincides with the set of all $n \times n$ skew-symmetric matrices. Notice that $N_P\mathbb{S}^+(n)$ does not depend explicitly on the base point P. ∎

The notion of normal space affords formulating the variational problem $\delta\mathscr{E}(\gamma) = 0$ as a *differential inclusion*. For every specific manifold, it will then be possible to turn such a differential inclusion into a differential equation to be solved under appropriate initial conditions. Let us consider a curve $\gamma : [0,\ 1] \to \mathbb{M}$ and let us define the *fundamental form* $\mathscr{F} : T\mathbb{M} \to \mathbb{R}$ as:

$$\mathscr{F}(x,v) := \langle v, v \rangle_x. \qquad (152)$$

The following result holds.

Theorem 2. *For a manifold $\mathbb{M}$, the variational equation $\delta\mathscr{E}(\gamma) = 0$ is equivalent to the following differential inclusion:*

$$\frac{\partial\mathscr{F}}{\partial\gamma} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathscr{F}}{\partial\dot{\gamma}} \in N_\gamma\mathbb{M}. \qquad (153)$$

Notice that such a relation appears as a generalized form of the well-known Euler–Lagrange equation of rational mechanics [48].

Proof. In order to prove such an important result, notice that the energy functional may be rewritten as:

$$\mathscr{E}(\gamma) = \int_0^1 \mathscr{F}(\gamma, \dot{\gamma})\mathrm{d}t. \qquad (154)$$

It is important to notice that, by virtue of the property (118), for every $(x, v) \in T\mathbb{M}$, there exists a neighborhood of (x, v) in $\mathbb{A}^2$, where the partial derivatives $\frac{\partial\mathscr{F}(x,v)}{\partial x}$ and $\frac{\partial\mathscr{F}(x,v)}{\partial v}$ exist (technically, one should use an extension $\overline{\mathscr{F}}$ in place of $\mathscr{F}$). Such partial derivatives are intended in the Gateaux sense, namely

$$\begin{aligned}
\lim_{\epsilon \to 0} \frac{\mathscr{F}(x + \epsilon u, v) - \mathscr{F}(x, v)}{\epsilon} &=: \left\langle \frac{\partial\mathscr{F}(x,v)}{\partial x}, u \right\rangle^{\mathbb{A}}, \text{ for every } u \in \mathbb{A}, \\
\lim_{\epsilon \to 0} \frac{\mathscr{F}(x, v + \epsilon u) - \mathscr{F}(x, v)}{\epsilon} &=: \left\langle \frac{\partial\mathscr{F}(x,v)}{\partial v}, u \right\rangle^{\mathbb{A}}, \text{ for every } u \in \mathbb{A}.
\end{aligned} \qquad (155)$$

Let us define an arbitrary smooth variation $\eta : [0,\ 1] \to T\mathbb{M}$, namely a deviation from the geodesic γ, such that $\eta(0) = \eta(1) = 0$.

For a sufficiently small constant $\epsilon > 0$, it is possible to evaluate the quantity $\mathcal{F}(\gamma + \epsilon\eta, \dot{\gamma} + \epsilon\dot{\eta})$. Applying the Taylor series expansion in $\mathbb{A}^2$ gives:

$$\mathcal{F}(\gamma + \epsilon\eta, \dot{\gamma} + \epsilon\dot{\eta}) = \mathcal{F}(\gamma, \dot{\gamma}) + \epsilon \left\langle \frac{\partial\mathcal{F}}{\partial\gamma}, \eta \right\rangle^{\mathbb{A}} + \epsilon \left\langle \frac{\partial\mathcal{F}}{\partial\dot{\gamma}}, \dot{\eta} \right\rangle^{\mathbb{A}} + \mathcal{O}(\epsilon^2), \quad (156)$$

where $\mathcal{O}$ denotes a Landau symbol and it is used to compactly represent the remainder of the series (e.g., in Lagrange form). The variation of the energy of a curve may be concretely defined as:

$$\delta\mathcal{E}(\gamma) := \lim_{\epsilon \to 0} \frac{\mathcal{E}(\gamma + \epsilon\eta) - \mathcal{E}(\gamma)}{\epsilon} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_0^1 [\mathcal{F}(\gamma + \epsilon\eta, \dot{\gamma} + \epsilon\dot{\eta}) - \mathcal{F}(\gamma, \dot{\gamma})] \mathrm{d}t$$
$$= \int_0^1 \left\langle \frac{\partial\mathcal{F}}{\partial\gamma}, \eta \right\rangle^{\mathbb{A}} \mathrm{d}t + \int_0^1 \left\langle \frac{\partial\mathcal{F}}{\partial\dot{\gamma}}, \dot{\eta} \right\rangle^{\mathbb{A}} \mathrm{d}t + \int_0^1 \lim_{\epsilon \to 0} \frac{\mathcal{O}(\epsilon^2)}{\epsilon} \mathrm{d}t. \quad (157)$$

The last integral is null. The second last integral may be rewritten, upon integration by parts, as:

$$\int_0^1 \left\langle \frac{\partial\mathcal{F}}{\partial\dot{\gamma}}, \dot{\eta} \right\rangle^{\mathbb{A}} \mathrm{d}t = \left\langle \frac{\partial\mathcal{F}}{\partial\dot{\gamma}}, \eta \right\rangle^{\mathbb{A}} \Big|_0^1 - \int_0^1 \left\langle \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial\mathcal{F}}{\partial\dot{\gamma}}, \eta \right\rangle^{\mathbb{A}} \mathrm{d}t. \quad (158)$$

The first addendum on the right-hand side is null because the variation η vanishes at the endpoints of the curve; therefore,

$$\delta\mathcal{E}(\gamma) = \int_0^1 \left\langle \frac{\partial\mathcal{F}}{\partial\gamma} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial\mathcal{F}}{\partial\dot{\gamma}}, \eta \right\rangle^{\mathbb{A}} \mathrm{d}t. \quad (159)$$

By imposing that $\delta\mathcal{E}(\gamma) = 0$ it is readily obtained that, since the variation η is arbitrary, it must hold the relation (153). $\square$

Let us apply the above theorem to a number of manifolds.

Example 20. *Let us determine the equation of geodesics on the hypersphere $\mathbb{S}^{n-1}$ through Equation (153). Let us recall that $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n \mid x^\top x = 1\}$; assume that $\langle v, w \rangle_x := v^\top w$ (independently of x) and let us recall that $N_x\mathbb{S}^{n-1} = \{\lambda x \mid \lambda \in \mathbb{R}\}$. Now, the equation of the geodesic reads:*

$$\frac{\partial(\dot{\gamma}^\top \dot{\gamma})}{\partial\gamma} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial(\dot{\gamma}^\top \dot{\gamma})}{\partial\dot{\gamma}} = 0 - 2\frac{\mathrm{d}\dot{\gamma}}{\mathrm{d}t} = -2\ddot{\gamma} \in \{\lambda\gamma \mid \lambda \in \mathbb{R}\} \quad (160)$$

In summary, the equation for the geodesic is

$$\ddot{\gamma}(t) = \lambda\gamma(t). \quad (161)$$

(Since λ is arbitrary, it absorbed the factor -2.) The geodesic equation is then a second-order differential equation in γ that needs two conditions to be solved.

As a further example, let us derive the geodesic equation for the manifold $\mathbb{SO}(n)$. Let us recall that $\mathbb{SO}(n) := \{R \in \mathbb{R}^{n \times n} \mid R^\top R = I_n, \det(R) = 1\}$; the canonical inner product in this space is $\langle V, W \rangle_R := \mathrm{tr}(V^\top W)$ (independently of R) and that $N_R\mathbb{SO}(n) = \{RS \mid S \in \mathbb{R}^{n \times n}, S = S^\top\}$. According to the general principle, the geodesic equation reads as

$$\frac{\partial(\dot{\gamma}^\top \dot{\gamma})}{\partial\gamma} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial(\dot{\gamma}^\top \dot{\gamma})}{\partial\dot{\gamma}} = 0 - 2\frac{\mathrm{d}\dot{\gamma}}{\mathrm{d}t} = -2\ddot{\gamma} \in \{\gamma S \mid S \in \mathbb{R}^{n \times n}, S = S^\top\} \quad (162)$$

In summary, the geodesic equation on the space $\mathbb{SO}(n)$ reads as

$$\ddot{\gamma} = \gamma S, \quad (163)$$

with S denoting an arbitrary symmetric function. ∎

Through the notion of geodesic arcs, one may define the notion of Riemannian (or geodesic) distance between two nearby points in a smooth manifold. In fact, let us denote by $x, y \in \mathbb{M}$ two nearby points and assume that there exists a unique geodesic arc $\gamma_x^y :$ $[0, 1] \to \mathbb{M}$ such that $\gamma_x^y(0) = x$ and $\gamma_x^y(1) = y$. By definition, the length of such geodesic arcs is taken as the distance between its two endpoints, namely:

$$d(x,y) := \int_0^1 \sqrt{\langle \dot{\gamma}_x^y(t), \dot{\gamma}_x^y(t) \rangle_{\gamma_x^y(t)}} \, \mathrm{d}t. \tag{164}$$

It is worth noticing, at this point, that the expression of the geodesic distance may be rewritten in terms of the fundamental form $\mathscr{F}$, namely

$$d(x,y) = \int_0^1 \sqrt{\mathscr{F}(\gamma_x^y, \dot{\gamma}_x^y)} \, \mathrm{d}t. \tag{165}$$

A noteworthy (and extremely useful) result that we are going to prove is that, along a geodesic, the fundamental form keeps constant. We shall see after the proof that such a result noticeably simplifies the computation of the geodesic distance.

Theorem 3. *Along a geodesic arc* $\gamma : [0, 1] \to \mathbb{M}$*, the function* $\mathscr{F}(\gamma(t), \dot{\gamma}(t))$ *stays constant with respect to t.*

Proof. Let us mention that, for any given pair $(x, v) \in T\mathbb{M}$, it holds that

$$\left\langle \frac{\partial \mathscr{F}(x,v)}{\partial v}, v \right\rangle^{\mathbb{A}} = 2\, \mathscr{F}(x,v), \tag{166}$$

which clearly stems from the fact that $\mathscr{F}(x, v)$ is quadratic in v. Such a property may be shown by recalling that $\mathscr{F}(x, v) = \langle v, G_x(v) \rangle^{\mathbb{A}}$, where G denotes again a metric kernel, and by noticing that

$$\frac{\partial \mathscr{F}(x,v)}{\partial v} = 2\, G_x(v), \tag{167}$$

from which it follows that

$$\left\langle \frac{\partial \mathscr{F}(x,v)}{\partial v}, v \right\rangle^{\mathbb{A}} = \langle 2\, G_x(v), v \rangle^{\mathbb{A}}, \tag{168}$$

hence the property (166). Let us now show that the differential inclusion (153) and the property (166) imply the constancy of the fundamental form $\mathscr{F}$ along a geodesic, namely that:

$$\mathscr{F}(\gamma(t), \dot{\gamma}(t)) = \mathscr{F}(\gamma(0), \dot{\gamma}(0)) \text{ for every value } t \in [0, 1] \tag{169}$$

on every geodesic. By multivariable calculus it is readily proven that:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathscr{F}(\gamma, \dot{\gamma}) = \left\langle \frac{\partial \mathscr{F}}{\partial \gamma}, \frac{\mathrm{d}\gamma}{\mathrm{d}t} \right\rangle^{\mathbb{A}} + \left\langle \frac{\partial \mathscr{F}}{\partial \dot{\gamma}}, \frac{\mathrm{d}\dot{\gamma}}{\mathrm{d}t} \right\rangle^{\mathbb{A}}. \tag{170}$$

Integrating such an equation over the interval $[0, t]$ gives:

$$\int_0^t \mathrm{d}\mathscr{F} = \int_0^t \left\langle \frac{\partial \mathscr{F}}{\partial \gamma}, \frac{\mathrm{d}\gamma}{\mathrm{d}t} \right\rangle^{\mathbb{A}} \mathrm{d}t + \int_0^t \left\langle \frac{\partial \mathscr{F}}{\partial \dot{\gamma}}, \frac{\mathrm{d}\dot{\gamma}}{\mathrm{d}t} \right\rangle^{\mathbb{A}} \mathrm{d}t. \tag{171}$$

The integral on the left-hand side is equal to $\mathscr{F}|_0^t$, while the last integral on the right-hand side may be evaluated through integration by parts and equals

$$\int_0^t \left\langle \frac{\partial \mathscr{F}}{\partial \dot{\gamma}}, \frac{d\dot{\gamma}}{dt} \right\rangle^{\mathbb{A}} dt = \left\langle \frac{\partial \mathscr{F}}{\partial \dot{\gamma}}, \frac{d\gamma}{dt} \right\rangle^{\mathbb{A}} \Big|_0^t - \int_0^t \left\langle \frac{d}{dt} \frac{\partial \mathscr{F}}{\partial \dot{\gamma}}, \frac{d\gamma}{dt} \right\rangle^{\mathbb{A}} dt$$

$$= 2\,\mathscr{F}|_0^t - \int_0^t \left\langle \frac{d}{dt} \frac{\partial \mathscr{F}}{\partial \dot{\gamma}}, \frac{d\gamma}{dt} \right\rangle^{\mathbb{A}} dt, \tag{172}$$

by Equation (166). From Equations (171) and (172), it follows that:

$$\mathscr{F}|_0^t = \int_0^t \left\langle \frac{\partial \mathscr{F}}{\partial \gamma} - \frac{d}{dt} \frac{\partial \mathscr{F}}{\partial \dot{\gamma}}, \frac{d\gamma}{dt} \right\rangle^{\mathbb{A}} dt + 2\,\mathscr{F}|_0^t, \tag{173}$$

hence

$$-\mathscr{F}|_0^t = \int_0^t \left\langle \frac{\partial \mathscr{F}}{\partial \gamma} - \frac{d}{dt} \frac{\partial \mathscr{F}}{\partial \dot{\gamma}}, \frac{d\gamma}{dt} \right\rangle^{\mathbb{A}} dt. \tag{174}$$

From the differential inclusion (153), it follows that the integrand is null; therefore, $\mathscr{F}|_0^t = 0$ for every $t \in [0, 1]$. $\square$

The above result means that any geodesic trajectory is traveled at a constant speed. Since, along a geodesic arc, the speed $\sqrt{\langle \dot{\gamma}_x^y(t), \dot{\gamma}_x^y(t) \rangle_{\gamma_x^y(t)}}$ is constant for any t, it holds that

$$d(x,y) = \sqrt{\langle \dot{\gamma}_x^y(0), \dot{\gamma}_x^y(0) \rangle_x}. \tag{175}$$

It is worth underlining that, provided that one knows the functional expression of the geodesic connecting two points, their distance does not require integration, but just algebraic operations. (The difficulty is hidden by the fact that finding the exact expression of a geodesic connecting two points may be more troublesome than one might expect.)

We make an observation in order to justify an idea presented at the beginning of this subsection about normal naïve acceleration for a geodesic.

Observation 2. *Let us take a closer look at the differential inclusion* (153)*, where* $\mathscr{F}(x,v) = \langle v, G_x(v) \rangle$*, namely*

$$\frac{\partial \langle \dot{\gamma}, G_\gamma(\dot{\gamma}) \rangle^{\mathbb{A}}}{\partial \gamma} - \frac{d}{dt} \frac{\partial \langle \dot{\gamma}, G_\gamma(\dot{\gamma}) \rangle^{\mathbb{A}}}{\partial \dot{\gamma}} \in N_\gamma \mathbb{M}. \tag{176}$$

Under the hypothesis that the metric kernel G_x *does not depend explicitly on the point* x*, the first term on the left-hand side vanishes. In this case, let us plainly assume that* $G_x \equiv \mathrm{id}_x$*. In addition, we have already established that* $\frac{\partial \langle \dot{\gamma}, G_\gamma(\dot{\gamma}) \rangle^{\mathbb{A}}}{\partial \dot{\gamma}} = 2G_\gamma(\dot{\gamma})$*. The above differential inclusion may hence be written as*

$$(G_x \equiv \mathrm{id}_x \Rightarrow)\ \ddot{\gamma} \in N_\gamma \mathbb{M}, \tag{177}$$

namely, the naïve acceleration is perpendicular to the velocity. This is perhaps the simplest form of a geodesic equation on a manifold. This is, in fact, the case for the hypersphere surveyed in an example.

A geodesic arc may be expressed in terms of two pieces of information, in addition to its endpoints, such as the initial values $\gamma(0) = x \in \mathbb{M}$ and $\dot{\gamma}(0) = v \in T_x\mathbb{M}$, which represent the point where a geodesic departs from and its initial speed, respectively. A geodesic arc determined by these two pieces of information will be denoted as $\gamma_{x,v} : [0, 1] \to \mathbb{M}$.

Example 21. *Let us consider a geodesic* $\gamma_{x,v} : [0, 1] \to \mathbb{M}$*, with* $(x,v) \in T\mathbb{M}$*, such that* $\dot{\gamma}_{x,v}(0) = v$*. In this case, it holds that*

$$L(\gamma_{x,v}) = \sqrt{\langle v, v \rangle_x} = \sqrt{\mathscr{F}(x,v)}. \tag{178}$$

■

8.2. Exponential and Logarithmic Maps

To a geodesic arc $\gamma_{x,v} : [0, 1] \to \mathbb{M}$ is associated a map $\exp_x : T_x\mathbb{M} \to \mathbb{M}$ defined as:

$$\exp_x(v) := \gamma_{x,v}(1). \tag{179}$$

The function (179) is termed *exponential map* with pole $x \in \mathbb{M}$ and is of paramount importance in manifold calculus. The exponential map $\exp : T\mathbb{M} \to \mathbb{M}$ takes two arguments, the point $x \in \mathbb{M}$ over the manifold and a vector $v \in T_x\mathbb{M}$ tangent to the manifold at the pole x, and returns a point on the manifold itself. A practical reading of the exponential map is that it advances a point x toward a direction v, like vector addition moves a point along a straight line in flat spaces. In other words, the following two expressions are the counterparts of each other in a flat space $\mathbb{R}^n$ and in a manifold $\mathbb{M}$:

$$\begin{cases} \gamma_{x,v}(t) = x + t\,v \text{ in } \mathbb{R}^n, \ t \in [0, 1], \\ \gamma_{x,v}(t) = \exp_x(t\,v) \text{ in } \mathbb{M}, \ t \in [0, 1]. \end{cases} \tag{180}$$

Much like the first expression denotes the shortest/straightest line on a flat space (endowed with a Euclidean metric), the second expression denotes the shortest/straightest line over a curved manifold.

The exponential map depends on the pair $(x, v) \in T\mathbb{M}$ and is locally invertible around $v = 0$. This follows from the fact that $\exp_x(0) = x$ and by local invertibility results. In the flat manifold $\mathbb{R}^n$, the exponential map may be inverted easily; in fact

$$\text{from } y = x + v \text{ it follows that } v = y - x. \tag{181}$$

In the flat manifold $\mathbb{R}^n$, we recover the classical and intuitive meaning of the exponential map and of its inverse:

- **Exponential map**: Given a point x and a vector v in $\mathbb{R}^n$, the exponential map $\exp_x(v) := x + v$ moves the point from x to $x + v$. (Indeed, the term 'vector' comes from the homonymus Latin term that means 'transporter'.)
- **Inverse exponential map**: Given two points x and y in $\mathbb{R}^n$, the inverse of the exponential map $\exp_x^{-1}(y)$ returns a vector v such that $\exp_x(v) = \exp_x(\exp_x^{-1}(y)) = x + (y - x) \equiv y$. In other words, the inverse exponential map applied to two points x, y determines the vector v that 'transports' x to y.

Let us observe that the inverse exponential map is non-symmetric in its arguments, namely the notation $\exp_x^{-1}(y)$ makes it immediately clear that it is not allowed to swap x with y. Indeed, even in the flat space $\mathbb{R}^n$, it holds that $\exp_x^{-1}(y) := y - x = -(x - y) = -\exp_y^{-1}(x)$. (In the case of a curved manifold, such a reciprocity relation is more convoluted.)

Generally, the inverse exponential map is termed *logarithmic map*. On a manifold $\mathbb{M}$ where it is defined an exponential map $\exp : T\mathbb{M} \to \mathbb{M}$ a logarithmic map is denoted as $\log : \mathbb{M}^2 \to T\mathbb{M}$. A logarithmic map takes as arguments two points x, $y \in \mathbb{M}$ and returns a *vector* $v = \log_x(y) \in T_x\mathbb{M}$. It is important to underline that a logarithmic map is defined only locally, namely, only if x, $y \in \mathbb{M}$ lay sufficiently close to one another. The lack of a global logarithmic map may be understood as follows. A logarithmic map is defined by the following equation:

$$\exp_x(\log_x y) = y, \tag{182}$$

and therefore, the existence of $\log_x y$ is tied to the chance of determining one and only one geodesic arc γ_x^y that connects the given points x, $y \in \mathbb{M}$, and then, by the equality $\gamma_x^y(t) = \gamma_{x,v}(t)$, it follows that $\log_x y \equiv v$. However, not every pair of points may be connected by a unique geodesic, and hence a global logarithmic map, in general, fails to exist. A quick example of such an unavoidable problem is found on the sphere $\mathbb{S}^2$; taking x as the North pole and y as the South pole, there exist infinitely many geodesic lines that connect the two poles. Hence, $\log_x y$ is undetermined.

Let us examine the expressions of geodesic arcs, geodesic distances and exponential maps for manifolds of interest in system theory and non-linear control.

- **Hypercube**: The space $\mathbb{R}^p$ endowed with the Euclidean metric admits straight lines as geodesics; in fact, since $T_x\mathbb{R}^p = \mathbb{R}^p$ it follows that $N_x\mathbb{R}^p = \{0\}$, and hence the geodesic equation is simply $\ddot{\gamma}_{x,v} = 0$ and its solution is $\gamma_{x,v}(t) = x + t\,v$ for every $x \in \mathbb{R}^p$, $v \in T_x\mathbb{R}^p$ and $t \in [0, 1]$. Now, take two points $x, y \in \mathbb{R}^p$ and look for a geodesic arc connecting them with $t \in [0, 1]$. It is necessary to find a vector $v \in T_x\mathbb{R}^p$ such that $\gamma_{x,v}(1) = y$. Such a vector is clearly $v = y - x$; hence, the unique geodesic arc connecting x to y is $\gamma_x^y(t) = x + t(y - x)$. Since $\sqrt{\langle \dot{\gamma}_x^y(0), \dot{\gamma}_x^y(0)\rangle_x} = \|v\|$, the Riemannian distance reads

$$d(x,y) = \|y - x\|, \tag{183}$$

which is a well-known result from calculus and geometry.

- **Hypersphere**: On the hypersphere $\mathbb{S}^{p-1}$ embedded in the Euclidean space $\mathbb{R}^p$, a geodesic line may be conceived as a curve on which a particle, departing from the point $x \in \mathbb{S}^{p-1}$ with velocity $v \in T_x\mathbb{S}^{p-1}$, slides with *constant* speed $\|v\|$, where $\|\cdot\|$ denotes the standard L_2 vector norm. On the hypersphere, we denote such a curve as $\gamma_{x,v}(t)$, where the variable $t \in [0, 1]$ provides a parametrization of the curve. The differential equation characterizing geodesics on the hypersphere may be determined by observing that, with the given conditions, in this case the naïve acceleration of the particle must be either null or normal to the tangent space at any point of the hypersphere itself, namely $\ddot{\gamma}_{x,v}(t) \in N_{\gamma_{x,v}(t)}\mathbb{S}^{p-1}$. Since the normal space to a hypersphere at a point x is radial along x, the geodesic equation reads as $\ddot{\gamma}_{x,v}(t) = \lambda\gamma_{x,v}(t)$. In explicit form, the equation of the geodesic on the unit hypersphere may be written as [49]:

$$\gamma_{x,v}(t) = x\cos(\|v\|t) + v\sin(\|v\|t)\|v\|^{-1}, \; t \in [0, 1], \tag{184}$$

as it is easy to verify by substitution. Additionally, it is easy to verify that $\gamma_{x,v}(0) = x$, $\dot{\gamma}_{x,v}(0) = v$ and that $\gamma_{x,v}(t)^\top\gamma_{x,v}(t) = 1$ for every t. The exponential map associated with the above geodesic is

$$\exp_x(v) := x\cos(\|v\|) + v\sin(\|v\|)\|v\|^{-1}. \tag{185}$$

The relationship (184) for the geodesic represents a 'great circle' on the hypersphere. Now let us take two points $x, y \in \mathbb{S}^{p-1}$ (non-antipodal, such that $x^\top y \neq -1$) and let us look for a geodesic arc of the form (184) connecting them. It is clearly necessary to find a vector $v \in T_x\mathbb{S}^{p-1}$ such that $\gamma_{x,v}(1) = y$. Such an equation in the unknown v may be expressed explicitly as

$$x\cos(\|v\|) + v\,\mathrm{sinc}(\|v\|) = y, \tag{186}$$

where 'sinc' denotes the cardinal sine function. Pre-multiplying the above equation by $x^\top$ gives $\cos(\|v\|) = x^\top y$, namely $\|v\| = \mathrm{acos}(x^\top y)$; hence

$$v = \frac{y - x(x^\top y)}{\mathrm{sinc}(\mathrm{acos}(x^\top y))}. \tag{187}$$

This expression represents the inverse of the exponential map applied to points $x, y \in \mathbb{S}^{p-1}$, namely

$$\log_x y := \frac{(I_p - xx^\top)y}{\mathrm{sinc}(\mathrm{acos}(x^\top y))}. \tag{188}$$

Notice that such a logarithmic map is defined only when $x^\top y \neq -1$, namely when the two points are antipodal. The unique geodesic arc connecting x to y is given by

$$\gamma_x^y(t) = x\cos(\mathrm{acos}(x^\top y)t) + \frac{y - x(x^\top y)}{\mathrm{sinc}(\mathrm{acos}(x^\top y))}\,\mathrm{sinc}(\mathrm{acos}(x^\top y)t). \qquad (189)$$

A noticeable consequence is that, since $\sqrt{\langle \dot\gamma_x^y(0), \dot\gamma_x^y(0)\rangle_x} = \|v\|$, the Riemannian distance between the points x and y reads

$$d(x,y) = \mathrm{acos}(x^\top y), \qquad (190)$$

where the inverse cosine function 'acos' returns a value in $[0,\ \pi]$.

- **Special orthogonal group**: In general, it is not easy to obtain the expression of a geodesic arc on a given manifold in closed form. In the present case, with the assumptions considered, the geodesic on $\mathbb{SO}(p)$ departing from the identity with velocity $H \in \mathfrak{so}(p)$ has expression $\tilde\gamma(t) = \exp(t\,H)$. (It is important to verify that $\tilde\gamma(0) = I_p$ and $\left.\frac{d\tilde\gamma(t)}{dt}\right|_{t=0} = H$.) It might be useful to verify such an essential result by the help of the following arguments. A geodesic $\tilde\gamma(t)$ on the Riemannian manifold $\mathbb{SO}(p)$, embedded in the Euclidean ambient space $\mathbb{A} := \mathbb{R}^{p\times p}$ and endowed with its canonical metric, departing from the identity I_p, should satisfy $\ddot{\tilde\gamma}(t) \in N_{\tilde\gamma(t)}\mathbb{SO}(p)$, and therefore it should hold:

$$\ddot{\tilde\gamma}(t) = \tilde\gamma(t)S(t)\,,\ \text{ with } S^\top(t) = S(t). \qquad (191)$$

Additionally, we know that any geodesic arc belongs entirely to the base manifold; therefore $\tilde\gamma^\top(t)\tilde\gamma(t) = I_p$. By differentiating such an expression two times with respect to the parameter t, one obtains:

$$\ddot{\tilde\gamma}^\top(t)\tilde\gamma(t) + 2\dot{\tilde\gamma}^\top(t)\dot{\tilde\gamma}(t) + \tilde\gamma^\top(t)\ddot{\tilde\gamma}(t) = 0_p. \qquad (192)$$

By plugging Equation (191) into Equation (192), we find that $S(t) = -\dot{\tilde\gamma}^\top(t)\dot{\tilde\gamma}(t)$, which leads to the second-order differential equation on the orthogonal group:

$$\ddot{\tilde\gamma}(t) = -\tilde\gamma(t)(\dot{\tilde\gamma}^\top(t)\dot{\tilde\gamma}(t)), \qquad (193)$$

to be solved with the initial conditions $\tilde\gamma(0) = I_p$ and $\dot{\tilde\gamma}(0) = H$. It is a straightforward task to verify that the solution to this second-order differential equation is given by the one-parameter curve $\tilde\gamma(t) = \mathrm{Exp}\,(t\,H)$, where 'Exp' denotes a matrix exponential. The expression of the geodesic arc in the position of interest may be made explicit by taking advantage of the Lie-group structure of the orthogonal group endowed with the canonical metric. In fact, let us consider the pair $X \in \mathbb{SO}(p)$ and $V \in T_X\mathbb{SO}(p)$ as well as the geodesic $\gamma(t)$ that emanates from X, namely $\gamma(0) = X$, with velocity V. We claim that the geodesic departing from $X \in \mathbb{SO}(p)$ in the direction $V \in T_X\mathbb{SO}(p)$ is:

$$\gamma_{X,V}(t) = X\,\mathrm{Exp}\,(t\,X^\top V). \qquad (194)$$

In fact, let us consider the left-translated curve $\tilde\gamma(t) = X^\top \gamma_{X,V}(t)$. It has the following properties:

1. The curve $\gamma_{X,V}(t)$ belongs to the orthogonal group at any time. This may be proven by computing the quantity $\gamma_{X,V}^\top(t)\gamma_{X,V}(t)$ and taking into account that the identity $\mathrm{Exp}^\top(t\,H) = \mathrm{Exp}\,(-t\,H)$ holds true. Therefore,

$$\begin{aligned}
\gamma_{X,V}^\top(t)\gamma_{X,V}(t) &= \mathrm{Exp}^\top(tX^\top V)X^\top X\,\mathrm{Exp}\,(tX^\top V)\\
&= \mathrm{Exp}\,(-tX^\top V)\,\mathrm{Exp}\,(tX^\top V) =\\
&= I_p.
\end{aligned}$$

2. It satisfies Equation (193); hence, it is a geodesic. In fact, notice that $\tilde{\gamma}(t) = X^\top X \operatorname{Exp}(tX^\top V) = \operatorname{Exp}(tX^\top V)$; hence, $\dot{\tilde{\gamma}}(t) = X^\top V \operatorname{Exp}(tX^\top V)$ and $\ddot{\tilde{\gamma}}(t) = (X^\top V)^2 \operatorname{Exp}(tX^\top V)$. Now, the right-hand side of Equation (193) has the expression $-\operatorname{Exp}(tX^\top V)(X^\top V \operatorname{Exp}(tX^\top V))^\top X^\top V \operatorname{Exp}(tX^\top V) = -\operatorname{Exp}(tX^\top V) \operatorname{Exp}(-tX^\top V)(X^\top V)^\top X^\top V \operatorname{Exp}(tX^\top V) = (X^\top V)^2 \operatorname{Exp}(tX^\top V)$ because $X^\top V \in \mathfrak{so}(p)$, hence the claim.

3. It satisfies $\tilde{\gamma}(0) = X$ and $\dot{\tilde{\gamma}}(0) = X(X^\top V) = V$; hence, it has the correct base point and direction.

Therefore, the exponential map associated with the above geodesic expression reads

$$\exp_X(V) := X \operatorname{Exp}(X^\top V). \tag{195}$$

Now, fixed two points $X, Y \in \mathbb{SO}(p)$, let us compute the geodesic arc $\gamma_X^Y : [0, 1] \to \mathbb{SO}(p)$ that joins them. It all boils down in finding a tangent vector $V \in T_X \mathbb{SO}(p)$ such that $Y = X \operatorname{Exp}(X^\top V)$. Pre-multiplying this expression by $X^\top$ gives $X^\top Y = \operatorname{Exp}(X^\top V)$; hence, $\operatorname{Log}(X^\top Y) = X^\top V$. Pre-multiplying the last equality by X gives the sought tangent vector as $V = X \operatorname{Log}(X^\top Y)$. The logarithmic map associated with the above exponential map therefore reads

$$\log_X Y := X \operatorname{Log}(X^\top Y). \tag{196}$$

The Riemannian distance between X and Y then takes the expression

$$d(X, Y) = \|X \operatorname{Log}(X^\top Y)\|_{\mathrm{F}} = \|\operatorname{Log}(X^\top Y)\|_{\mathrm{F}}, \tag{197}$$

where 'Log' denotes, as usual, the matrix logarithm.

- **Stiefel manifold**: Let us consider the expression of geodesics corresponding to two metrics.

  1. *Euclidean metric*: The solution of the geodesic equation, with the initial conditions $\gamma_{X,V}(0) = X \in \operatorname{St}(n, p)$ and $\dot{\gamma}_{X,V}(0) = V \in T_X \operatorname{St}(n, p)$, reads [33]:

  $$\gamma_{X,V}(t) = [X\ V] \operatorname{Exp}\left(t\begin{bmatrix} X^\top V & -V^\top V \\ I_p & X^\top V \end{bmatrix}\right) I_{2p,p} \operatorname{Exp}(-tX^\top V), \tag{198}$$

  for $t \in [0, 1]$. The expression of the exponential map corresponding to the Euclidean metric reads, therefore,

  $$\exp_X(V) := [X\ V] \operatorname{Exp}\left(\begin{bmatrix} X^\top V & -V^\top V \\ I_p & X^\top V \end{bmatrix}\right) I_{2p,p} \operatorname{Exp}(-X^\top V), \tag{199}$$

  while the expression of the logarithmic map and of the Riemannian distance between two given points, in closed form, are unknown at present (to the best of the author's knowledge). The pseudo-identity matrix $I_{2p,p}$, with $2p$ rows and p columns, is just an identity matrix topping a zero matrix.

  2. *Canonical metric*: The geodesic arc $\gamma_{X,V}$ may be computed as follows. Let Q and R denote the factors of the thin QR factorization of the matrix V, then:

  $$\gamma_{X,V}(t) = [X\ Q] \operatorname{Exp}\left(t\begin{bmatrix} X^\top V & -R^\top \\ R & 0_p \end{bmatrix}\right)\begin{bmatrix} I_p \\ 0_p \end{bmatrix}, \tag{200}$$

  for $t \in [0, 1]$. The expression of the exponential map corresponding to the canonical metric is, therefore,

  $$\exp_X(V) := [X\ Q] \operatorname{Exp}\left(\begin{bmatrix} X^\top V & -R^\top \\ R & 0_p \end{bmatrix}\right)\begin{bmatrix} I_p \\ 0_p \end{bmatrix}, \tag{201}$$

while the expressions of the logarithmic map and of the Riemannian distance between two points are still unknown.

In fact, neither the logarithmic map nor the geodesic distance are known in closed form for a Stiefel manifold.

- **Real symplectic group**: According to the two considered metrics, we have:

  1. *KM metric*: Under the pseudo-Riemannian metric (115), it is indeed possible to solve the geodesic equation in closed form. The geodesic curve $\gamma_{Q,V} : [0, 1] \to \mathrm{Sp}(2n)$ with $Q \in \mathrm{Sp}(2n)$ and $V \in T_Q\mathrm{Sp}(2n)$ corresponding to the indefinite Khvedelidze–Mladenov metric (115) has the expression:

  $$\gamma_{Q,V}(t) = Q\,\mathrm{Exp}\,(tQ^{-1}V). \tag{202}$$

  In fact, the geodesic equation in variational form is:

  $$\delta \int_0^1 \mathrm{tr}(\gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma})\mathrm{d}t = 0. \tag{203}$$

  The calculation of this variation is facilitated by the following rules of the calculus of variations:

  $$\delta(Q^{-1}) = -Q^{-1}(\delta Q)Q^{-1}, \tag{204}$$

  $$\delta\left(\frac{\mathrm{d}Q}{\mathrm{d}t}\right) = \frac{\mathrm{d}}{\mathrm{d}t}(\delta Q), \tag{205}$$

  $$\delta(QZ) = (\delta Q)Z + Q(\delta Z), \tag{206}$$

  for curves $Q, Z : [0, 1] \to \mathrm{Sp}(2n)$. By computing the variations, integrating by parts and recalling that the variations vanish at endpoints, it is found that the geodesic equation in variational form reads:

  $$\int_0^1 \mathrm{tr}(\delta\gamma(\gamma^{-1}\ddot{\gamma}\gamma^{-1} - \gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1}))\mathrm{d}t = 0. \tag{207}$$

  The variation $\delta\gamma \in T_\gamma\mathrm{Sp}(2n)$ is arbitrary. By the structure of the normal space $N_Q\mathrm{Sp}(2n)$, the equation $\mathrm{tr}(P^\top\delta\gamma) = 0$, with $\delta\gamma \in T_\gamma\mathrm{Sp}(2n)$, implies that $P^\top = HJ\gamma^{-1}$ with $H \in \mathfrak{so}(2n)$. Therefore, Equation (207) is satisfied if and only if:

  $$\gamma^{-1}\ddot{\gamma}\gamma^{-1} - \gamma^{-1}\dot{\gamma}\gamma^{-1}\dot{\gamma}\gamma^{-1} = HJ\gamma^{-1}, \; H \in \mathfrak{so}(2n), \tag{208}$$

  or, equivalently,

  $$\ddot{\gamma} - \dot{\gamma}\gamma^{-1}\dot{\gamma} = \gamma HJ, \tag{209}$$

  for some $H \in \mathfrak{so}(2n)$. In order to determine the value of matrix H, note that:

  $$\gamma^\top J\gamma - J = 0 \Rightarrow \ddot{\gamma}^\top J\gamma + 2\dot{\gamma}^\top J\dot{\gamma} + \gamma^\top\ddot{\gamma} = 0.$$

  Substituting the expression $\ddot{\gamma} = \dot{\gamma}\gamma^{-1}\dot{\gamma} + \gamma HJ$ into the above equation yields the condition $JHJ = 0$. Hence, $H = 0$ and the geodesic equation reads:

  $$\ddot{\gamma} - \dot{\gamma}\gamma^{-1}\dot{\gamma} = 0. \tag{210}$$

  Its solution, with the initial conditions $\gamma(0) = Q \in \mathrm{Sp}(2n)$ and $\dot{\gamma}(0) = V \in T_X\mathrm{Sp}(2n)$, is found to be of the form (202). By definition of matrix exponential, it follows that $\dot{\gamma}_{Q,V}(t) = V\,\mathrm{Exp}\,(t\,Q^{-1}V)$.

  2. *Euclidean metric*: The expression of the geodesic corresponding to the Euclidean metric was derived in [50]. Let $\gamma : [0,1] \to \mathrm{Sp}(2n)$ be a geodesic arc connecting

the points $X, Y \in \mathrm{Sp}(2n)$. Let us define $h(t) := \gamma^{-1}(t)\dot\gamma(t)$. The geodesic that minimizes the following energy functional

$$\int_0^1 \mathrm{tr}(h^\top(t)h(t))\mathrm{d}t, \tag{211}$$

is the solution of the differential (Lax) equation:

$$\dot h(t) = h^\top(t)h(t) - h(t)h^\top(t). \tag{212}$$

Furthermore, for the initial conditions $\gamma(0) = X \in \mathrm{Sp}(2n)$ and $\dot\gamma(0) = V \in T_X\mathrm{Sp}(2n)$, the geodesic on the real symplectic group is given by

$$\gamma_{X,V}(t) = X\,\mathrm{Exp}\,(t(X^{-1}V)^\top)\,\mathrm{Exp}\,(t[(X^{-1}V) - (X^{-1}V)^\top]), \tag{213}$$

in the case that the real symplectic group is equipped by a Euclidean metric.

- **Space of symmetric, positive-definite matrices**: The geodesic arc, corresponding to the canonical metric, emanating from a point $P \in \mathbb{S}^+(p)$ in the direction $V \in T_P\mathbb{S}^+(p)$ has the expression:

$$\gamma_{P,V}(t) = \sqrt[\mathbb{S}]{P}\,\mathrm{Exp}\left(t\sqrt[\mathbb{S}]{P^{-1}}V\sqrt[\mathbb{S}]{P^{-1}}\right)\sqrt[\mathbb{S}]{P}, \tag{214}$$

where $\sqrt[\mathbb{S}]{\cdot}$ denotes a symmetric matrix square root. The exponential and the logarithmic maps for $P, Q \in \mathbb{S}^+(p)$ and $V \in T_P\mathbb{S}^+(p)$ thus read:

$$\exp_P(V) = \sqrt[\mathbb{S}]{P}\,\mathrm{Exp}\left(\sqrt[\mathbb{S}]{P^{-1}}V\sqrt[\mathbb{S}]{P^{-1}}\right)\sqrt[\mathbb{S}]{P}, \tag{215}$$

$$\log_P(Q) = \sqrt[\mathbb{S}]{P}\,\mathrm{Log}\left(\sqrt[\mathbb{S}]{P^{-1}}Q\sqrt[\mathbb{S}]{P^{-1}}\right)\sqrt[\mathbb{S}]{P}. \tag{216}$$

The symmetric matrix square root of a $\mathbb{S}^+(p)$ matrix P may be computed by means of its eigenvalue factorization. In fact, if the matrix P is factored as $X\,\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p)X^\top$, with $X \in \mathbb{O}(p)$ and every $\lambda_k > 0$, then it holds that $\sqrt[\mathbb{S}]{P} = X\,\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3}, \ldots, \sqrt{\lambda_p})X^\top$. The squared Riemannian distance between two points $P, Q \in \mathbb{S}^+(p)$ is given by

$$\begin{aligned}
d^2(P,Q) &= \langle\log_P Q, \log_P Q\rangle_P = \mathrm{tr}((\log_P Q)P^{-1}(\log_P Q)P^{-1}) \\
&= \mathrm{tr}\left(\sqrt[\mathbb{S}]{P}\,\mathrm{Log}\left(\sqrt[\mathbb{S}]{P^{-1}}Q\sqrt[\mathbb{S}]{P^{-1}}\right)\sqrt[\mathbb{S}]{P}P^{-1}\sqrt[\mathbb{S}]{P}\,\mathrm{Log}\left(\sqrt[\mathbb{S}]{P^{-1}}Q\sqrt[\mathbb{S}]{P^{-1}}\right)\sqrt[\mathbb{S}]{P}P^{-1}\right) \\
&= \mathrm{tr}\left(\mathrm{Log}^2\left(\sqrt[\mathbb{S}]{P^{-1}}Q\sqrt[\mathbb{S}]{P^{-1}}\right)\sqrt[\mathbb{S}]{P}P^{-1}\sqrt[\mathbb{S}]{P}\right) \\
&= \mathrm{tr}\left(\mathrm{Log}^2\left(\sqrt[\mathbb{S}]{P^{-1}}Q\sqrt[\mathbb{S}]{P^{-1}}\right)\right).
\end{aligned}$$

Notice that the following identity holds true:

$$\mathrm{tr}(I_p - \sqrt[\mathbb{S}]{P^{-1}}Q\sqrt[\mathbb{S}]{P^{-1}})^k = \mathrm{tr}(I_p - QP^{-1})^k, \tag{217}$$

and hence the Riemannian distance between two SPD matrices may be written equivalently as

$$d^2(P,Q) = \mathrm{tr}(\mathrm{Log}^2(QP^{-1})), \tag{218}$$

as obtained by the series expansion of the matrix logarithm function (cf. Section 2.1).

- **Grassmann manifold**: A geodesic arc on a Grassmann manifold emanating from $[X] \in \mathrm{Gr}(n, p)$ with the velocity $V \in T_{[X]}\mathrm{Gr}(n, p)$ may be written as:

$$
\gamma_{[X],V}(t) = [XB\ A] \begin{bmatrix} \cos(Dt) \\ \sin(Dt) \end{bmatrix} B^\top,
$$
$$
= XB \cos(Dt)B^\top + A\sin(Dt)B^\top. \tag{219}
$$

where $X \in \mathrm{St}(n, p)$ denotes a Stiefel matrix whose columns form an orthonormal basis of the subspace $[X]$ and $ADB^\top$ denotes the compact singular value factorization of the matrix V. The sin/cos functions applied to a diagonal matrix simply act, in terms of components, on the entries of the main diagonal. As a compact, smooth Riemannian manifold, a Grassmann manifold is *geodesically complete*, which implies that *any two points on* $\mathrm{Gr}(n, p)$ *may be connected by a geodesic arc.*
The exponential map associated with the canonical metric reads:

$$
\exp_{[X]}(V) = XB\cos(D)B^\top + A\sin(D)B^\top. \tag{220}
$$

The logarithmic map $\log_{[X]}[Y]$ of two subspaces $[X], [Y] \in \mathrm{Gr}(n, p)$ is not easy to compute in general. In [40], it is shown that if their Stiefel representatives $X, Y \in \mathrm{St}(n, p)$ are such that the product $X^\top Y$ is symmetric, then $\log_{[X]}[Y] = ADB^\top$, where $B(\cos D)B^\top$ denotes the spectral factorization of the matrix $X^\top Y$ and $A := (I - XX^\top)YB(\sin D)^{-1}$.

Example 22. *Let us compute the length of a geodesic curve $\gamma_x^y : [0,\ 1] \to \mathbb{S}^{n-1}$ that, by definition, represents the distance between points $x, y \in \mathbb{S}^{n-1}$:*

$$
d(x, y) := \mathrm{L}(\gamma_x^y). \tag{221}
$$

Let us start by showing that:

$$
\dot{\gamma}_x^y(t) = \frac{1}{\sin\theta}[\theta\cos(\theta t)y - \theta\cos(\theta(1 - t))x] \tag{222}
$$

with $\cos\theta := x^\top y$; hence,

$$
\|\dot{\gamma}_x^y(t)\|^2 = \frac{\theta^2}{\sin^2\theta}\left[\cos^2(\theta t) + \cos^2(\theta(1 - t)) - 2\cos(\theta t)\cos(\theta(1 - t))x^\top y\right]
$$
$$
= \theta^2 = \mathrm{acos}^2(x^\top y). \tag{223}
$$

It is instructive to verify that $\|\dot{\gamma}_x^y(t)\|^2$ does not actually depend on t, which facilitates the computation of the length. In fact, the length of $\gamma_x^y : [0,\ 1] \to \mathbb{S}^{n-1}$ is given by:

$$
\mathrm{L}(\gamma_x^y) = \int_0^1 \|\dot{\gamma}_x^y(t)\|\mathrm{d}t = \mathrm{acos}(x^\top y)\int_0^1 \mathrm{d}t. \tag{224}
$$

Therefore, the distance between $x, y \in \mathbb{S}^{n-1}$ equals:

$$
d(x, y) = \mathrm{acos}(x^\top y). \tag{225}
$$

As a further example, let us determine the length of the geodesic curve $\gamma_{R,V} : [0,\ 1] \to \mathbb{SO}(n)$ given by the relationship (194). By definition, it holds that

$$
\mathrm{L}(\gamma_{R,V}) = \int_0^1 \|\dot{\gamma}_{R,V}(t)\|\mathrm{d}t. \tag{226}
$$

Let us show that

$$\dot{\gamma}_{R,V}(t) = (R\,R^\top)V\operatorname{Exp}(R^\top V\,t) = V\operatorname{Exp}(R^\top V\,t),\tag{227}$$

therefore

$$
\begin{aligned}
\mathrm{L}(\gamma_{R,V}) &= \int_0^1 \sqrt{\operatorname{tr}[(V\operatorname{Exp}(R^\top V\,t))^\top V\operatorname{Exp}(R^\top V\,t)]}\,\mathrm{d}t \\
&= \int_0^1 \sqrt{\operatorname{tr}[\operatorname{Exp}(-R^\top V\,t)V^\top V\operatorname{Exp}(R^\top V\,t)]}\,\mathrm{d}t \\
&= \int_0^1 \|V\|\,\mathrm{d}t = \|V\|,
\end{aligned}
\tag{228}
$$

as proven in the general theoretical developments. ∎

The following observation clarifies a technical argument.

Observation 3. *The parameter t in the expressions of geodesic lines, which we may identify as time, indicates which point of a geodesic one is referring to. Normally, the value $t = 0$ corresponds to the initial point x, namely $\gamma_{x,v}(0) \equiv x$. It is instructive to observe that, in all examined cases, the velocity v and the time parameter t appear to be multiplied to one another. For example, the equation of the geodesic $\gamma_{x,v}$ on a hypersphere may be rewritten as*

$$\gamma_{x,v}(t) = \cos(\|v\|\,t)\,x + \sin(\|v\|\,t)\frac{v\,t}{\|v\|\,t},\ t > 0.\tag{229}$$

This means that it is always possible to re-scale the parameter t as $\alpha\,t$ as long as the velocity v is re-scaled as v/α, where $\alpha > 0$. This is the reason why the time parameter t normally ranges in the interval $[0, 1]$.

8.3. Geodesic Interpolation

The notions of metric, distance, geodesics, are related to the notion of interpolation. In fact, the interpolation between two points $x, y \in \mathbb{M}$ may be defined through the optimization problem [51]:

$$\iota_x^y(t) := \arg\min_{z\in\mathbb{M}}[(1 - t)d^2(x, z) + td^2(z, y)],\tag{230}$$

with parameter $t \in [0, 1]$ providing the degree of interpolation between the two points. The one-parameter curve $\iota_x^y : [0, 1] \to \mathbb{M}$ describes a trajectory over the manifold $\mathbb{M}$, having x and y as endpoints, which may be regarded as an interpolation curve. For example, the value $\iota_x^y(\frac{1}{2})$ is regarded as a midpoint between points x and y.

Example 23. *Let us consider two close points $X, Y \in \mathbb{SO}(p)$ and let us look for an interpolation of such points in $\mathbb{SO}(p)$ endowed with its canonical metric. We are looking for a minimizer of the criterion:*

$$Z \mapsto (1 - t)\|\operatorname{Log}(X^\top Z)\|_{\mathrm{F}}^2 + t\|\operatorname{Log}(Y^\top Z)\|_{\mathrm{F}}^2,\tag{231}$$

with $t \in [0, 1]$ assigned. We may look for an explicit solution of the above minimization problem by reasoning as follows. The points X and Y may be connected by a geodesic arc $\gamma_{X,V}(\tau) = X\operatorname{Exp}(\tau X^\top V)$ with $\tau \in [0, 1]$ and $V \in T_X\mathbb{SO}(p)$ such that $\gamma_{X,V}(1) = Y$, namely, with V satisfying condition $\operatorname{Exp}(X^\top V) = X^\top Y$. As the interpolating point $\iota_X^Y(t)$ must be as close as possible to both X and Y, it should belong to the geodesic $\gamma_{X,V}(\tau)$; therefore, we are left with the problem of minimizing the criterion function:

$$
\begin{aligned}
\tau \ \mapsto\ & (1 - t)\|\operatorname{Log}(X^\top X\operatorname{Exp}(\tau X^\top V))\|_{\mathrm{F}}^2 + t\|\operatorname{Log}(Y^\top X\operatorname{Exp}(\tau X^\top V))\|_{\mathrm{F}}^2 \\
&= (1 - t)\|\tau X^\top V\|_{\mathrm{F}}^2 + t\|(\tau - 1)X^\top V\|_{\mathrm{F}}^2 \\
&= ((1 - t)\tau^2 + t(\tau - 1)^2)\|\operatorname{Log}(X^\top Y)\|_{\mathrm{F}}^2,
\end{aligned}
$$

with respect to the variable τ. In the second line, we have used the identity $Y^\top X = \text{Exp}\left(-X^\top V\right)$. The minimum of the above criterion function in $[0, 1]$ is readily seen to be achieved in $\tau = t$; thus, the interpolating curve reads:

$$\iota_X^Y(t) = X\,\text{Exp}\left(t\,\text{Log}\left(X^\top Y\right)\right) = X(X^\top Y)^t \tag{232}$$

for $0 \leq t \leq 1$. Notice that a matrix power A^α, with $A \in \mathbb{R}^{n \times n}$ and $\alpha \in \mathbb{R}$, is defined (inasmuch as it exists) in terms of matrix exponential/logarithm, hence the last identity. ∎

Generalizing the notion of midpoint, one comes up with the notion of *mean value* (and of dispersion of a sample set around its mean value) in a metrizable manifold $\mathbb{M}$. We may consider what follows:

- The notion of 'mean value' of objects in a metrizable space should reflect the intuitive understanding that the mean value is an element of the space that locates 'amidst' the available objects. Therefore, a fundamental tool in the definition of mean value is a measure of 'how far apart' elements in the sample space lie to one another.
- The notion of 'metric variance' of objects in a metrizable space should be defined in a way that accounts for the dispersion of these objects about their mean values and also depends on how the dissimilarity of such objects is measured.

A way of defining the mean value of a set of objects $x_k \in \mathbb{M}$, $k = 1, \ldots, K$, is provided by the notion of Fréchet sample mean. The Fréchet sample mean and associated sample metric variance [52] may be defined as:

$$\mu := \arg\min_{x \in \mathbb{M}} \frac{1}{K} \sum_{k=1}^{K} d^2(x, x_k), \tag{233}$$

$$\sigma^2 := \min_{x \in \mathbb{M}} \frac{1}{K} \sum_{k=1}^{K} d^2(x, x_k), \tag{234}$$

where d denotes a distance function in the Riemannian manifold $\mathbb{M}$. It is worth noting that there is no guarantee, in general, that the optimization problem (233) admits a unique solution. For example, consider the space $\mathbb{S}^{p-1}$ and a sample set consisting of two antipodal points; then, every point over the equator is a mean value. (Notice that a mean is Fréchet only if it is a global minimizer, otherwise it is a Karcher mean.) If the sample points are close enough to each other, it is known that the mean value is unique [53].

The notion of mean value of a set of points belonging to a curved manifold is utterly important in system theory and non-liner control. In fact, the mean value μ is, by definition, close to all points in a collection of points. Therefore, the tangent space $T_\mu\mathbb{M}$ associated with a cloud of data points may serve as a reference tangent space (see, for example, [54]).

8.4. Coordinate-Prone Geodesy, Christoffel Symbols*

In terms of the variation of an energy functional, a geodesic on a manifold $\mathbb{M}$ is defined as the unique curve $\gamma(t) \in \mathbb{M}$, $t \in [0, 1]$ that meets the following requirement

$$\delta \int_0^1 \langle \dot{\gamma}, \dot{\gamma} \rangle_\gamma \mathrm{d}t = 0, \tag{235}$$

where the symbol δ denotes the variation of the integral functional. The integral functional in (235) represents the total action associated with the parametrized smooth curve $\gamma : [0, 1] \to \mathbb{M}$ of local coordinates $x^k = \gamma^k(t)$ and may be written explicitly as:

$$\int_0^1 \langle \dot{\gamma}, \dot{\gamma} \rangle_\gamma \mathrm{d}t = \int_0^1 g_{ij}(\gamma(t))\dot{\gamma}^i(t)\dot{\gamma}^j(t)\mathrm{d}t. \tag{236}$$

The variation δ in the expression (235) corresponds to a perturbation of the total action (236). Let $\eta : [0, 1] \to \mathbb{M}$ denote an arbitrary parametrized smooth curve of local coordinates $\eta^k(t)$ such that $\eta^k(0) = \eta^k(1) = 0$. Define the perturbed action as

$$\mathscr{A}_\eta^*(h) := \int_0^1 g_{ij}(\gamma + h\,\eta)(\dot\gamma^i + h\,\dot\eta^i)(\dot\gamma^j + h\,\dot\eta^j)\mathrm{d}t, \tag{237}$$

with $h \in [-\epsilon, \epsilon]$, $\epsilon > 0$. The condition (235) may be expressed explicitly in terms of the perturbed action as:

$$\lim_{h \to 0} \frac{\mathscr{A}_\eta^*(h) - \mathscr{A}_\eta^*(0)}{h} = 0, \ \forall \text{ perturbation } \eta. \tag{238}$$

The perturbed total action may be expanded around $h = 0$ as follows:

$$\mathscr{A}_\eta^*(h) = \int_0^1 \left[g_{ij}(x) + h\frac{\partial g_{ij}}{\partial x^k}\eta^k + \mathscr{O}(h^2) \right]\bigg|_{x=\gamma(t)} [\dot\gamma^i\dot\gamma^j + h(\dot\gamma^i\dot\eta^j + \dot\gamma^j\dot\eta^i) + \mathscr{O}(h^2)]\mathrm{d}t$$

$$= \int_0^1 g_{ij}(\gamma)\dot\gamma^i\dot\gamma^j\mathrm{d}t + h\int_0^1 \left[g_{ij}(x)(\dot\gamma^i\dot\eta^j + \dot\gamma^j\dot\eta^i) + \eta^k\frac{\partial g_{ij}}{\partial x^k}\dot\gamma^i\dot\gamma^j \right]\bigg|_{x=\gamma(t)} \mathrm{d}t + \mathscr{O}(h^2)$$

$$= \mathscr{A}_\eta^*(0) + h\int_0^1 \left[g_{ik}(x)\dot\gamma^i\dot\eta^k + g_{kj}(x)\dot\gamma^j\dot\eta^k + \frac{\partial g_{ij}}{\partial x^k}\dot\gamma^i\dot\gamma^j\eta^k \right]\bigg|_{x=\gamma(t)} \mathrm{d}t + \mathscr{O}(h^2),$$

where $\mathscr{O}(h^2)$ is such that $\lim_{h\to 0} h^{-1}\mathscr{O}(h^2) = 0$. The first term in the integral on the right-hand side of the last line may be integrated by parts, namely:

$$\int_0^1 g_{ik}(\gamma)\dot\gamma^i\dot\eta^k\mathrm{d}t = g_{ik}(\gamma)\dot\gamma^i\eta^k\Big|_0^1 - \int_0^1 \frac{d(g_{ik}(\gamma)\dot\gamma^i)}{dt}\eta^k\mathrm{d}t, \tag{239}$$

whose first term on the right-hand side vanishes to zero because $\eta^k(0) = \eta^k(1) = 0$, hence:

$$\int_0^1 g_{ik}(\gamma)\dot\gamma^i\dot\eta^k\mathrm{d}t = -\int_0^1 \left(\frac{\partial g_{ik}}{\partial x^j}\dot\gamma^i\dot\gamma^j + g_{ik}(x)\ddot\gamma^i \right)\bigg|_{x=\gamma(t)}\eta^k\mathrm{d}t. \tag{240}$$

A similar result holds for the second term within the integral on the right-hand side. Therefore, it holds that:

$$\frac{\mathscr{A}_\eta^*(h) - \mathscr{A}_\eta^*(0)}{h} = \int_0^1 \left[\left(\frac{\partial g_{ij}}{\partial x^k} - \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{kj}}{\partial x^i} \right)\dot\gamma^i\dot\gamma^j - 2g_{ik}(x)\ddot\gamma^i \right]\bigg|_{x=\gamma(t)}\eta^k\mathrm{d}t + \frac{\mathscr{O}(h^2)}{h}.$$

The condition (238), therefore, implies that:

$$g_{ik}\ddot\gamma^i + \frac{1}{2}\left(-\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{kj}}{\partial x^i} \right)\dot\gamma^i\dot\gamma^j = 0. \tag{241}$$

Let us now recall the notion of *Christoffel symbols of the first kind* (named after Elwin Bruno Christoffel) associated with a metric tensor of components g_{ij}. These coefficients are defined as:

$$\Gamma_{ijk} := \frac{1}{2}\left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} \right). \tag{242}$$

On top of it, the *Christoffel symbols of the second kind* are defined as $\Gamma_{ij}^h := g^{hk}\Gamma_{ijk}$. The Christoffel symbols of the second kind are symmetric in the covariant indices, namely, $\Gamma_{ij}^h = \Gamma_{ji}^h$.

On the basis of the Christoffel symbols, Equation (241) may be rewritten as

$$g_{ik}\ddot\gamma^i + \Gamma_{ijk}\dot\gamma^i\dot\gamma^j = 0. \tag{243}$$

Multiplying by g^{hk} and introducing the Christoffel symbols of the second kind, Equation (243) may be written as:

$$\ddot{\gamma}^i + \Gamma^i_{jk}\dot{\gamma}^j\dot{\gamma}^k = 0. \tag{244}$$

Such a system of second-order differential equations needs two initial conditions to be solved. Typical initial conditions are $\gamma^i(0) = x^i$ and $\dot{\gamma}^i(0) = v^i$.

Example 24. *One might wonder if the Christoffel symbols defined in* (242) *are the components of a tensor. The answer is negative, because none of its lower index denote covariancy.* ∎

9. Riemannian Gradient of a Manifold-to-Scalar Function

The gradient of a scalar function in a point of its manifold-type domain represents a degree of variability of the function in a vicinity of such a point. In the present section, we are going to survey the notion of 'Riemannian gradient' starting from more familiar cases to obtain a full definition for manifold-to-scalar functions.

9.1. Riemannian Gradient: Motivation and Definition

In the simplest case of a scalar function of a scalar variable $f : \mathbb{R} \to \mathbb{R}$, its degree of variability is quantified by its slope $\frac{\mathrm{d}f}{\mathrm{d}x}$. In fact, let us recall that, given a point $x \in \mathbb{R}$, moving slightly away from it of a little amount $h \in \mathbb{R}$, the value of the function $f(x + h)$ may be related to the value of $f(x)$ through a Taylor series expansion. Such an expansion, truncated to a first-order term by a Lagrange-type remainder, reads

$$f(x + h) = f(x) + \frac{\mathrm{d}f(x)}{\mathrm{d}x}h + \mathcal{O}(h^2), \tag{245}$$

and hence the first derivative of the function (that is, its gradient) truly quantifies the variability of the function around the point x. In the above expression, $\mathcal{O}$ again denotes a Landau symbol.

In the more involved case of a scalar function of several variables, $f : \mathbb{R}^p \to \mathbb{R}$, the gradient $\frac{\partial f}{\partial x}$ quantifies again the degree of variability of the function around a point, although such information is spread along p axes. In fact, in this case, there exist multiple directions along which the variability of a function may be explored. After choosing a direction, one may move away along a straight line. Starting from a point $x \in \mathbb{R}^p$ and moving away along a direction $v \in \mathbb{R}^p$ of a small fraction t, the value of $f(x + tv)$ may be related to the value of $f(x)$ through a multivariable Taylor series expansion which, in the first order, reads

$$f(x + tv) = f(x) + \left(\frac{\partial f(x)}{\partial x}\right)^\top (tv) + \mathcal{O}(t^2). \tag{246}$$

Therefore, the gradient of the function f (which, in this case, may be termed 'Euclidean gradient' and coincides with the 'Gateaux derivative' and with the 'Jacobian' of the function) encodes the variability of the function and the quantity $\left(\frac{\partial f(x)}{\partial x}\right)^\top v$ represents the directional derivative of the function f at x along the direction v. In the following, we shall utilize the shorter notation $\partial_x := \frac{\partial}{\partial x}$.

In the case of a scalar function whose domain is a curved manifold $f : \mathbb{M} \to \mathbb{R}$, its gradient, which we shall denote as $\mathrm{grad}_x f$, serves again to quantify the degree of variability of a function in the neighborhood of point $x \in \mathbb{M}$ along a given tangent direction. On the same line of reasoning as before, starting from a point $x \in \mathbb{M}$ one may think of moving away slightly along a given tangent direction $v \in T_x\mathbb{M}$. After choosing a direction, the simplest way to move away from a point along a given direction is to travel along a geodesic line for a short time. Thus, taking a geodesic $\gamma_{x,v} : [-\epsilon, \epsilon] \to \mathbb{M}$, the value of $f(\gamma_{x,v})$ may be related to the value $f(x)$ through a Taylor-type expansion, which is written as

$$f(\gamma_{x,v}(t)) = f(x) + \langle \mathrm{grad}_x f, tv \rangle_x + \mathcal{O}(t^2). \tag{247}$$

Therefore, the gradient of the function f quantifies the variability of a function around a point, while $\langle \mathrm{grad}_x f, v \rangle_x$ may be interpreted as a directional derivative of the function f at x along a direction v. From the relationship (247), we see that

$$\frac{f(\gamma_{x,v}(t)) - f(x)}{t} = \langle \mathrm{grad}_x f, v \rangle_x + \frac{\mathscr{O}(t^2)}{t}. \tag{248}$$

Taking the limit as t approaches zero, the second addendum on the right-hand side vanishes to zero; hence, we obtain

$$\langle \mathrm{grad}_x f, v \rangle_x = \lim_{t \to 0} \frac{f(\gamma_{x,v}(t)) - f(x)}{t} = \left. \frac{\mathrm{d}f(\gamma_{x,v}(t))}{\mathrm{d}t} \right|_{t=0}. \tag{249}$$

The right-hand side is resemblant of a Gateaux derivative of the function f in the direction v. The relationship (249) may be indeed taken as a definition of the gradient of the function with the proviso that it should hold for every $v \in T_x \mathbb{M}$.

By taking a closer look at the expression (249), it is readily seen that it holds even if the geodesic line is replaced by any smooth curve $\gamma : [-\epsilon, \epsilon] \to \mathbb{M}$ as long as $\gamma(0) = x$ and $\dot{\gamma}(0) = v$, in which case we can write

$$\langle \mathrm{grad}_x f, v \rangle_x = \lim_{t \to 0} \frac{f(\gamma(t)) - f(x)}{t}. \tag{250}$$

Now let us underline an illusory contradiction in the above expression.

Observation 4. *The curve γ takes part only in closed proximity of the point x; therefore, the right-hand side of the above relation only depends on the point x and the tangent v, as well as the function f, and, in particular, it does not depend on the metric. In other words, the result of the expression $\lim_{t \to 0} \frac{f(\gamma(t)) - f(x)}{t}$ quantifies how much the function f varies upon moving away from the point x toward the direction v. As opposed to this, the left-hand side of the relation (250) seems to depend on the metric $\langle \cdot, \cdot \rangle$. Since the metric may be chosen arbitrarily and independently of the function f and of the quantities x and v, the above relation looks like a conceptual clash.*

The only possible way to fix such an illusory contradiction is to recognize that the gradient itself must depend on the metric in such a way that the directional derivative $\langle \mathrm{grad}_x f, v \rangle_x$ does not. In other words, if we denote by $\langle \cdot, \cdot \rangle^{(1)}$ and $\langle \cdot, \cdot \rangle^{(2)}$ two metrics for the manifold $\mathbb{M}$, then it must hold that

$$\langle \mathrm{grad}_x^{(1)} f, v \rangle_x^{(1)} = \langle \mathrm{grad}_x^{(2)} f, v \rangle_x^{(2)} \tag{251}$$

for every $v \in T_x \mathbb{M}$, where $\mathrm{grad}_x^{(1)} f$ denotes the gradient of the function f at x corresponding to the metric $\langle \cdot, \cdot \rangle^{(1)}$ and $\mathrm{grad}_x^{(2)} f$ denotes the gradient of the function f at x corresponding to the metric $\langle \cdot, \cdot \rangle^{(2)}$.

Let us formalize the above discussion with the aim of formalizing the notion of *Riemannian gradient* on a Riemannian manifold. Let us consider a Riemannian manifold $\mathbb{M}$ embedded in an ambient space $\mathbb{A}$ and, for every point x, the tangent space $T_x \mathbb{M}$. Let us now consider a smooth function $f : \mathbb{M} \to \mathbb{R}$ and an inner product $\langle \cdot, \cdot \rangle^{\mathbb{A}}$ in $\mathbb{A}$. We shall require that the function f is extendable to a neighborhood of x in $\mathbb{A}$ so as to be able to compute the Gateaux derivative ∂_x of the function f with respect to x. (As usual, we are supposed to use an extension $\bar{f}$ in the following expressions.) Let us recall that

$$\lim_{h \to 0} \frac{f(x + h\,v) - f(x)}{h} =: \langle \partial_x f, v \rangle^{\mathbb{A}}. \tag{252}$$

We denote the inner product that the tangent bundle $T\mathbb{M}$ is endowed with at $x \in \mathbb{M}$ as $\langle \cdot, \cdot \rangle_x$. The Riemannian gradient $\mathrm{grad}_x^{\mathbb{M}} f$ of the function f over the manifold $\mathbb{M}$ at the point x is uniquely defined by the following two conditions:

- **Tangency condition**: For every $x \in \mathbb{M}$, $\mathrm{grad}_x^{\mathbb{M}} f \in T_x \mathbb{M}$.
- **Metric compatibility condition**: For every $x \in \mathbb{M}$ and every $v \in T_x \mathbb{M} \subset \mathbb{A}$, it holds that $\langle \mathrm{grad}_x^{\mathbb{M}} f, v \rangle_x = \langle \partial_x f, v \rangle^{\mathbb{A}}$.

The tangency condition codifies the requirement that a Riemannian gradient must always be a tangent vector to the manifold, namely $(x, \mathrm{grad}_x f) \in T\mathbb{M}$, while the metric compatibility condition codifies the requirement that the inner product between a gradient vector and any other tangent vector belonging to the same tangent space takes a value that is invariant with respect to the chosen metric. The latter condition is better understood by considering that the linear component of the variation of a function when moving away from a point x of a quantity v is given by

$$\mathrm{d}_x f(v) = \langle \mathrm{grad}_x^{\mathbb{M}} f, v \rangle_x. \tag{253}$$

Clearly, the amount $\mathrm{d}_x f(v)$ depends only on x, f and the displacement v, certainly not on the metric. As a consequence, the gradient needs to depend on the metric. The 'reference' inner product is taken as the inner product that the ambient space is endowed with. In fact, the Riemannian gradient of a manifold-to-scalar function represents the rate of change of such function at a given point toward a given direction. The superscript $^{\mathbb{M}}$ may be removed for simplicity.

Let us survey the calculation of the Riemannian gradient in a number of spaces of interest in applications.

- **Hypercube**: In the space $\mathbb{R}^p$ endowed with the Euclidean metric, the Gateaux derivative of a regular function $f : \mathbb{R}^p \to \mathbb{R}$ is simply the column array of partial derivatives of function f with respect to the entries of the column array $x = [x^{(1)} \, x^{(2)} \, x^{(3)} \, \cdots \, x^{(p)}]^\top$, namely

$$\mathrm{grad}_x^{\mathbb{R}^p} f = \partial_x f := \begin{bmatrix} \frac{\partial f}{\partial x^{(1)}} \\ \frac{\partial f}{\partial x^{(2)}} \\ \vdots \\ \frac{\partial f}{\partial x^{(p)}} \end{bmatrix}. \tag{254}$$

Likewise, the Gateaux derivative of a regular function $f : \mathbb{R}^{p \times q} \to \mathbb{R}$ is the Jacobian matrix of partial derivatives of function f with respect to the entries of the matrix X, denoted by $X^{(i,j)}$, namely

$$\mathrm{grad}_X^{\mathbb{R}^{p \times q}} f = \partial_X f := \begin{bmatrix} \frac{\partial f}{\partial X^{(1,1)}} & \frac{\partial f}{\partial X^{(1,2)}} & \cdots & \frac{\partial f}{\partial X^{(1,q)}} \\ \frac{\partial f}{\partial X^{(2,1)}} & \frac{\partial f}{\partial X^{(2,2)}} & \cdots & \frac{\partial f}{\partial X^{(2,q)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial X^{(p,1)}} & \frac{\partial f}{\partial X^{(p,2)}} & \cdots & \frac{\partial f}{\partial X^{(p,q)}} \end{bmatrix}. \tag{255}$$

- **Hypersphere**: Given a regular function $f : \mathbb{S}^{p-1} \to \mathbb{R}$, its Riemannian gradient at $x \in \mathbb{S}^{p-1}$ is denoted as $\mathrm{grad}_x^{\mathbb{S}^{p-1}} f \in T_x \mathbb{S}^{p-1}$. The Riemannian gradient associated with the canonical metric has the expression:

$$\mathrm{grad}_x^{\mathbb{S}^{p-1}} f = (I_p - xx^\top) \partial_x f. \tag{256}$$

In fact, the metric compatibility condition prescribes that $\langle \mathrm{grad}_x^{\mathbb{S}^{p-1}} f, v \rangle = \langle \partial_x f, v \rangle$, for every $v \in T_x \mathbb{S}^{p-1}$; hence, $\langle \mathrm{grad}_x^{\mathbb{S}^{p-1}} f - \partial_x f, v \rangle = 0$ for every $v \in T_x \mathbb{S}^{p-1}$, and therefore $\mathrm{grad}_x^{\mathbb{S}^{p-1}} f - \partial_x f \in N_x \mathbb{S}^{p-1}$. It follows that

$$\mathrm{grad}_x^{\mathbb{S}^{p-1}} f = \partial_x f + \lambda x, \tag{257}$$

for some $\lambda \in \mathbb{R}$. The tangency condition then entails $(\mathrm{grad}_x^{\mathbb{S}^{p-1}} f)^\top x = 0$, hence that $(\partial_x f + \lambda\, x)^\top x = 0$, namely $\lambda = -x^\top \partial_x f$, which gives the stated result.

- **Special orthogonal group**: Let us compute the Riemannian gradient of a regular function $f : \mathbb{SO}(p) \to \mathbb{R}$. Let the manifold $\mathbb{SO}(p)$ be equipped with its canonical metric. Let $\mathrm{grad}_X^{\mathbb{SO}(p)} f$ denote the gradient vector of a function f at $R \in \mathbb{SO}(p)$ derived from the canonical metric. According to the compatibility condition for the Riemannian gradient it must hold that:

$$\langle V, \mathrm{grad}_R^{\mathbb{SO}(p)} f \rangle = \langle V, \partial_R f \rangle, \ \forall V \in T_R \mathbb{SO}(p)$$

and therefore:

$$\langle V, \mathrm{grad}_R^{\mathbb{SO}(p)} f - \partial_R f \rangle = 0, \ \forall V \in T_R \mathbb{SO}(p).$$

This implies that the quantity $\mathrm{grad}_R^{\mathbb{SO}(p)} f - \partial_R f$ belongs to the normal space $N_R \mathbb{SO}(p)$, namely:

$$\mathrm{grad}_R^{\mathbb{SO}(p)} f = \partial_R f + RS, \text{ with } S \text{ being } p \times p \text{ symmetric.} \tag{258}$$

In order to determine the unknown matrix S, we may exploit the tangency condition, namely $(\mathrm{grad}_R^{\mathbb{SO}(p)} f)^\top R + R^\top (\mathrm{grad}_R^{\mathbb{SO}(p)} f) = 0_p$. Let us first pre-multiply both sides of Equation (258) by $R^\top$, which gives:

$$R^\top \mathrm{grad}_R^{\mathbb{SO}(p)} f = R^\top \partial_R f + S.$$

The above equation, transposed hand by hand, gives:

$$(\mathrm{grad}_R^{\mathbb{SO}(p)} f)^\top R = \partial_R^\top f R + S.$$

Hand-by-hand summation of the last two equations gives:

$$R^\top \partial_R f + \partial_R^\top f R = -2S,$$

that is:

$$S = -\frac{1}{2} \left(R^\top \partial_R f + \partial_R^\top f R \right). \tag{259}$$

By plugging the expression (259) into the expression (258), we obtain the Riemannian gradient in the orthogonal group, corresponding to its canonical metric:

$$\mathrm{grad}_R^{\mathbb{SO}(p)} f = \frac{1}{2} \left(\partial_R f - R \partial_R^\top f R \right). \tag{260}$$

- **Stiefel manifold**: Let us compute the expression of the Riemannian gradient of a regular function $f : \mathrm{St}(n,p) \to \mathbb{R}$ at a point $X \in \mathrm{St}(n,p)$ corresponding to the Euclidean and the canonical metrics. Recall that the Riemannian gradient $\mathrm{grad}_X^{\mathrm{St}(n,p)} f$ in a Stiefel manifold embedded in the Euclidean space $\mathbb{R}^{n \times p}$ is the unique matrix in $T_X \mathrm{St}(n,p)$ such that:

$$\mathrm{tr}\left(U^\top \partial_X f \right) = \langle U, \mathrm{grad}_X f \rangle_X, \ \forall U \in T_X \mathrm{St}(n,p). \tag{261}$$

1. *Euclidean metric*: The metric compatibility condition becomes $\mathrm{tr}(U^\top (\partial_X f - \mathrm{grad}_X^{\mathrm{St}(n,p)} f)) = 0$, which implies that $\mathrm{grad}_X^{\mathrm{St}(n,p)} f = \partial_X f + XS$, with S being $p \times p$ symmetric. Pre-multiplying both sides of this equation by the matrix $X^\top$ yields $S + X^\top \partial_X f = X^\top \mathrm{grad}_X^{\mathrm{St}(n,p)} f$. Transposing both hands of the above equation and summing hand by hand yields:

$$S = -\frac{1}{2}\left(\partial_X^\top f X + X^\top \partial_X f\right) + \frac{1}{2}\left(\left(\text{grad}_X^{\text{St}(n,p)} f\right)^\top X + X^\top \text{grad}_X^{\text{St}(n,p)} f\right). \quad (262)$$

From the condition $\text{grad}_X^{\text{St}(n,p)} f \in T_X \text{St}(n,p)$, according to Equation (46), it follows that $(\text{grad}_X^{\text{St}(n,p)} f)^\top X + X^\top \text{grad}_X^{\text{St}(n,p)} f = 0$. In conclusion, the sought Riemannian gradient reads:

$$\text{grad}_X^{\text{St}(n,p)} f = \partial_X f - \frac{1}{2} X\left(\partial_X^\top f X + X^\top \partial_X f\right). \quad (263)$$

2. *Canonical metric*: The metric compatibility condition prescribes that:

$$\text{tr}\left(U^\top \left(\partial_X f - \left(I_n - \frac{1}{2} X X^\top\right) \text{grad}_X^{\text{St}(n,p)} f\right)\right) = 0, \ \forall U \in T_X \text{St}(n,p), \quad (264)$$

and therefore, invoking the tangency condition as well, it turns out that:

$$\begin{cases} \partial_X f - \left(I_n - \frac{1}{2} X X^\top\right) \text{grad}_X^{\text{St}(n,p)} f = XS, \\ S = \frac{1}{2}\left(X^\top \partial_X f + \partial_X^\top f X\right). \end{cases} \quad (265)$$

Solving for the Riemannian gradient yields the final expression:

$$\text{grad}_X^{\text{St}(n,p)} f = \partial_x f - X \partial_X^\top f X. \quad (266)$$

- **Real symplectic group**: According to the two considered metrics, the expression of the gradient may be computed as outlined below.

  1. *KM metric*: The structure of the pseudo-Riemannian gradient of a regular function $f : \text{Sp}(2n) \to \mathbb{R}$ associated with the Khvedelidze–Mladenov metric (115) is given by

$$\nabla_Q f = \frac{1}{2} QJ\left(Q^\top \partial_Q f J - J \partial_Q^\top f Q\right). \quad (267)$$

  In fact, the pseudo-Riemannian gradient of a regular function $f : \text{Sp}(2n) \to \mathbb{R}$ associated with the metric (115) is computed as the solution of the following system of equations:

$$\begin{cases} \text{tr}(Q^{-1}\text{grad}_Q f \, Q^{-1} V) = \text{tr}(\partial_Q^\top f \, V), \ \forall V \in T_Q \text{Sp}(2n), \\ \text{grad}_Q^\top f \, J \, Q + Q^\top J \text{grad}_Q f = 0. \end{cases} \quad (268)$$

  The first constraint ensures the compatibility of the pseudo-Riemannian gradient with the chosen metric, while the second constraint enforces the requirement for the pseudo-Riemannian gradient to lay on a specific tangent space. In particular, the metric compatibility condition may be recast as:

$$\text{tr}(V^\top(\partial_Q f - Q^{-\top}\text{grad}_Q^\top f \, Q^{-\top})) = 0. \quad (269)$$

  The above condition implies that $\partial_Q f - Q^{-\top}\text{grad}_Q^\top f \, Q^{-\top} \in N_Q \text{Sp}(2n)$, and hence that $\partial_Q f - Q^{-\top}\text{grad}_Q^\top f \, Q^{-\top} = JQH$ with $H \in \mathfrak{so}(2n)$. Therefore, the pseudo-Riemannian gradient of the criterion function f has the expression:

$$\nabla_Q f = Q \, \partial_Q^\top f \, Q - QHJ. \quad (270)$$

In order to determine the value of the unknown skew-symmetric matrix H, it is sufficient to plug the expression (270) of the gradient within the tangency condition, which becomes:

$$\left(Q\partial_Q^\top f\, Q - QHJ\right)^\top JQ + Q^\top J\left(Q\partial_Q^\top f\, Q - QHJ\right) = 0. \tag{271}$$

Solving for H gives:

$$H = -\tfrac{1}{2}\left(JQ^\top \partial_Q f + \partial_Q^\top f\, QJ\right). \tag{272}$$

Plugging the above expression into (270) gives the result (267).

2. *Euclidean metric*: The structure of the Riemannian gradient $\nabla_Q f$ of a regular function $f : \mathrm{Sp}(2n) \to \mathbb{R}$ corresponding to the metric (116) reads:

$$\nabla_Q f = \tfrac{1}{2}QJ\left(\partial_Q^\top f\, QJ - JQ^\top \partial_Q f\right). \tag{273}$$

In fact, the Riemannian gradient $\nabla_Q f$ must satisfy the conditions:

$$\begin{cases} \nabla_Q f = QJ(H - Q^{-1}J\partial_Q f), \\ H = \tfrac{1}{2}(Q^{-1}J\partial_Q f + \partial_Q^\top f\, QJ), \end{cases} \tag{274}$$

from which the expression of the Riemannian gradient associated with the metric (116) follows.

- **Manifold of symmetric, positive-definite matrices**: The Riemannian gradient $\mathrm{grad}_P f$ of the function $f : \mathbb{S}^+(p) \to \mathbb{R}$ may be calculated as the unique vector in $T_P\mathbb{S}^+(p)$ that satisfies the following equation:

$$\mathrm{tr}(V\partial_P f) = \mathrm{tr}\left(VP^{-1}(\mathrm{grad}_P^{\mathbb{S}^+(p)} f)P^{-1}\right), \ \forall V \in T_P\mathbb{S}^+(n). \tag{275}$$

The solution of the above equation satisfies:

$$\partial_P f - P^{-1}(\mathrm{grad}_P^{\mathbb{S}^+(p)} f)P^{-1} = \tfrac{1}{2}\left(\partial_P f - \partial_P^\top f\right), \tag{276}$$

and hence the expression of the sought Riemannian gradient follows:

$$\mathrm{grad}_P^{\mathbb{S}^+(p)} f = \tfrac{1}{2}P\left(\partial_P f + \partial_P^\top f\right)P. \tag{277}$$

- **Grassmann manifold**: The Riemannian gradient $\mathrm{grad}_{[X]} f$ may be calculated by its definition and reads as:

$$\nabla_{[X]}^{\mathrm{Gr}(n,p)} f = (I_n - XX^\top)\partial_X f. \tag{278}$$

The above calculations may be conveniently unified by recalling the notion of metric kernel, which affords turning the metric compatibility condition into a projection. Let us recall that the inner product $\langle v, u\rangle_x$ may be written as $\langle v, G_x(u)\rangle^{\mathbb{A}}$. On the basis of such an equality, the metric compatibility condition may be rewritten as

$$\langle G_x(\mathrm{grad}_x^{\mathbb{M}} f), v\rangle^{\mathbb{A}} = \langle \partial_x f, v\rangle^{\mathbb{A}} \text{ for every } v \in T_x\mathbb{M}. \tag{279}$$

Equivalently, $\langle G_x(\mathrm{grad}_x^{\mathbb{M}} f) - \partial_x f, v\rangle^{\mathbb{A}} = 0$ for every $v \in T_x\mathbb{M}$. By the definition of a normal space, such a condition may be rewritten as

$$G_x(\mathrm{grad}_x^{\mathbb{M}} f) - \partial_x f \in N_x\mathbb{M}. \tag{280}$$

In practical terms, the Gateaux derivative $\partial_x f \in \mathbb{A}$ may be uniquely decomposed into the sum of two terms, a tangent one (that it, the gradient) and a normal one. Upon discarding the normal component, the remainder is the sought gradient.

The latter observation may be expressed using the notion of orthogonal projection. Let us define an *orthogonal projector* $\Pi_x : \mathbb{A} \to T_x\mathbb{M}$ as an operator that maps an element of the ambient space to a tangent vector in a specific tangent space. Let us observe that

$$\Pi_x(a) \in T_x\mathbb{M} \Leftrightarrow a - \Pi_x(a) \in N_x\mathbb{M}. \tag{281}$$

The projection operator $\Pi_x : \mathbb{A} \to T_x\mathbb{M}$ is defined by the two conditions:

1. **Tangency**: $\Pi_x(a) \in T_x\mathbb{M}$ for $x \in \mathbb{M}$ and $a \in \mathbb{A}$.
2. **Complementarity**: $\langle v, a - \Pi_x(a)\rangle^{\mathbb{A}} = 0$ for all $v \in T_x\mathbb{M}$.

Notice that if $\nu \in N_x\mathbb{M}$ then $\Pi_x(\nu) = 0$, while if $v \in T_x\mathbb{M}$ then $\Pi_x(v) = v$.

Observation 5. *Orthogonal projection stems from a minimization problem, namely, given a point $a \in \mathbb{A}$ and a vector space $\mathbb{V}$, the orthogonal projection of a on $\mathbb{V}$ is the shortest vector in $\mathbb{V}$ that approximates a, namely*

$$\min_{v \in \mathbb{V}} \langle a - v, a - v\rangle^{\mathbb{A}}. \tag{282}$$

Applying orthogonal projection to both sides of the relationship (280) leads to

$$\Pi_x(G_x(\mathrm{grad}_x^{\mathbb{M}} f)) - \Pi_x(\partial_x f) = 0, \tag{283}$$

but since $\Pi_x(G_x(\mathrm{grad}_x^{\mathbb{M}} f)) = G_x(\mathrm{grad}_x^{\mathbb{M}} f)$, we may write the explicit formula

$$\mathrm{grad}_x^{\mathbb{M}} f = G_x^{-1}(\Pi_x(\partial_x f)). \tag{284}$$

The latter expression is quite suggestive, as it states that the Riemannian gradient of a function may be computed as the projection of its Gateaux derivative over a tangent space, to make it a tangent vector, further compensated by the metric kernel to take into account the effect of the metric (namely, to make it sure that the inner product of gradient with any tangent vector is independent of the metric). Notice that $\langle \mathrm{grad}_x^{\mathbb{M}} f, v\rangle_x = \langle G_x^{-1}(\Pi_x(\partial_x f)), G_x(v)\rangle^{\mathbb{A}}$. Since G_x is self-adjoint, such expression simplifies into $\langle \Pi_x(\partial_x f), v\rangle^{\mathbb{A}}$ which is, in fact, independent of the metric kernel.

Let us summarize a few expressions of interest of orthogonal projection.

- **Hypersphere**: An expression of an orthogonal projector $\Pi_x : \mathbb{R}^n \to T_x\mathbb{S}^{p-1}$, for $x \in \mathbb{S}^{p-1}$, where the ambient space $\mathbb{R}^p$ is endowed with a Euclidean metric, is:

$$\Pi_x(a) := (I_p - xx^\top)a. \tag{285}$$

  Notice that $a - \Pi_x(a) = (x^\top a)x$ is radial (that is, directed along x) and hence normal.
- **Stiefel manifold**: It might be useful to define an orthogonal projection operator $\Pi_X : \mathbb{R}^{n\times p} \to T_X\mathrm{St}(n, p)$, for $X \in \mathrm{St}(n, p)$. Let us assume the ambient space $\mathbb{A} := \mathbb{R}^{n\times p}$ be endowed with a Euclidean metric $\langle U, V\rangle^{\mathbb{A}} := \mathrm{tr}(U^\top V)$. In this case, the orthogonal projection takes the expression

$$\Pi_X(U) := U - \tfrac{1}{2}X(X^\top U + U^\top X), \tag{286}$$

- **Grassmann manifold**: An expression of orthogonal projection $\Pi_{[X]} : \mathbb{R}^{n\times p} \to T_{[X]}\mathrm{Gr}(n, p)$, for $[X] \in \mathrm{Gr}(n, p)$ is

$$\Pi_{[X]}(V) := (I_n - XX^\top)V. \tag{287}$$

Let us apply the above considerations to a quadratic function on the hypersphere.

Example 25. *Let us consider the function $f : \mathbb{S}^{n-1} \to \mathbb{R}$ defined as:*

$$f(x) := \tfrac{1}{2} x^\top P_0 \, x, \tag{288}$$

with $P_0 \in \mathbb{S}^+(n)$ being constant. We assume the hypersphere to be embedded in the ambient space $\mathbb{R}^n$ endowed with a Euclidean metric. The Euclidean gradient (and also the Gateaux derivative) of such a function reads as

$$\frac{\partial f}{\partial x} = P_0 \, x. \tag{289}$$

Since, on the hypersphere, $\Pi_x(a) = (I - xx^\top)a$, the Riemannian gradient of the function f takes the expression

$$\mathrm{grad}_x f = (I_n - xx^\top)P_0 \, x = P_0 \, x - (x^\top P_0 x)x. \tag{290}$$

∎

Let us verify that the expression (286) is indeed an orthogonal projection over the tangent bundle $T\mathrm{St}(n,p)$.

Example 26. *The expression $\Pi_X(U) := U - \tfrac{1}{2} X(X^\top U + U^\top X)$ realizes an orthogonal projection over the tangent space $T_X\mathrm{St}(n,p)$. To prove such a statement, it suffices to show that $\Pi_X(U) \in T_X\mathrm{St}(n,p)$ and that $\langle U - \Pi_X(U), V \rangle^\mathbb{A} = 0$ for every $V \in T_X\mathrm{St}(n,p)$, which represent the conditions of tangency and of orthogonality, respectively. To what concerns the first property:*

$$\begin{aligned}
& X^\top \Pi_X(U) + \Pi_X^\top(U)X \\
& = X^\top U + U^\top X - \tfrac{1}{2}(X^\top X)(X^\top U + U^\top X) - \tfrac{1}{2}(X^\top U + U^\top X)(X^\top X) \\
& = 0.
\end{aligned} \tag{291}$$

To what concerns the second property, notice that

$$\begin{aligned}
& \langle U - \Pi_X(U), V \rangle^\mathbb{A} \\
& = \mathrm{tr}(V^\top(U - U + \tfrac{1}{2}X(X^\top U + U^\top X))) \\
& = \tfrac{1}{2}\mathrm{tr}(V^\top X(X^\top U + U^\top X)).
\end{aligned} \tag{292}$$

Now observe that $H := V^\top X$ is skew-symmetric, while $S := X^\top U + U^\top X$ is symmetric, for every $X \in \mathrm{St}(n,p)$, $U \in \mathbb{A}$ and $V \in T_X\mathrm{St}(n,p)$. Since $\mathrm{tr}(HS) \equiv 0$, the statement holds true. ∎

A further interesting counterexample clarifies the notion of orthogonality in orthogonal projection.

Example 27. *Let us consider the function $\Pi_X(U) := U - XU^\top X$ which realizes a projection over the tangent bundle $T\mathrm{St}(n,p)$ but not an orthogonal one as long as the Euclidean metric is concerned. Showing that it indeed realizes a projection is fairly easy; in fact,*

$$\begin{aligned}
& X^\top \Pi_X(U) + \Pi_X^\top(U)X \\
& = X^\top U - U^\top X + U^\top X - X^\top U \\
& = 0.
\end{aligned} \tag{293}$$

Likewise, we can show that such a projection is not *orthogonal in the Euclidean metric; in fact,*

$$
\begin{aligned}
&\langle U - \Pi_X(U), V \rangle^{\mathbb{A}} \\
&= \mathrm{tr}(V^\top (U - U + XU^\top X)) \\
&= \mathrm{tr}((V^\top X)(U^\top X)).
\end{aligned}
\tag{294}
$$

Since the matrix product $U^\top X$ is arbitrary, the latter expression is not guaranteed to be null. ∎

As a last example, let us consider the case of the symplectic group endowed with its canonical metric. The orthogonal projection is not easy to express in closed form.

Example 28. *Let us consider the symplectic group* $\mathrm{Sp}(2n)$ *embedded into the ambient space* $\mathbb{A} := \mathbb{R}^{2n \times 2n}$ *endowed with a Euclidean metric* $\langle A, B \rangle^{\mathbb{A}} := \mathrm{tr}(A^\top B)$. *By definition of orthogonal projection,* $\Pi_Q(A)$ *must be an element of the tangent space* $T_Q\mathrm{Sp}(2n)$, *while* $A - \Pi_Q(A)$ *must be an element of the normal space* $N_Q\mathrm{Sp}(2n)$. *According to the characterization (49), it should then hold*

$$
A - \Pi_Q(A) = JQH,
\tag{295}
$$

namely $\Pi_Q(A) = A - JQH$, *with H skew-symmetric to be determined. From the tangency condition, it follows that*

$$
(A - JQH)^\top JQ + Q^\top J(A - JQH) = 0.
\tag{296}
$$

Through some manipulations, the above equation may be rewritten as

$$
HQ^\top Q + Q^\top QH = -A^\top JQ + Q^\top J^\top A.
\tag{297}
$$

All quantities in the above equation are known except for the unknown skew-symmetric matrix H. The above equation is an instance of the more general Sylvester equation (for a review, see, e.g., [55]).

According to [55], there exist a number of expressions to write the solution to such an equation. One solution is based on the Kronecker operations. In particular, in this case we can use the Kronecker sum $I_{2n} \oplus Q^\top Q$ *and the vectorization operator 'vec' to rewrite Equation (297) as*

$$
(I_{2n} \oplus Q^\top Q) \mathrm{vec}(H) = \mathrm{vec}(-A^\top JQ + Q^\top J^\top A).
\tag{298}
$$

The orthogonal projection operation may therefore be expressed as

$$
\Pi_Q(A) = A - JQ\,\mathrm{vec}^{-1}\{(I_{2n} \oplus Q^\top Q)^{-1}\mathrm{vec}(-A^\top JQ + Q^\top J^\top A)\}.
\tag{299}
$$

A more elegant expression stems from an integral representation of the solution of the Sylvester Equation (297), namely

$$
\Pi_Q(A) = A + JQ \int_0^{+\infty} \mathrm{Exp}\,(Q^\top Q\,t)(-A^\top JQ + Q^\top J^\top A)\,\mathrm{Exp}\,(Q^\top Q\,t)\mathrm{d}t.
\tag{300}
$$

For the integral representation of the solution of a Sylvester equation, we refer the readers to the aforementioned review paper [55]. We mention that none of these solutions are efficient from a computational point of view.

9.2. Application of Riemannian Gradient to Optimization on Manifold

An important application of the Riemannian gradient, which is of specific interest in system theory and non-linear control, is gradient-based mathematical optimization, which we are going to briefly survey in the remainder of the present section.

Let us preliminarily observe that the relationship (250) that defines the notion of Riemannian gradient possess a further interpretation. Let us take a regular function

$f : \mathbb{M} \to \mathbb{R}$, a regular curve $\gamma : [-\epsilon, \epsilon] \to \mathbb{M}$ and a metric $\langle \cdot, \cdot \rangle$ in $T\mathbb{M}$. The functions f and γ may be composed as $f \circ \gamma : [-\epsilon, \epsilon] \to \mathbb{R}$. Therefore, from the relation (250), it follows that:

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t)) = \langle \mathrm{grad}_{\gamma(t)} f, \dot{\gamma}(t) \rangle_{\gamma(t)} \tag{301}$$

along the curve γ. As a matter of fact, such a relation is a counterpart of the familiar rule of derivation for composite functions

$$\frac{\mathrm{d}}{\mathrm{d}t} f(g(t)) = \frac{\mathrm{d}f}{\mathrm{d}g}(g(t)) \frac{\mathrm{d}g}{\mathrm{d}t}(t). \tag{302}$$

To determine the maximum of a function on a manifold it is possible to make use of a *gradient method*. Such a method is based on the following important property: given a regular function $f : \mathbb{M} \to \mathbb{R}$, the solution $x(t)$ of the following initial value problem

$$\dot{x}(t) = \pm \mathrm{grad}_{x(t)} f, \ x(0) = x_0 \in \mathbb{M}, \tag{303}$$

is such that

$$\lim_{t \to +\infty} x(t) = x_\star \tag{304}$$

where $x_\star$ denotes a point of local minimum of the function f near x_0 if one chooses the sign '$-$' in front of the gradient; otherwise, $x_\star$ denotes a local maximum of the function f near x_0 if one chooses the sign '$+$' in front of the gradient.

For a function with domain $\mathbb{R}^n$, the extremal points are among those that make its partial derivatives vanish to zero. A similar result holds for functions whose domain is a smooth manifold. Given a differentiable function $f : \mathbb{M} \to \mathbb{R}$, the extremal points are those $x \in \mathbb{M}$ for which:

$$\mathrm{grad}_x f = 0. \tag{305}$$

This observation may be exploited

- To determine those points $x_\star$ to which the solutions of the differential Equation (303) will tend toward;
- Considering that Equation (305) is often non-linear and difficult to solve in closed form, while the differential Equation (303) may be solved (albeit approximately) by numerical recipes, the differential Equation (303) may be considered as a way to solve an equation such as (305).

On the basis of the relation (301), it is quite straightforward to show the following fundamental result.

Theorem 4. *The differential equation*

$$\dot{x}(t) = -\mathrm{grad}_{x(t)} f, \ x(0) = x_0 \in \mathbb{M}, \tag{306}$$

generates trajectories toward points of the domain corresponding to ever-decreasing values of the function f.

Proof. Let us observe that

$$\dot{f} = \langle \mathrm{grad}_x f, \dot{x} \rangle_x = \langle \mathrm{grad}_x f, -\mathrm{grad}_x f \rangle_x = -\|\mathrm{grad}_x f\|_x^2 \leq 0, \tag{307}$$

where equality holds if and only if $\mathrm{grad}_x f = 0$. □

In a fully analogous way, upon choosing the sign $+$ in front of the gradient, the corresponding differential equation generates trajectories tending toward ever-increasing values of the function f. Notice that the above facts are completely independent of the initial value x_0 and from the chosen metric.

In practical terms, it is safe to assume that the function f admits maxima and minima. For example, in the case of a continuous function f defined on a compact manifold, such as the hypersphere $\mathbb{S}^{n-1}$, a theorem by Weierstrass ensures that f admits at least a maximum and a minimum. (In fact, the *Weierstrass extreme value theorem* holds for topological spaces which any manifold is an instance of.)

To summarize, the differential Equation (303) may be exploited to look for the extremal points of a function:

- The dynamical system $\dot{x}(t) = +\mathrm{grad}_{x(t)}f$, $x(0) = x_0 \in \mathbb{M}$ generates a trajectory in the state space $\mathbb{M}$ that tends toward a point of maximum of the function f located near the initial state x_0,
- Conversely, the dynamical system $\dot{x}(t) = -\mathrm{grad}_{x(t)}f$, $x(0) = x_0 \in \mathbb{M}$ generates a trajectory in the space $\mathbb{M}$ that tends toward a point of minimum of the function f located near the initial state x_0.

The locality property of such a gradient-based extremum search is due to the monotonicity property (307), which essentially shows that the trajectory of such systems fall in the basin of attraction of the extremal point that the initial condition also belongs to. In practice, the choice of an initial condition (also termed, in this context, *initial guess*), is a sensitive step that determines the success or the failure of extremal-point searching. In fact, a function f might possess several extremal points (namely, local maxima and minima). It pays to keep in mind that the extremal-point searching method based on the first-order dynamical systems (303) is able to determine only an extremal point at a time and, in particular, only the one nearest to the initial guess x_0. It is therefore important to put into effect some pre-estimation technique to initialize the search procedure correctly.

From an algorithmic point of view, as opposed to what is required in numerical simulation where precision is a common demand, in the case of optimization by a gradient method, only the last point of the trajectory, namely $x_\star$, needs to be approximated with good precision, and hence the employed numerical method does not need to be of a high order.

Example 29. *Let us consider Oja's equation [56,57]*

$$\dot{x}(t) = (I_n - x(t)x^\top(t))P_0\, x(t), \ x(0) = x_0 \in \mathbb{S}^{n-1}, \tag{308}$$

with $P_0 \in \mathbb{S}^+(n)$. As we have seen in the previous example, the right-hand side of Oja's equation coincides with the Riemannian gradient of the function (often termed 'Rayleigh quotient') $f(x) := \frac{1}{2}x^\top P_0\, x$ over the unit hypersphere endowed with its canonical metric. Such a function is quadratic, continuous and differentiable and is defined over a compact space; hence, it admits a maximum over the space $\mathbb{S}^{n-1}$. Indeed, it admits n local maxima. The differential Equation (308) hence affords determining one of its local maxima depending on the initial condition.

The extremal points are those that make the gradient $\mathrm{grad}_x f = P_0\, x - (x^\top P_0\, x)\, x$ vanish to zero, namely the solutions $x \in \mathbb{S}^{n-1}$ to the equation

$$P_0\, x = (x^\top P_0\, x)\, x. \tag{309}$$

Clearly, the solutions $x_\star$ coincide with the eigenvectors of the symmetric, positive-definite matrix P_0. Let us observe that the eigenvalue associated with the eigenvector $x_\star$ is $\lambda_\star := x_\star^\top P_0\, x_\star$; hence, $f(x_\star) = \frac{1}{2}\lambda_\star$. Since the dynamical system (308) looks for the maximal value of the function f, Oja's equation generates a trajectory that tends toward the eigenvector of P_0 corresponding to its maximal eigenvalue. ∎

The differential equations of the kind

$$\dot{x}(t) = \mathrm{grad}_{x(t)}f, \tag{310}$$

represent an instance of dynamical systems of the first order, on a manifold, of the gradient type. (Notice that to switch the sign in front of the gradient, it suffices to take the function $-f$ instead of f. Likewise, to rescale the gradient of a factor $\eta > 0$, it suffices to take ηf instead of f.)

9.3. A Golden Gradient Rule: Gradient of Squared Distance

A golden calculation rule involving the notion of Riemannian gradient and that of Riemannian distance is as follows:

$$\operatorname{grad}_x d^2(x,y) = -2\log_x y. \tag{311}$$

The formal proof of this result which, of course, holds under technical conditions, traces back to [53]. It is instructive to verify this property in a couple of cases.

Example 30. *Let us verify the calculation rule* (311) *for the familiar case that* $\mathbb{M} = \mathbb{R}^p$. *We have already shown that* $d^2(x,y) = \|\log_x y\|^2$. *In the case of the hypercube* $\mathbb{R}^p$, *endowed with a Euclidean metric, it holds that* $\log_x y = y - x$; *therefore,* $d^2(x,y) = \|x - y\|^2$, *the familiar Euclidean distance in* $\mathbb{R}^p$. *Now, let us observe that*

$$\frac{\partial \|x - y\|^2}{\partial x} = 2(x - y), \tag{312}$$

and henceforth

$$\frac{\partial d^2(x,y)}{\partial x} = -2(y - x) = -2\log_x y. \tag{313}$$

In addition, we can verify such a property for the hypersphere $\mathbb{S}^{p-1}$. *Let us recall that* $d^2(x,y) = \arccos^2(x^\top y)$; *therefore, it holds that*

$$\frac{\partial d^2(x,y)}{\partial x} = 2\,d(x,y)\left(-\frac{1}{\sqrt{1 - (x^\top y)^2}}\right)y. \tag{314}$$

By the calculation rule for the gradient on the unit hypersphere, we obtain

$$\operatorname{grad}_x d^2(x,y) = -\frac{2\,d(x,y)}{\sin d(x,y)}(I_p - xx^\top)y = -2\frac{(I_p - xx^\top)y}{\operatorname{sinc} d(x,y)} \tag{315}$$

which coincides precisely with $-2\log_x y$ *on the unit hypersphere.* ∎

For those who are mathematically minded, it is interesting to read a sketch of proof of the golden rule (311) based on properties that we have already surveyed.

Theorem 5. *Let us consider a Riemannian manifold* $(\mathbb{M}, \langle \cdot, \cdot \rangle)$ *embedded in an ambient space* $(\mathbb{A}, \langle \cdot, \cdot \rangle^{\mathbb{A}})$ *and a pair of points* $x, y \in \mathbb{M}$ *such that* $d(x,y)$ *and* $\log_x y$ *are well defined. The Riemannian gradient of the squared distance* $d^2(x,y)$ *with respect to the variable* x *equals* $-2\log_x y$.

Proof. Let us define the following functions:

- An arbitrary sufficiently smooth curve $\alpha_x(t)$ such that $\alpha_x(0) = x$, while $\dot{\alpha}_x(t)$ is arbitrary;
- The fundamental form $\mathscr{F}(x,v) := \langle v, v \rangle_x$;
- A geodesic curve $c(t,s)$ connecting the point y to the point $\alpha_x(t)$, emanating from the former, with parameter $s \in [0, 1]$;
- The partial derivatives $\dot{c}(t,s) := \frac{\partial c}{\partial t}(t,s)$ and $c'(t,s) := \frac{\partial c}{\partial s}(t,s)$.

In addition, notice that the following properties hold:

(P$_1$) $c(t,0) = y$, therefore $\dot{c}(t,0) = 0$;

(P$_2$) $c(t,1) = \alpha_x(t)$, therefore $\dot{c}(t,1) = \dot{\alpha}_x(t)$;

(P$_3$) $c'(t,1) = -\log_{\alpha_x(t)} y$. (In fact, notice that if $\gamma_y^x : [0,\ 1] \to \mathbb{M}$ denotes a geodesic from x to y then it holds that $\dot{\gamma}_x^y(0) = \log_y x$ and $\gamma_y^x(s) = \gamma_x^y(1-s)$; therefore, $\dot{\gamma}_y^x(1) = -\dot{\gamma}_x^y(0) = -\log_y x$.)

Let us recall that the squared Riemannian distance between two points may be written in terms of the fundamental form, namely

$$d^2(\alpha_x(t), y) = \mathscr{F}(c(t,s), c'(t,s)), \tag{316}$$

where the right-hand side of the above equation is independent of s, as we have shown in Section 8. Let us integrate both sides of the above equation with respect to s in the interval $[0,\ 1]$. This gives:

$$d^2(\alpha_x(t), y) = \int_0^1 \mathscr{F}(c(t,s), c'(t,s)) \mathrm{d}s. \tag{317}$$

Now, let us compute the derivative of both sides with respect to the parameter t:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} d^2(\alpha_x(t), y) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \mathscr{F}(c(t,s), c'(t,s)) \mathrm{d}s \\
&= \int_0^1 \left\langle \frac{\partial \mathscr{F}}{\partial c}(t,s), \frac{\partial c}{\partial t}(t,s) \right\rangle^{\mathbb{A}} \mathrm{d}s + \int_0^1 \left\langle \frac{\partial \mathscr{F}}{\partial c'}(t,s), \frac{\partial c'}{\partial t}(t,s) \right\rangle^{\mathbb{A}} \mathrm{d}s.
\end{aligned}
\tag{318}
$$

The second integral may be rewritten through the rule of integration by parts upon observing that $\frac{\partial c'}{\partial t} = \frac{\partial \dot{c}}{\partial s}$ (Schwarz) and recalling that $\frac{\partial \mathscr{F}(x,v)}{\partial v} = 2\,G_x(v)$, where G denotes the metric kernel associated with the metrics in $\mathbb{M}$ and $\mathbb{A}$. The application of such rule gives:

$$
\begin{aligned}
\int_0^1 \left\langle \frac{\partial \mathscr{F}}{\partial c'}(t,s), \frac{\partial c'}{\partial t}(t,s) \right\rangle^{\mathbb{A}} \mathrm{d}s &= \langle 2\,G_x(c'(t,s)), \dot{c}(t,s) \rangle^{\mathbb{A}} \Big|_{s=0}^{1} \\
&\quad - \int_0^1 \left\langle \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathscr{F}}{\partial c'}(t,s), \dot{c}(t,s) \right\rangle^{\mathbb{A}} \mathrm{d}s.
\end{aligned}
\tag{319}
$$

Putting the pieces together gives:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} d^2(\alpha_x(t), y) &= \int_0^1 \left\langle \frac{\partial \mathscr{F}}{\partial c}(t,s) - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathscr{F}}{\partial c'}(t,s), \dot{c}(t,s) \right\rangle^{\mathbb{A}} \mathrm{d}s \\
&\quad + \langle 2\,c'(t,s), \dot{c}(t,s) \rangle_{c(t,s)} \Big|_{s=0}^{1}
\end{aligned}
\tag{320}
$$

Since $c(\cdot, s)$ is a geodesic, the first integral is null due to the differential inclusion (153); therefore, we rest with

$$\frac{\mathrm{d}}{\mathrm{d}t} d^2(\alpha_x(t), y) = \langle 2\,c'(t,1), \dot{c}(t,1) \rangle_{c(t,1)} - \langle 2\,c'(t,0), \dot{c}(t,0) \rangle_{c(t,0)}. \tag{321}$$

Now, due to property P$_1$, the second term on the right-hand side of the above equality is null; hence, thanks to properties P$_2$–P$_3$, the above equality simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}t} d^2(\alpha_x(t), y) = 2\langle -\log_{\alpha_x(t)} y, \dot{\alpha}_x(t) \rangle_{\alpha_x(t)}. \tag{322}$$

Setting $t = 0$ and recalling the definition of gradient (249) proves the golden rule. $\square$

Let us emphasize how useful such a result is, since in non-linear control a number of functions to differentiate are based on distances. Let us apply the golden calculation rule to a non-linear control problem known as *consensus optimization*.

Example 31. *A category of dynamical systems of gradient type on manifolds are those associated with* consensus optimization *among a set of agents [58,59]. Let us consider, as an example, a dynamical system that affords reaching attitudinal consensus in a fleet of flying objects (such as drones).*

The attitude of a flying agent (namely, its orientation in space) may be represented by a rotation matrix $R \in \mathbb{SO}(3)$ that summarizes the three attitudinal coordinates roll, pitch *and* yaw. *A fleet of N flying agents may be described as a set of N rotation matrices $R_i \in \mathbb{SO}(3)$ with $i = 1, 2, 3, \ldots, N$.*

The attitude of each agent in the fleet is then described by a time-function $R_i(t)$. Consensus optimization may be formulated as follows:

- *Given a metric for the manifold $\mathbb{SO}(3)$, it is possible to define a distance function $d : \mathbb{SO}(3) \times \mathbb{SO}(3) \to \mathbb{R}_0^+$ between any pair of attitude matrices in $\mathbb{SO}(3)$;*
- *A hierarchy is established among the agents in a fleet through a set of weights $w_{ij} \geq 0$;*
- *Then, consensus optimization consists of determining an evolution law for each agent to minimize the distance between any pair of attitude matrices weighted according to the assigned hierarchy, namely, to minimize the function*

$$f_i := \frac{1}{2} \sum_{j=1}^{N} w_{ij} \, d^2(R_i, R_j), \; with \; i = 1, 2, 3, \ldots, N. \tag{323}$$

Notice that $d(R_i, R_i) = 0$; hence, the values assigned to the diagonal coefficients w_{ii} are unimportant.

On the basis of the calculation rule (311), it turns out that

$$\mathrm{grad}_{R_i} f_i = - \sum_{j=1}^{N} w_{ij} \, \log_{R_i} R_j. \tag{324}$$

The set of gradient-type control systems to achieve consensus optimization within the fleet reads:

$$\dot{R}_i(t) = \sum_{j=1}^{N} w_{ij} \, \log_{R_i(t)} R_j(t), \; R_i(0) = R_i^0 \in \mathbb{SO}(3), \tag{325}$$

with $i = 1, 2, 3, \ldots, N$ and $t \geq 0$. (Notice that $\log_R R = 0$.) ■

9.4. Riemannian Gradient in Coordinates*

Let $f : \mathbb{M} \to \mathbb{R}$ denote a smooth manifold-to-real function, whose gradient at a point $x \in \mathbb{M}$ is sought. Let us fix a coordinate system on a local chart that includes the point x whose local coordinates are $x^1, x^2, \ldots x^p$.

In order to determine the gradient $\mathrm{grad}_x f$, let us make use of the metric compatibility condition, recalling that

- The Euclidean inner product reads $\langle \partial_x f, v \rangle = \frac{\partial f}{\partial x^k} v^k$.
- The inner product reads $\langle \mathrm{grad}_x f, v \rangle_x = g_{ik} (\mathrm{grad}_x f)^i v^k$.

The metric compatibility condition requires that

$$\frac{\partial f_x}{\partial x^k} v^k = g_{ik} (\mathrm{grad}_x f)^i v^k \; for \; every \; choice \; of \; v, \tag{326}$$

hence, $g_{ik} (\mathrm{grad}_x f)^i = \frac{\partial f_x}{\partial x^k}$. Multiplying both sides by g^{hk} one obtains $g^{hk} g_{ik} (\mathrm{grad}_x f)^i = g^{hk} \frac{\partial f_x}{\partial x^k}$. Since $g^{hk} g_{ik} = \delta_i^h$, we obtain

$$(\mathrm{grad}_x f)^h = g^{hk} \frac{\partial f_x}{\partial x^k}. \tag{327}$$

The Riemannian gradient in intrinsic coordinates is therefore expressed as

$$\mathrm{grad}_x f = g^{hk} \frac{\partial f_x}{\partial x^k} \partial_h^x. \tag{328}$$

In manifold calculus, the Riemannian gradient has yet another interpretation in relation to the differential of a manifold-to-real function and to so-called *musical isomorphisms*. In fact, there exist canonical operators to convert contravariant components to covariant (also referred to as *lowering an index*) or vice versa (*raising an index*).

Given a regular function $f : \mathbb{M} \to \mathbb{R}$, its differential $\mathrm{d}f : T\mathbb{M} \to T^*\mathbb{M}$ is expressed by:

$$\mathrm{d}f_x(v) = \langle \mathrm{grad}_x f, v \rangle_x, \ v \in T_x\mathbb{M}, \tag{329}$$

where $\mathrm{grad}_x f : \mathbb{M} \to T_x\mathbb{M}$ denotes its Riemannian gradient. Namely, the differential represents the linear part of the change of a function when moving away from a point x in the direction v. In local coordinates:

$$\mathrm{d}f_x(v) = g_{ij}(x)(\nabla_x f)^i v^j. \tag{330}$$

Since the differential does not depend on the choice of coordinates, it must hold that $\mathrm{d}f_x(v) = (\partial_i f)_x v^i$. It follows the metric-compatibility condition $g_{ij}(x)(\nabla_x f)^j = (\partial_i f)_x$. Such expressions may be interpreted through 'musical' isomorphisms:

- **Sharp isomorphism ($\sharp$)**: The 'sharp' isomorphism $\sharp : T^*\mathbb{M} \to TM$ takes a cotangent vector and returns a tangent vector. Namely, given a cotangent vector $\omega = \omega_i \, \mathrm{d}x^i$, the sharp isomorphism acts like $v := \omega^\sharp = g^{ij}\omega_i \partial_j$. For example, in the case of gradient, one may say that $\mathrm{grad}_x f = (df_x)^\sharp$.

- **Flat isomorphism ($\flat$)**: The 'flat' isomorphism $\flat : T\mathbb{M} \to T^*M$ takes a tangent vector and returns a cotangent vector. Namely, given a tangent vector $v = v^i \partial_i^x$, the flat isomorphism acts like $\omega := v^\flat = g_{ij}v^i\mathrm{d}x^j$. Then, the differential is the dual of gradient via a flat isomorphism.

10. Parallelism and Parallel Transport along a Curve

Parallel transport along a curve on a manifold is a fundamental concept that emerges from the curved nature of manifolds and is deeply interwoven in the fabric of manifold calculus.

10.1. Properties and Definition of Parallel Transport

Given a curve on a Riemannian manifold $\mathbb{M}$ embedded into a Euclidean space $\mathbb{A}$, a point $x \in \mathbb{M}$ on the curve and a tangent vector $v \in T_x\mathbb{M}$ to the curve, if such a vector is moved to another point on the curve by ordinary rigid translation (which is available in $\mathbb{A}$ and may be referred to as 'parallel translation'), in general it will not be tangent to the curve anymore. This observation suggests that it is necessary to define a notion of *transport* that is compatible with the structure of the manifold.

Another reason to motivate the notion of parallel transport is the lack of existence of 'uniform' vector fields on curved manifolds. On a flat space, such as $\mathbb{R}^p$, a uniform vector field $u : \mathbb{R}^p \to \mathbb{R}^p$ is a field $x \mapsto u_x$ such that u_x is constant (same direction, length and orientation) for every $x \in \mathbb{R}^p$. On a curved manifold embedded in an ambient space $\mathbb{A}$, however, it is unlikely that a tangent vector field may take the same value in two points, because tangency at one point would most likely imply lack of tangency in another point. The closest notion to uniformity would be *parallelism*: Informally speaking, a vector field is said to be parallel along a curve if it keeps the same orientation with respect to such a curve's own velocity field. Notice that this property is expressed in quite vague terms: such vagueness is indeed intended and implies that there exist infinitely many ways to interpret the notion of parallel transport.

Example 32. *Let $\mathbb{M}$ denote a Riemannian manifold and $x \mapsto w_x$ a tangent vector field on the tangent bundle $T\mathbb{M}$; namely, the function w assigns a tangent vector to each point x on the manifold. In the familiar case that $\mathbb{M} = \mathbb{R}^3$, there exists a notion of* uniform *vector field; namely, a vector field may take the same amplitude and direction in any point of the base-space $\mathbb{M}$. This is the case, for example, of an electric field within an ideal capacitor or an induction field within an ideal electrical coil. A distinguishing feature of a uniform vector field in $\mathbb{R}^3$ is that its directional derivative is zero everywhere; namely, given any point $x \in \mathbb{R}^3$ and any direction $v \in T_x\mathbb{R}^3$, it holds that*

$$\lim_{h \to 0} \frac{w_{x+hv} - w_x}{h} = 0, \tag{331}$$

where the left-hand side represents the directional derivative of the vector field $w \in \Gamma(\mathbb{M})$ along a direction v at the point x.

On a curved manifold $\mathbb{M}$, uniform vector fields hardly exist, as the rigid translation of a tangent vector to a different point of a manifold will most likely result in a non-tangent vector. The closest notion to uniformity that may be recovered on a curved manifold is that of parallelism. A parallel vector field needs to satisfy a generalized version of the condition (331) (which we shall not see in this survey since it requires covariant derivation, which will be covered in a separate survey). ∎

The notion of parallelism of a vector field gives rise to an important canonical operator in manifold calculus, namely, *parallel transport along a curve*. Parallel transport allows one to, e.g., compare vectors belonging to different tangent spaces. On a smooth Riemannian manifold $\mathbb{M}$, fix a smooth curve $\gamma : [0, 1] \to \mathbb{M}$. The parallel transport map associated with the curve γ is denoted as $\mathbb{P}_{\gamma}^{s \to t} : T_{\gamma(s)}\mathbb{M} \to T_{\gamma(t)}\mathbb{M}$, which is a linear map for every $s, t \in [0, 1]$, namely

$$\mathbb{P}_{\gamma}^{s \to t}(\alpha\, w + \beta\, v) = \alpha\, \mathbb{P}_{\gamma}^{s \to t}(w) + \beta\, \mathbb{P}_{\gamma}^{s \to t}(v) \tag{332}$$

for every $\alpha, \beta \in \mathbb{R}$ and $u, v \in T_{\gamma(s)}\mathbb{M}$. The parallel transport map depends smoothly on its arguments and is such that $\mathbb{P}_{\gamma}^{s \to s}$ coincides with the identity map and $\mathbb{P}_{\gamma}^{u \to t} \circ \mathbb{P}_{\gamma}^{s \to u} = \mathbb{P}_{\gamma}^{s \to t}$ for every $s, u, t \in [0, 1]$. (Notice that the last property holds on the same curve only, while it does not hold for different arcs.)

In most cases of importance in system theory and control, the main interest lies in the parallel transport of a tangent vector *along a geodesic curve* on a Riemannian manifold, rather than along a generic smooth curve. Parallel transport along a geodesic arc may be conjugated in at least two ways:

- **Parallel transport along a geodesic arc joining two given points**: Given two (sufficiently close) points $x, y \in \mathbb{M}$, and a tangent vector $w \in T_x\mathbb{M}$, the parallel transport of w from x to y along the geodesic arc connecting x to y is denoted by $\mathbb{P}^{x \to y}(w)$, since this notation shows all relevant information. In fact, the notation $\mathbb{P}^{x \to y}$ for the parallel transport operator is a shortened version of $\mathbb{P}_{\gamma_x^y}^{0 \to 1} : T_x\mathbb{M} \to T_y\mathbb{M}$, where $\gamma_x^y : [0, 1] \to \mathbb{M}$ would denote the geodesic arc such that $\gamma_x^y(0) = x$ and $\gamma_x^y(1) = y$.
- **Parallel transport along a geodesic specified by a point and a tangent direction**: Given a point $x \in \mathbb{M}$, and a tangent direction $v \in T_x\mathbb{M}$, the parallel transport of a vector $w \in T_x\mathbb{M}$ along the geodesic arc departing from the point x toward the direction v is denoted by $\mathbb{P}_{\gamma_{x,v}}^{0 \to t}(w)$, where $\gamma_{x,v} : [-\epsilon, \epsilon] \to \mathbb{M}$, with $\epsilon > 0$ and $t \in [-\epsilon, \epsilon]$, would denote the geodesic curve such that $\gamma_{x,v}(0) = x$ and $\dot{\gamma}_{x,v}(0) = v$.

There exist other special cases of interest in control and automation. One such special case supplies a partial answer to the problem of effecting parallel transport on a manifold for which the structure of the parallel transport operator is unknown. Another special case is parallel transport along a closed loop, which may arise when dealing with periodic trajectories or self-intersecting trajectories.

- **Self parallel transport along a geodesic arc**: A distinguishing feature of a geodesic arc is that it parallel-transports its own initial slope. On a setting where the parallel

transport operator is unknown in closed form, there exists no known way to transport a given vector w along a geodesic curve $\gamma_{x,v}$. However, it is possible to parallel-transport a given vector v along a geodesic $\gamma_{x,v}$, namely to compute $\mathbb{P}^{0 \to t}_{\gamma_{x,v}}(v)$: it coincides exactly with $\dot{\gamma}_{x,v}(t)$. Such a numerical trick was invoked, for instance, in [39], Subsection III.A.

- **Parallel transport along a closed loop**: Parallel transport of a vector along a (piece-wise continuous) *loop* ℓ is denoted as $\mathbb{P}_\ell : T_x \mathbb{M} \to T_x \mathbb{M}$, where x is termed *base* of the loop. In general, $\mathbb{P}_\ell(v) \neq v$, for $v \in T_x \mathbb{M}$. This phenomenon is referred to as *anholonomy*. Whenever $\mathbb{P}_\ell$ realizes an isometry, since $\|\mathbb{P}_\ell(v)\|_x = \|v\|_x$, the operator $\mathbb{P}_\ell$ changes only the orientation of v. For example, if $\mathbb{M} \subset \mathbb{R}^3$, then we may say that $\mathbb{P}_\ell(v)$ is a rotated version of a tangent vector v laying in the same tangent space, namely, $\mathbb{P}_\ell$ may be represented as an element of the orthogonal group $\mathbb{O}(3)$. Intuitively, holonomy is specifically related to curved spaces, and therefore holonomy must be related to curvature. This conjecture is made precise by the Ambrose–Singer theorem [60].

In the following, we shall go through a detailed derivation of the notion of parallel transport along a curve on the tangent bundle of a manifold embedded in an ambient space. In general, the result of parallel transport depends on the curve; moreover, parallel transport may be conceived in a way that preserves the angle between transported vectors but not their lengths, or to preserve both.

10.2. Coordinate-Free Derivation of Parallel Transport

Let us develop a notion of parallel transport $\mathbb{P}$ such that, in every point of the curve γ, it holds that $\mathbb{P}^{0 \to t}_\gamma(w) \in T_\gamma \mathbb{M}$. In addition, a fundamental requirement that parallel transport should fulfill is that, given two tangent vectors $u, w \in T_x \mathbb{M}$ that form an angle α, parallel transport should preserve the value of such angle along the whole curve, namely the angle between $\mathbb{P}^{0 \to t}_\gamma(u)$ and $\mathbb{P}^{0 \to t}_\gamma(w)$. Let us recall that the angle α between two vectors u, w in $T_x \mathbb{M}$ is defined by

$$\cos(\alpha) := \frac{\langle u, w \rangle_x}{\|u\|_x \|w\|_x}. \tag{333}$$

We may thus require parallel transport to fulfill:

1. **Tangency condition**: for every $t \in [0, 1]$ it should hold that $\mathbb{P}^{0 \to t}_\gamma(w) \in T_{\gamma(t)} \mathbb{M}$,
2. **Conformal isometry condition**: for every value of the parameter $t \in [0, 1]$ it should hold that $\langle \mathbb{P}^{0 \to t}_\gamma(u), \mathbb{P}^{0 \to t}_\gamma(w) \rangle_{\gamma(t)}$ keeps constant to $\langle u, w \rangle_x$ along the curve.

Let us verify that the second condition implies both conformality and isometry. To what concerns isometry, let us show that taking $w = u$ gives:

$$\langle \mathbb{P}^{0 \to t}_\gamma(u), \mathbb{P}^{0 \to t}_\gamma(u) \rangle_{\gamma(t)} = \|\mathbb{P}^{0 \to t}_\gamma(u)\|^2_{\gamma(t)}, \tag{334}$$

from which it follows that $\|\mathbb{P}^{0 \to t}_\gamma(u)\|_{\gamma(t)}$ keeps constant along the curve γ to $\|u\|_x$. It is straightforward to prove conformality; in fact,

$$\frac{\langle \mathbb{P}^{0 \to t}_\gamma(u), \mathbb{P}^{0 \to t}_\gamma(w) \rangle_{\gamma(t)}}{\|\mathbb{P}^{0 \to t}_\gamma(u)\|_{\gamma(t)} \|\mathbb{P}^{0 \to t}_\gamma(w)\|_{\gamma(t)}} = \frac{\langle u, w \rangle_x}{\|u\|_x \|w\|_x}, \ \forall t \in [0, 1]. \tag{335}$$

(Let us recall that a linear transformation that preserves angles is termed *conformal*.) It is important to underline that the above conditions do not define parallel transport univocally, but they represent minimal requirements for an operator to be qualified as parallel transport. In other words, parallel transport may be defined in several ways as long as it meets the above conditions, which turn out to be of great importance in non-linear control.

To ease notation, let us define the following fields:

- $\overline{u}(t) := \mathbb{P}_\gamma^{0 \to t}(u)$: a vector field $(\gamma(t), \overline{u}(t))$ that represents the evolution, over the curve γ, of the vector u that is tangent to the manifold at $\gamma(0)$ and after transport it will be tangent at $\gamma(1)$,
- $\overline{w}(t) := \mathbb{P}_{\gamma(t)}^{0 \to t}(w)$: analogous to $\overline{u}(t)$, both are *transport fields*,
- $\mathcal{Q}(t) := \langle \overline{u}(t), \overline{w}(t) \rangle_{\gamma(t)}$: a scalar field that represents the inner product between the above two transport fields along the curve γ.

For a manifold $\mathbb{M}$ embedded in an ambient space $\mathbb{A}$ endowed with a metric $\langle \cdot, \cdot \rangle^\mathbb{A}$, the condition of conformality and isometry may be written in a more detailed way. Let us recall from Section 7 that the inner product of two tangent vectors may be written as $\langle u, w \rangle_x = \langle u, G_x(w) \rangle^\mathbb{A}$, where the metric kernel has key properties, namely $G_x(v)$ is linear in v, G_x is self-adjoint (namely, $\langle u, G_x(w) \rangle^\mathbb{A} = \langle G_x(u), w \rangle^\mathbb{A}$), G_x is invertible and its inverse is denoted as G_x^{-1}, and G_x maps $T_x\mathbb{M}$ to itself.

In order to perform the following calculations seamlessly, it pays to highlight a few calculation rules about the metric kernel G and the fundamental form $\mathscr{F}$:

1. **Taylor series expansion of the kernel** G: Given $G_x(v)$ with $x \in \mathbb{M}$, $v \in T_x\mathbb{M}$, let us write

$$G_{\gamma_{x,w}(t)}(v) = G_x(v) + G_x^\bullet(v, t\,w) + \mathcal{O}(t^2), \tag{336}$$

where $\gamma_{x,w} : [-\epsilon, \epsilon] \to \mathbb{M}$ denotes any smooth curve such that $\gamma_{x,w}(0) = x$ and $\dot{\gamma}_{x,w}(0) = w$, with $w \in T_x\mathbb{M}$ denoting a direction along which a variation of the metric kernel is sought. The quantity $G_x^\bullet(v, w)$ represents a first-order derivative of the kernel G_x calculated with respect to the variable x. (Notation-wise, the 'bullet' derivative is a convenient way to denote a derivative that cannot be expressed in more specific terms. For example, if f, g are real functions of a real variable, one might denote $\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = \dot{f}(g(x))\dot{g}(x)$ as $f^\bullet(g(x), \dot{g}(x))$.) The formal definition goes like

$$G_x^\bullet(v, w) := \left. \frac{\mathrm{d}}{\mathrm{d}t} G_{\gamma_{x,w}(t)}(v) \right|_{t=0}, \tag{337}$$

where again $\gamma_{x,w} : [-\epsilon, \epsilon] \to \mathbb{M}$ denotes any smooth curve such that $\gamma_{x,w}(0) = x$ and $\dot{\gamma}_{x,w}(0) = w$. The function $G_x^\bullet(v, w)$ is linear in the arguments v and w and, in general, non-linear in x. Notice that the extendability property of G_x from $T_x\mathbb{M}$ to $\mathbb{A}$ holds also for the first argument of the derivative $G_x^\bullet$.

2. **Commutativity with the inner product in the ambient space**: In the expression $\langle u, G_x^\bullet(v, w) \rangle^\mathbb{A}$ it is allowed to swap the arguments u and v. In fact, notice that

$$\begin{aligned}
\langle u, G_x^\bullet(v, w) \rangle^\mathbb{A} &= \left\langle u, \left. \frac{\mathrm{d}}{\mathrm{d}t} G_{\gamma_{x,w}(t)}(v) \right|_{t=0} \right\rangle^\mathbb{A} = \left. \frac{\mathrm{d}}{\mathrm{d}t} \langle u, G_{\gamma_{x,w}(t)}(v) \rangle^\mathbb{A} \right|_{t=0} \\
&= \left. \frac{\mathrm{d}}{\mathrm{d}t} \langle G_{\gamma_{x,w}(t)}(u), v \rangle^\mathbb{A} \right|_{t=0} = \left\langle \left. \frac{\mathrm{d}}{\mathrm{d}t} G_{\gamma_{x,w}(t)}(u) \right|_{t=0}, v \right\rangle^\mathbb{A} \\
&= \langle v, G_x^\bullet(u, w) \rangle^\mathbb{A}.
\end{aligned} \tag{338}$$

This property of $G_x^\bullet$ holds even if the first argument of the derivative belongs to $\mathbb{A}$.

3. **Partial derivatives of the fundamental form**: Let us recall the definition of the fundamental form $\mathscr{F}(x, v) := \langle v, G_x(v) \rangle^\mathbb{A}$. Its partial derivatives read

$$\frac{\partial \mathscr{F}}{\partial x}(x, v) = G_x^\bullet(v, v), \ \frac{\partial \mathscr{F}}{\partial v}(x, v) = 2\,G_x(v). \tag{339}$$

Let us see a few examples to clarify the above definitions and properties.

Example 33. *Let us start considering the case of the manifold* $\mathbb{S}^+(n)$ *endowed with its canonical metric, embedded in the ambient space* $\mathbb{A} := \mathbb{R}^{n \times n}$ *endowed with a Euclidean metric. In this setting, we have seen that* $G_P(V) = P^{-1}VP^{-1}$. *By definition, its 'bullet derivative' reads*

$$G_P^\bullet(V, W) = \frac{\mathrm{d}}{\mathrm{d}t}(\gamma^{-1}V\gamma^{-1})\Big|_{t=0}, \tag{340}$$

where $\gamma : [-\epsilon, \epsilon] \to \mathbb{S}^+(n)$ *denotes any smooth curve such that* $\gamma(0) = P \in \mathbb{S}^+(n)$ *and* $\dot{\gamma}(0) = W \in T_P\mathbb{S}^+(n)$. *By using the noticeable matrix-flow derivation rule* (104), *we obtain*

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma^{-1}V\gamma^{-1}) = (-\gamma^{-1}\dot{\gamma}\gamma^{-1})V\gamma^{-1} + \gamma^{-1}V(-\gamma^{-1}\dot{\gamma}\gamma^{-1}). \tag{341}$$

Setting $t = 0$ *leads to the sought expression*

$$G_P^\bullet(V, W) = -P^{-1}WP^{-1}VP^{-1} - P^{-1}VP^{-1}WP^{-1}. \tag{342}$$

Let us verify the property of commutativity with the inner product. We may calculate that

$$\begin{aligned} \langle G_P^\bullet(V, W), U \rangle^\mathbb{A} &= -\operatorname{tr}(P^{-1}WP^{-1}VP^{-1}U) - \operatorname{tr}(P^{-1}VP^{-1}WP^{-1}U), \\ \langle G_P^\bullet(U, W), V \rangle^\mathbb{A} &= -\operatorname{tr}(P^{-1}UP^{-1}WP^{-1}V) - \operatorname{tr}(P^{-1}WP^{-1}UP^{-1}V). \end{aligned} \tag{343}$$

The two quantities may be proven equal by the cyclic permutation property of the trace.

Another example that we are surveying concerns the manifold $\mathrm{Sp}(2n)$ *endowed with its canonical metric again embedded in the ambient space* $\mathbb{A} := \mathbb{R}^{n \times n}$ *endowed with a Euclidean metric. In this setting, we have seen that* $G_Q(V) = Q^{-\top}Q^{-1}V$. *Its 'bullet derivative' reads*

$$G_Q^\bullet(V, W) = \frac{\mathrm{d}}{\mathrm{d}t}(\gamma^{-\top}\gamma^{-1}V)\Big|_{t=0}, \tag{344}$$

where $\gamma : [-\epsilon, \epsilon] \to \mathrm{Sp}(2n)$ *denotes any smooth curve such that* $\gamma(0) = Q \in \mathrm{Sp}(2n)$ *and* $\dot{\gamma}(0) = W \in T_Q\mathrm{Sp}(2n)$.

By invoking the identity (104) *once again, we obtain*

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma^{-\top}\gamma^{-1}) = (-\gamma^{-1}\dot{\gamma}\gamma^{-1})^\top\gamma^{-1} + \gamma^{-\top}(-\gamma^{-1}\dot{\gamma}\gamma^{-1}). \tag{345}$$

Setting $t = 0$ *leads to the expression*

$$G_Q^\bullet(V, W) = -Q^{-\top}(W^\top Q^{-\top} + Q^{-1}W)Q^{-1}V. \tag{346}$$

Let us verify the commutativity property with the inner product. We may calculate that

$$\begin{aligned} \langle U, G_P^\bullet(V, W) \rangle^\mathbb{A} &= -\operatorname{tr}(U^\top Q^{-\top}(W^\top Q^{-\top} + Q^{-1}W)Q^{-1}V), \\ \langle V, G_P^\bullet(U, W) \rangle^\mathbb{A} &= -\operatorname{tr}(V^\top Q^{-\top}(W^\top Q^{-\top} + Q^{-1}W)Q^{-1}U). \end{aligned} \tag{347}$$

These two quantities may be proven equal by transposing the arguments of the traces. ∎

The condition of conformality and isometry may be written very compactly as $\dot{\mathscr{Q}} = 0$. By multivariable calculus and the above definition and properties, we obtain:

$$\dot{\mathscr{Q}} = \langle \dot{\overline{u}}, G_\gamma(\overline{w}) \rangle^\mathbb{A} + \langle \overline{u}, G_\gamma^\bullet(\overline{w}, \dot{\gamma}) \rangle^\mathbb{A} + \langle \overline{u}, G_\gamma(\dot{\overline{w}}) \rangle^\mathbb{A}. \tag{348}$$

(This is one situation in which we need to invoke the 'extendability' property of the metric kernel; in fact, notice that $\dot{\overline{w}}$ does not necessarily belong to $T_\gamma\mathbb{M}$.) In the second

term on the right-hand side, the vector fields $\overline{u}$ and $\overline{w}$ may be swapped, and hence we may write:

$$\dot{\mathscr{Q}} = \langle \dot{\overline{u}}, G_\gamma(\overline{w}) \rangle^{\mathbb{A}} + \langle \overline{w}, G_\gamma^\bullet(\overline{u}, \dot{\gamma}) \rangle^{\mathbb{A}} + \langle \overline{u}, G_\gamma(\dot{\overline{w}}) \rangle^{\mathbb{A}}. \tag{349}$$

Since the above development is still too general to lead to a computable notion of parallel transport, let us refer parallelism to the velocity field associated with the curve γ, namely, let us set $\overline{w} := \dot{\gamma}$. This means that the parallel transport operation will keep fixed the angle between the transported vector and the velocity vector associated with the curve. The expression of $\dot{\mathscr{Q}}$ becomes then:

$$\dot{\mathscr{Q}} = \langle \dot{\overline{u}}, G_\gamma(\dot{\gamma}) \rangle^{\mathbb{A}} + \langle \dot{\gamma}, G_\gamma^\bullet(\overline{u}, \dot{\gamma}) \rangle^{\mathbb{A}} + \langle \overline{u}, G_\gamma(\ddot{\gamma}) \rangle^{\mathbb{A}}. \tag{350}$$

In the above relation, all quantities are known except for the transport field $\overline{u}$ of which we are seeking a temporal evolution law.

A special case, of sure interest in applications since it dispenses us from choosing a curve γ, is parallel transport along a geodesic line. Let thus γ denote a geodesic and let us recall that, in this case, the expression $G_\gamma(\ddot{\gamma})$ may be rewritten in terms of the velocity $\dot{\gamma}$ through the differential inclusion that characterizes a geodesic line. Let us then revamp the geodesic equation with the aim of writing it in a *Christoffel normal form*. Recall that, given the fundamental form for a line $\mathscr{F}(\gamma, \dot{\gamma}) = \langle \dot{\gamma}, G_\gamma(\dot{\gamma}) \rangle^{\mathbb{A}}$, whenever it is a geodesic line it holds that:

$$\frac{\partial \mathscr{F}}{\partial \gamma} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathscr{F}}{\partial \dot{\gamma}} \in N_\gamma \mathbb{M}. \tag{351}$$

According to the relationships in (339)

$$\frac{\partial \mathscr{F}}{\partial \gamma} = G_\gamma^\bullet(\dot{\gamma}, \dot{\gamma}), \quad \frac{\partial \mathscr{F}}{\partial \dot{\gamma}} = 2\, G_\gamma(\dot{\gamma}), \tag{352}$$

and then

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathscr{F}}{\partial \dot{\gamma}} = 2\, G_\gamma^\bullet(\dot{\gamma}, \dot{\gamma}) + 2\, G_\gamma(\ddot{\gamma}), \tag{353}$$

therefore the differential inclusion (351) that defines a geodesic line becomes

$$- G_\gamma^\bullet(\dot{\gamma}, \dot{\gamma}) - 2\, G_\gamma(\ddot{\gamma}) = B_\gamma(\dot{\gamma}, \dot{\gamma}), \tag{354}$$

where $B_x : (T_x\mathbb{M})^2 \to N_x\mathbb{M}$ is a bilinear function of its vector arguments and returns an element of the normal space $N_x\mathbb{M}$. Ultimately, we obtained

$$G_\gamma(\ddot{\gamma}) = -\tfrac{1}{2} G_\gamma^\bullet(\dot{\gamma}, \dot{\gamma}) - \tfrac{1}{2} B_\gamma(\dot{\gamma}, \dot{\gamma}). \tag{355}$$

This same expression may be rewritten compactly in one of the following *normal Christoffel forms*:

$$\begin{cases} G_\gamma(\ddot{\gamma}) + \bar{\Gamma}_\gamma(\dot{\gamma}, \dot{\gamma}) = 0, \text{ or} \\ \ddot{\gamma} + \bar{\bar{\Gamma}}_\gamma(\dot{\gamma}, \dot{\gamma}) = 0, \end{cases} \tag{356}$$

where the function $\bar{\Gamma}_x$ is termed *Christoffel form of the* 1st *kind*, while the quantity $\bar{\bar{\Gamma}}_x := G_x^{-1} \circ \bar{\Gamma}_x$ is termed *Christoffel form of the* 2nd *kind*. Notice that only the restricted Christoffel form $\bar{\Gamma}_x(v, v)$ with $x \in \mathbb{M}$ and $v \in T_x\mathbb{M}$ has been defined so far, namely

$$\bar{\Gamma}_x(v, v) := \tfrac{1}{2} G_x^\bullet(v, v) + \tfrac{1}{2} B_x(v, v). \tag{357}$$

Since both terms on the right-hand side are bilinear in their vector arguments, the Christoffel form is quadratic. It is now necessary to define the full Christoffel form $\Gamma_x(u, w)$, with $x \in \mathbb{M}$ and $u, w \in T_x\mathbb{M}$ on the basis of its restricted version. Such a result may be achieved through the following *polarization formula*:

$$\bar{\Gamma}_x(u, w) := \tfrac{1}{4} \{ \bar{\Gamma}_x(u + w, u + w) - \bar{\Gamma}_x(u - w, u - w) \}, \tag{358}$$

which stems from the requirements that the full Christoffel form is *bilinear* and *symmetric* in its vector argument, namely $\bar{\Gamma}_x(u, w) = \bar{\Gamma}_x(w, u)$.

Employing the normal Christoffel form in the latest expression of $\dot{\mathcal{Q}}$ gives:

$$\dot{\mathcal{Q}} = \langle \dot{\bar{u}}, G_\gamma(\dot{\gamma}) \rangle^{\mathbb{A}} + \langle \dot{\gamma}, G_\gamma^\bullet(\bar{u}, \dot{\gamma}) \rangle^{\mathbb{A}} - \langle \bar{u}, \bar{\Gamma}_\gamma(\dot{\gamma}, \dot{\gamma}) \rangle^{\mathbb{A}}. \tag{359}$$

Since $\bar{\Gamma}$ is derived by polarization from a restricted bilinear function, it is self-adjoint in both its arguments; hence, it commutes with the inner product $\langle \cdot, \cdot \rangle^{\mathbb{A}}$. Therefore, in the third term on the right hand side it is possible to swap the transport field $\bar{u}$ with any of the two fields $\dot{\gamma}$, which leads to:

$$\begin{aligned} \dot{\mathcal{Q}} &= \langle \dot{\bar{u}}, G_\gamma(\dot{\gamma}) \rangle^{\mathbb{A}} + \langle \dot{\gamma}, G_\gamma^\bullet(\bar{u}, \dot{\gamma}) \rangle^{\mathbb{A}} - \langle \dot{\gamma}, \bar{\Gamma}_\gamma(\bar{u}, \dot{\gamma}) \rangle^{\mathbb{A}} \\ &= \langle \dot{\bar{u}}, G_\gamma(\dot{\gamma}) \rangle^{\mathbb{A}} + \langle \dot{\gamma}, \bar{\Gamma}_\gamma(\bar{u}, \dot{\gamma}) \rangle^{\mathbb{A}} \end{aligned} \tag{360}$$

where we have used the fact that $\bar{\Gamma}$ and $\frac{1}{2}G^\bullet$ differ by a purely normal component.

In order to make the above addenda uniform to one another, let us introduce in the above expression twice the identity $G_\gamma \circ G_\gamma^{-1}(= \mathrm{id}_\gamma)$ obtaining:

$$\dot{\mathcal{Q}} = \langle G_\gamma(\dot{\gamma}), \dot{\bar{u}} \rangle^{\mathbb{A}} + \langle G_\gamma(\dot{\gamma}), G_\gamma^{-1} \circ \bar{\Gamma}_\gamma(\bar{u}, \dot{\gamma}) \rangle^{\mathbb{A}}.$$

Now it is possible to gather the terms in $G_\gamma(\dot{\gamma})$ and to write the above sum as a single inner product:

$$\dot{\mathcal{Q}} = \langle G_\gamma(\dot{\gamma}), \dot{\bar{u}} + G_\gamma^{-1} \circ \bar{\Gamma}_\gamma(\bar{u}, \dot{\gamma}) \rangle^{\mathbb{A}} = 0. \tag{361}$$

Ultimately, the last relationship still expresses the basic property of parallel transport to realize a conformal isometry, which still leaves much room open for a precise definition of parallel transport. The condition $\dot{\mathcal{Q}} = 0$ alone would give rise to infinitely many equations to compute a transport field, and hence parallel transport. Among these infinitely many, one arises simply by requiring the right-hand side of the above inner product to vanish to zero. Namely:

$$\dot{\bar{u}} + \bar{\bar{\Gamma}}_\gamma(\bar{u}, \dot{\gamma}) = 0. \tag{362}$$

The solution of such equation supplies an expression for the transport field which, in turn, supplies a notion of parallel transport along a geodesic line, namely

$$\mathbb{P}^{x \to y}(u) := \bar{u}(1), \tag{363}$$

where $\bar{u}(t)$ denotes the solution of the differential Equation (363) with the initial condition $\bar{u}(0) = u$. Such a differential equation in $\mathbb{A}$ is first-order and linear in the unknown transport field. Since the set of solutions to a linear first-order differential equation is a linear space, parallel transport is a linear isomorphism of the tangent bundle $T\mathbb{M}$.

Example 34. *As anticipated, a noteworthy property of any geodesic line is that it self-transports its own velocity along. In other words, given a geodesic $\gamma_x^y : [0, 1] \to \mathbb{M}$ that connects two points $x, y \in \mathbb{M}$, the following function turns out to realize parallel transport:*

$$\mathbb{P}_{\gamma_x^y}^{0 \to t}(\dot{\gamma}_x^y(0)) := \dot{\gamma}_x^y(t). \tag{364}$$

Let us verify that it meets the fundamental properties any parallel transport should meet:

1. ***Tangency****: for every $t \in [0, 1]$ it holds that $\mathbb{P}_{\gamma_x^y}^{0 \to t}(\dot{\gamma}_x^y(0)) \in T_{\gamma_x^y(t)}\mathbb{M}$; in fact, $\mathbb{P}_{\gamma_x^y}^{0 \to t}(\dot{\gamma}_x^y(0)) = \dot{\gamma}_x^y(t)$ is tangent to the curve at every point $\gamma_x^y(t)$.*
2. ***Conformal isometry****: for every $t \in [0, 1]$ it holds that $\langle \mathbb{P}_{\gamma_x^y}^{0 \to t}(\dot{\gamma}_x^y(0)), \mathbb{P}_{\gamma_x^y}^{0 \to t}(\dot{\gamma}_x^y(0)) \rangle_{\gamma_x^y(t)} = \|\dot{\gamma}_x^y(t)\|^2_{\gamma_x^y(t)}$ which is constant along a geodesic,*

as claimed. ∎

As a special case, it is worth considering a manifold $\mathbb{M}$ endowed with a uniform metric. In this instance, the metric kernel G_x is independent of the point x, and therefore it holds that $G^\bullet \equiv 0$ along every (geodesic) line. The transport field equation hence simplifies into

$$\dot{\bar{u}} = -\tfrac{1}{2} G^{-1} \circ B_\gamma(\bar{u}, \dot{\gamma}). \tag{365}$$

In the ever simpler instance where the metric is Euclidean, the metric kernel coincides with the identity (namely $G_x \equiv \mathrm{id}_x$), and hence it turns out that

$$\dot{\bar{u}} = -\tfrac{1}{2} B_\gamma(\bar{u}, \dot{\gamma}). \tag{366}$$

Recalling that the function B maps the double tangent bundle $(T\mathbb{M})^2$ to the normal bundle $N\mathbb{M}$, it is quite straightforward to figure out a practical interpretation of the differential Equation (366). It basically prescribes that, when traveling from a point over the curve γ to another infinitesimally close point, the transport field looses its normal component to stay tangent.

The following two examples aim at clarifying the above theoretical developments.

Example 35. *Let us apply the theoretical development just analyzed to determine a transport field equation on a hypersphere $\mathbb{S}^{n-1} \subset \mathbb{A}$ with $\mathbb{A} := \mathbb{R}^n$ endowed with the metric $\langle u, w \rangle_x := \langle u, w \rangle^{\mathbb{A}} = u^\top w$.*

Let us start by determining the Christoffel form (of the first or second kind is unessential, in this case, because the metric kernel coincides with the identity). The easiest way to determine Christoffel forms is via a geodesic equation. The fundamental form, in this case, reads $\mathscr{F} := \dot{\gamma}^\top \dot{\gamma}$. From the fundamental form we find that

$$\frac{\partial \mathscr{F}}{\partial \gamma} = 0 \text{ and } \frac{\partial \mathscr{F}}{\partial \dot{\gamma}} = 2\,\dot{\gamma}, \tag{367}$$

therefore

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathscr{F}}{\partial \dot{\gamma}} = 2\,\ddot{\gamma}. \tag{368}$$

The geodesic equation then reads $\ddot{\gamma} = \lambda\,\gamma$; hence, $\lambda = \ddot{\gamma}^\top \gamma$. From the tangency of the velocity $\dot{\gamma}$, it follows that:

$$\dot{\gamma}^\top \gamma = 0 \overset{\mathrm{d}/\mathrm{d}t}{\Rightarrow} \ddot{\gamma}^\top \gamma + \dot{\gamma}^\top \dot{\gamma} = 0, \tag{369}$$

and henceforth $\ddot{\gamma}^\top \gamma = -\|\dot{\gamma}\|^2$. As a consequence, $\lambda = -\|\dot{\gamma}\|^2$ and:

$$\bar{\Gamma}_x(u, w) = \bar{\bar{\Gamma}}_x(u, w) = -(u^\top w)x. \tag{370}$$

We may now verify that $\langle \bar{\Gamma}_x(u, w), v \rangle^{\mathbb{A}} = \langle \bar{\Gamma}_x(v, w), u \rangle^{\mathbb{A}}$. It holds that

$$\begin{aligned} \langle \bar{\Gamma}_x(u, w), v \rangle^{\mathbb{A}} &= -(u^\top w)(v^\top x) = 0, \\ \langle \bar{\Gamma}_x(v, w), u \rangle^{\mathbb{A}} &= -(v^\top w)(u^\top x) = 0, \end{aligned} \tag{371}$$

hence the assertion.

The equation of the transport field, in this case, reads as:

$$\dot{\bar{u}} + (\bar{u}^\top \dot{\gamma})\gamma = 0. \tag{372}$$

Such differential equation may be solved and returns an expression for the transport field $\bar{u}$. The definition (363) returns then the expression for the parallel transport from the point $x \in \mathbb{S}^{n-1}$ of a vector $w \in T_x \mathbb{S}^{n-1}$ to the point $y \in \mathbb{S}^{n-1}$:

$$\mathbb{P}^{x \to y}(w) := \left[I_n - \frac{(x + y)y^\top}{1 + x^\top y} \right] w. \tag{373}$$

Let us verify some characteristic features of the above map:

- **Linearity**: The linearity of $\mathbb{P}^{x\to y}(w)$ with respect to w (but not with respect to x ad y) is quite apparent.
- **Identity**: Letting $y \equiv x$ leads to an identity map. In fact, it holds that $\mathbb{P}^{x\to x}(w) = \left[I_n - \frac{(x+x)x^\top}{1+x^\top x}\right]w = \left[I_n - \frac{2xx^\top}{2}\right]w = (I_n - xx^\top)w = w - x(x^\top w) = w - 0 = w.$
- **Tangency**: It should hold that $y^\top \mathbb{P}^{x\to y}(w) = 0$. In fact: $y^\top \left[I_n - \frac{(x+y)y^\top}{1+x^\top y}\right]w = y^\top w - \frac{y^\top(x+y)y^\top w}{1+x^\top y} = y^\top w - \frac{(y^\top x+1)y^\top w}{1+x^\top y} = y^\top w - y^\top w = 0.$

Important properties of the map (373) *have been hence verified.* ■

Example 36. *Let us consider the case of the manifold $\mathbb{S}^+(n) \subset \mathbb{A}$ endowed with the metric $\langle U, W\rangle_P := \mathrm{tr}(UP^{-1}WP^{-1})$ embedded in the ambient space $\mathbb{A} := \mathbb{R}^{n\times n}$ endowed with the metric $\langle U, V\rangle^\mathbb{A} := \mathrm{tr}(UV)$. Recall from Section 7 that the corresponding metric kernel takes the expression $G_P(W) := P^{-1}WP^{-1}$ (which, as a function of W, is well-defined in the whole ambient space $\mathbb{A}$).*

Let us determine the Christoffel form of the second kind through calculations related to geodesy. Define the fundamental form on a curve as $\mathscr{F}(\gamma, \dot\gamma) := \mathrm{tr}(\dot\gamma\gamma^{-1}\dot\gamma\gamma^{-1})$ and observe that:

$$\frac{\partial\mathscr{F}}{\partial\dot\gamma} = 2\gamma^{-1}\dot\gamma\gamma^{-1} \ [\equiv 2\,G_\gamma(\dot\gamma)], \tag{374}$$

$$\frac{\partial\mathscr{F}}{\partial\gamma} = -2\,\gamma^{-1}\dot\gamma\gamma^{-1}\dot\gamma\gamma^{-1} \ [\equiv G_\gamma^\bullet(\dot\gamma, \dot\gamma)], \tag{375}$$

from which it follows that:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathscr{F}}{\partial\dot\gamma} = 2\frac{\mathrm{d}}{\mathrm{d}t}(\gamma^{-1})\dot\gamma\gamma^{-1} + 2\gamma^{-1}\ddot\gamma\gamma^{-1} + 2\gamma^{-1}\dot\gamma\frac{\mathrm{d}}{\mathrm{d}t}(\gamma^{-1}). \tag{376}$$

Recalling the matrix identity

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma^{-1}) = -\gamma^{-1}\dot\gamma\gamma^{-1},$$

leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathscr{F}}{\partial\dot\gamma} = -4\,\gamma^{-1}\dot\gamma\gamma^{-1}\dot\gamma\gamma^{-1} + 2\,\gamma^{-1}\ddot\gamma\gamma^{-1}, \tag{377}$$

which, in turn, implies that:

$$\begin{aligned}\frac{\partial\mathscr{F}}{\partial\gamma} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathscr{F}}{\partial\dot\gamma} &= -2\,\gamma^{-1}\dot\gamma\gamma^{-1}\dot\gamma\gamma^{-1} + 4\,\gamma^{-1}\dot\gamma\gamma^{-1}\dot\gamma\gamma^{-1} - 2\,\gamma^{-1}\ddot\gamma\gamma^{-1}, \\ &= 2\,\gamma^{-1}\dot\gamma\gamma^{-1}\dot\gamma\gamma^{-1} - 2\,\gamma^{-1}\ddot\gamma\gamma^{-1}, \\ &= 2\,\gamma^{-1}(\dot\gamma\gamma^{-1}\dot\gamma - \ddot\gamma)\gamma^{-1}.\end{aligned} \tag{378}$$

The transport equation, in normal Christoffel form, hence reads:

$$\ddot\gamma - \dot\gamma\gamma^{-1}\dot\gamma - \omega\gamma = 0, \tag{379}$$

where ω denotes a skew-symmetric matrix function (namely $\omega^\top = -\omega$) such that the product $\omega\gamma$ belongs to the normal space at γ.

In order to determine a function ω, let us invoke the condition of tangency of the velocity field $\dot\gamma$ that, in this case, is expressed as $\dot\gamma^\top = \dot\gamma$. From such a condition, it follows that $\ddot\gamma^\top = \ddot\gamma$; hence:

$$(\ddot\gamma - \dot\gamma\gamma^{-1}\dot\gamma - \omega\gamma)^\top = \ddot\gamma - \dot\gamma\gamma^{-1}\dot\gamma - \gamma\omega^\top = 0, \tag{380}$$

from which it follows that $\omega = 0$. The restricted Christoffel form of the second kind, in this case, hence reads as

$$\bar{\bar{\Gamma}}_P(V, V) := -VP^{-1}V. \tag{381}$$

To determine the complete Christoffel form of the second kind, let us make use of the polarization formula (358):

$$
\begin{aligned}
\bar{\bar{\Gamma}}_P(V, W) &:= \tfrac{1}{4}\left\{\bar{\bar{\Gamma}}_P(V + W, V + W) + \bar{\bar{\Gamma}}_P(V - W, V - W)\right\} \\
&= \tfrac{1}{4}\left\{-(V + W)P^{-1}(V + W) + (V - W)P^{-1}(V - W)\right\} \\
&= -\tfrac{1}{2}VP^{-1}W - \tfrac{1}{2}WP^{-1}V.
\end{aligned} \tag{382}
$$

Let us now verify that the form $\bar{\Gamma}$ commutes with the ambient inner product. We have $\bar{\Gamma} = G \circ \bar{\bar{\Gamma}}$; therefore,

$$\bar{\Gamma}_P(V, W) = -\tfrac{1}{2}P^{-1}(VP^{-1}W + WP^{-1}V)P^{-1}. \tag{383}$$

As a consequence, we have

$$
\begin{aligned}
\langle \bar{\Gamma}_P(V, W), U \rangle^{\mathbb{A}} &= -\tfrac{1}{2}\mathrm{tr}(UP^{-1}VP^{-1}WP^{-1} + UP^{-1}WP^{-1}VP^{-1}), \\
\langle \bar{\Gamma}_P(U, W), V \rangle^{\mathbb{A}} &= -\tfrac{1}{2}\mathrm{tr}(VP^{-1}UP^{-1}WP^{-1} + VP^{-1}WP^{-1}UP^{-1}),
\end{aligned} \tag{384}
$$

which may be easily proven equal by the cyclic commutation property of matrix trace.

As a consequence, and employing the relation (342), *the parallel transport equation is found to be:*

$$\dot{\overline{U}} + \gamma(-\gamma^{-1}\dot{\gamma}\gamma^{-1}\overline{U}\gamma^{-1} - \gamma^{-1}\overline{U}\gamma^{-1}\dot{\gamma}\gamma^{-1})\gamma - \left(-\tfrac{1}{2}\dot{\gamma}\gamma^{-1}\overline{U} - \tfrac{1}{2}\overline{U}\gamma^{-1}\dot{\gamma}\right) = 0, \tag{385}$$

or, after a few straightforward simplifications,

$$\dot{\overline{U}} - \tfrac{1}{2}\left(\dot{\gamma}\gamma^{-1}\overline{U} + \overline{U}\gamma^{-1}\dot{\gamma}\right) = 0. \tag{386}$$

Through the definition (363), *we may obtain the expression of the parallel transport from a point $P \in \mathbb{S}^+(n)$ of a tangent vector $W \in T_P\mathbb{S}^+(n)$ to a point $Q \in \mathbb{S}^+(n)$:*

$$\mathbb{P}^{P \to Q}(W) := \sqrt{QP^{-1}}W\sqrt{P^{-1}Q}, \tag{387}$$

where the matrix square root $\sqrt{\cdot}$ returns a symmetric matrix. Let us verify two properties of the found operator:

- ***Identity**: Setting $Q \equiv P$ leads to parallel transport collapse to an identity map. Indeed, it holds that $\mathbb{P}^{P \to P}(W) := \sqrt{PP^{-1}}W\sqrt{P^{-1}P} = W$.*
- ***Tangency**: It holds that $(\mathbb{P}^{P \to Q}(W))^{\top} = \mathbb{P}^{P \to Q}(W)$. In fact, we have $(\mathbb{P}^{P \to Q}(W))^{\top} = (P^{-1}Q)^{\frac{\top}{2}}W^{\top}(QP^{-1})^{\frac{\top}{2}} = (Q^{\top}P^{-\top})^{\frac{1}{2}}W(P^{-\top}Q^{\top})^{\frac{1}{2}} = \sqrt{QP^{-1}}W\sqrt{P^{-1}Q} = \mathbb{P}^{P \to Q}(W)$.*

Linearity is apparent as well. ∎

The calculations illustrated in the previous example requires the evaluation of the partial derivatives of the fundamental form $\mathscr{F}$, determined in (374) and (375). It is quite instructive to sketch such calculations.

Example 37. *Let us survey in detail the calculations that led to relations* (374) *and* (375). *Such calculations may be carried out through calculus (exact way) or by analytic approximations (which is more clear from a computational viewpoint).*

To begin with, it is useful to analytically justify a matrix approximation, namely

$$(P + E)^{-1} \approx P^{-1} - P^{-1}EP^{-1}, \tag{388}$$

with $P \in \mathbb{S}^+(n)$ and $E \in \mathbb{R}^{n \times n}$ with $\|E\|$ small. Such approximation may be proven through the notion of analytic matrix function. *By definition, a matrix-to-matrix function $f : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ is analytic in a point $X_0 \in \mathbb{R}^{n \times n}$ if it admits, in a neighborhood of such a point, a polynomial series expansion, namely:*

$$f(X) = \sum_{n=0}^{\infty} a_n (X - X_0)^n, \text{ for } \|X - X_0\| < \rho \tag{389}$$

where $a_n \in \mathbb{R}$ are coefficients of the series and $\rho > 0$ denotes the radius of convergence of the polynomial series.

Let us now consider the function $f(X) := (I_n + X)^{-1}$. It is analytic in a neighborhood of the point $X = 0_n$ and the associated polynomial series, truncated to the second term, reads:

$$(I_n + X)^{-1} \approx I_n - X, \tag{390}$$

from which it readily follows that

$$
\begin{aligned}
(P + E)^{-1} &= (P(I_n + P^{-1}E))^{-1} \\
&= (I_n + P^{-1}E)^{-1}P^{-1} \\
&\approx (I_n - P^{-1}E)P^{-1} \\
&= P^{-1} - P^{-1}EP^{-1}.
\end{aligned} \tag{391}
$$

To justify the partial derivatives (374) and (375), let us define $\mathscr{F}(P, V) := \mathrm{tr}(VP^{-1}VP^{-1})$ and evaluate its partial derivatives with respect to P and V. The partial derivative of a matrix-to-scalar function may be thought of as the 'coefficient' of the linear term arising from an additive perturbation.

Applying such a practical idea to evaluate the partial derivative with respect to the first argument gives:

$$\mathscr{F}(P + E, V) = \mathscr{F}(P, V) + \left\langle \frac{\partial \mathscr{F}}{\partial P}, E \right\rangle^{\mathbb{A}} + \textit{higher order terms}, \tag{392}$$

where E denotes a matrix (termed perturbation*) whose entries are small numbers (such that $\|E\|$ is very small). In the present case, we have:*

$$
\begin{aligned}
\mathscr{F}(P + E, V) &= \mathrm{tr}(V(P + E)^{-1}V(P + E)^{-1}), \\
&\approx \mathrm{tr}(V(P^{-1} - P^{-1}EP^{-1})V(P^{-1} - P^{-1}EP^{-1})) \\
&\approx \mathrm{tr}(VP^{-1}VP^{-1}) - \mathrm{tr}(VP^{-1}EP^{-1}VP^{-1}) - \mathrm{tr}(VP^{-1}VP^{-1}EP^{-1}) \\
&\quad + \textit{higher order terms} \\
&= \mathscr{F}(P, V) - \mathrm{tr}(P^{-1}VP^{-1}VP^{-1}E) - \mathrm{tr}(P^{-1}VP^{-1}VP^{-1}E) \\
&\quad + \textit{higher order terms} \\
&= \mathscr{F}(P, V) - 2\,\mathrm{tr}(P^{-1}VP^{-1}VP^{-1}E) + \textit{higher order terms}.
\end{aligned} \tag{393}
$$

Comparing this expression with (392), one obtains the sought result, namely:

$$\frac{\partial \mathscr{F}}{\partial P} = -2\,P^{-1}VP^{-1}VP^{-1}. \tag{394}$$

To what concerns the second argument, we may write that:

$$\mathscr{F}(P, V + E) = \mathscr{F}(P, V) + \left\langle \frac{\partial \mathscr{F}}{\partial V}, E \right\rangle^{\mathbb{A}} + \text{higher order terms,} \tag{395}$$

where E again denotes a perturbation. In this case, we have:

$$
\begin{aligned}
\mathscr{F}(P, V + E) &= \text{tr}((V + E)P^{-1}(V + E)P^{-1}), \\
&= \text{tr}(VP^{-1}VP^{-1}) + \text{tr}(EP^{-1}VP^{-1}) + \text{tr}(VP^{-1}EP^{-1}) \\
&\quad + \text{higher order terms} \\
&= \mathscr{F}(P, V) + \text{tr}(P^{-1}VP^{-1}E) + \text{tr}(P^{-1}VP^{-1}E) \\
&\quad + \text{higher order terms} \\
&= \mathscr{F}(P, V) + 2\,\text{tr}(P^{-1}VP^{-1}E) + \text{higher order terms.}
\end{aligned}
\tag{396}
$$

Comparing such an expression with (395) *yields:*

$$\frac{\partial \mathscr{F}}{\partial V} = 2\, P^{-1}VP^{-1}, \tag{397}$$

which is the sought result. ∎

Let us summarizes a few known formulas about parallel transport. All the formulas given below hold under the provision that the manifolds are endowed with their canonical metrics.

- **Hypercube**: The hypercube endowed with the standard Euclidean metric is a flat space; hence, parallel transport may be realized as a rigid translation. Namely parallel transport is an identity map.
- **Hypersphere**: Parallel transport on the hypersphere $\mathbb{S}^{p-1}$ of the tangent vector $w \in T_x\mathbb{S}^{p-1}$ along the geodesic arc $\gamma_{x,v}$ of an extent t may be computed by [39]:

$$\mathbb{P}_{\gamma_{x,v}}^{0 \to t}(w) = \left[I_p + \frac{\cos(\|v\|t) - 1}{\|v\|^2} vv^\top - \sin(\|v\|t)xv^\top \right] w. \tag{398}$$

For a reference, readers might want to consult [61]. Starting from this, some mathematical work leads to the following result: given two points $x, y \in \mathbb{S}^{p-1}$ and a vector $w \in T_x\mathbb{S}^{n-1}$, the parallel transport operator $\mathbb{P}^{x \to y} : T_x\mathbb{S}^{p-1} \to T_y\mathbb{S}^{p-1}$ has the following structure:

$$\mathbb{P}^{x \to y}(w) = \left[I_p - \frac{(x + y)y^\top}{1 + x^\top y} \right] w, \tag{399}$$

provided that $x^\top y \neq -1$ (namely, the points x and y are not antipodal one to another).

- **Special orthogonal group**: Parallel transport along a geodesic curve on the special orthogonal group $\mathbb{SO}(p)$ may be implemented through the following formula:

$$\mathbb{P}^{X \to Y}(W) = X\sqrt{X^\top Y}X^\top W\sqrt{X^\top Y}. \tag{400}$$

- **Stiefel manifold**: To the best of this author's knowledge, there appear to be no closed formulas about parallel transport on a Stiefel manifold $\text{St}(n, p)$. This is one of those cases in which self-parallel transport might be invoked. A self-parallel transport expression, corresponding to the canonical metric, is

$$\mathbb{P}_{\gamma_{X,V}}^{0 \to t}(V) = [V - XR^\top] \, \text{Exp}\left(t \begin{bmatrix} X^\top V & -R^\top \\ R & 0_p \end{bmatrix} \right) \begin{bmatrix} I_p \\ 0_p \end{bmatrix}, \tag{401}$$

where R denotes the R-factor of the thin QR factorization of the matrix $(I_n - XX^\top)V$.

- **Manifold of symmetric, positive-definite (SPD) matrices**: Given a geodesic arc $\gamma_{X,V} : [0,1] \to \mathbb{S}^+(p)$ and a tangent vector $W \in T_X\mathbb{S}^+(p)$, the expression of the parallel transport operator that shifts the tangent vector W along the geodesic arc $\gamma_{X,V}(t)$ toward the endpoint $t = 1$ reads:

$$\mathbb{P}_{\gamma_{X,V}}^{0 \to 1}(W) = \text{Exp}\left(\frac{VX^{-1}}{2}\right) W \, \text{Exp}\left(\frac{X^{-1}V}{2}\right). \tag{402}$$

  Some mathematical work leads to the following result: given two points $X, Y \in \mathbb{S}^+(p)$ and a tangent vector $W \in T_X\mathbb{S}^+(p)$, the parallel transport operator $\mathbb{P}^{X \to Y} : T_X\mathbb{S}^+(p) \to T_Y\mathbb{S}^+(p)$ has the following structure:

$$\mathbb{P}^{X \to Y}(W) = \sqrt{YX^{-1}} W \sqrt{X^{-1}Y}. \tag{403}$$

- **Grassmann manifold**: Parallel transport of a vector $U \in T_{[X]}\text{Gr}(n, p)$ along the geodesic $\gamma_{[X],V}$ with $V \in T_{[X]}\text{Gr}(n, p)$ is given by [33]:

$$\mathbb{P}_{\gamma_{[X],V}}^{0 \to t}(U) = \left([XB\ A] \begin{bmatrix} -\sin(Dt) \\ \cos(Dt) \end{bmatrix} A^\top + I_n - AA^\top \right) U, \tag{404}$$

  where $ADB^\top$ denotes the compact singular value decomposition of the matrix V.

  Let us derive the expression of the Christoffel form of the second kind for the special orthogonal group.

Example 38. *According to the canonical metric for the special orthogonal group $\mathbb{SO}(n)$, the fundamental form is $\mathscr{F}(R, V) := \text{tr}(V^\top V)$; hence, the geodesic equation stemming from the differential inclusion reads as $\ddot{\gamma} = \gamma S$, where S is a symmetric function. Recalling that any geodesic line must satisfy the constraint $\gamma^\top \gamma = I_n$, deriving this twice with respect to time gives*

$$\ddot{\gamma}^\top \gamma + 2\dot{\gamma}^\top \dot{\gamma} + \gamma^\top \ddot{\gamma} = 0. \tag{405}$$

Plugging in the geodesic equation leads to $S = -\dot{\gamma}^\top \dot{\gamma}$, and hence the complete geodesic equation reads as $\ddot{\gamma} = -\gamma \dot{\gamma}^\top \dot{\gamma}$. Consequently, the restricted Christoffel form of the second kind associated with the special orthogonal group endowed with the canonical metric takes the expression

$$\bar{\bar{\Gamma}}_R(V, V) := R(V^\top V). \tag{406}$$

The full Christoffel form descends from the polarization Formula (358), which leads to

$$\bar{\bar{\Gamma}}_R(V, W) = \tfrac{1}{2} R(V^\top W + W^\top V). \tag{407}$$

It is interesting to notice that $\bar{\bar{\Gamma}}_R(V, W) \in N_R\mathbb{SO}(n)$. ∎

Let us derive the expression of the Christoffel form of the second kind for the Stiefel manifold.

Example 39. *As we have seen, the Stiefel manifold $\text{St}(n, p)$ may be endowed both with a Euclidean metric and its canonical metric.*

In the Euclidean metric, the fundamental form reads as $\mathscr{F}(X, V) := \text{tr}(V^\top V)$; hence, the geodesic equation stemming from the differential inclusion reads as $\ddot{\gamma} = \gamma S$, where S is a symmetric function. Recalling that any geodesic must satisfy the constraint $\gamma^\top \gamma = I_p$, similarly to the calculations shown in the previous example, we find that

$$\bar{\bar{\Gamma}}_X(V, V) := X(V^\top V). \tag{408}$$

The full Christoffel form descends from the polarization Formula (358), *which leads to*

$$\bar{\bar{\Gamma}}_X(V,W) = \tfrac{1}{2}X(V^\top W + W^\top V). \tag{409}$$

In the canonical metric, the fundamental form reads $\mathscr{F}(X,V) := \mathrm{tr}\left(V^\top\left(I_p - \tfrac{1}{2}XX^\top\right)V\right)$. *Calculations of the geodesic equation show that*

$$\begin{aligned}
\frac{\partial\mathscr{F}}{\partial\dot{\gamma}} &= 2(I_p - \tfrac{1}{2}\gamma\gamma^\top)\dot{\gamma}, \\
\frac{\partial\mathscr{F}}{\partial\gamma} &= -\dot{\gamma}\dot{\gamma}^\top\gamma.
\end{aligned} \tag{410}$$

From the first derivative, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathscr{F}}{\partial\dot{\gamma}} = 2(I_p - \tfrac{1}{2}\gamma\gamma^\top)\ddot{\gamma} - (\dot{\gamma}\gamma^\top + \gamma\dot{\gamma}^\top)\dot{\gamma}. \tag{411}$$

From the differential inclusion (153), *it hence follows the equation*

$$2(I_p - \tfrac{1}{2}\gamma\gamma^\top)\ddot{\gamma} = -\dot{\gamma}\dot{\gamma}^\top\gamma + \dot{\gamma}\gamma^\top\dot{\gamma} + \gamma\dot{\gamma}^\top\dot{\gamma} - \gamma S, \tag{412}$$

where S is a symmetric matrix function to be determined. To solve for the naive derivative $\ddot{\gamma}$, *let us recall that* $(I_p - \tfrac{1}{2}\gamma\gamma^\top)^{-1} = I_p + \gamma\gamma^\top$. *Hence, the above equation may be rewritten as*

$$\ddot{\gamma} = \dot{\gamma}\gamma^\top\dot{\gamma} + \gamma(\gamma^\top\dot{\gamma})^2 - \gamma\dot{\gamma}^\top\dot{\gamma} - \gamma S, \tag{413}$$

where we have repeatedly used the fact that $\dot{\gamma}^\top\gamma = -\gamma^\top\dot{\gamma}$. *Plugging the above relationship into the condition* (405) *leads to the expression* $S = 2(\gamma^\top\dot{\gamma})^2$. *Plugging back such an expression into the relationship* (413) *leads to the geodesic equation*

$$\ddot{\gamma} + \dot{\gamma}\gamma^\top\dot{\gamma} + \gamma((\gamma^\top\dot{\gamma})^2 + \dot{\gamma}^\top\dot{\gamma}) = 0, \tag{414}$$

from which stems the restricted Christoffel form of the second kind

$$\bar{\bar{\Gamma}}_X(V,V) = VV^\top X + XV^\top(I_n - XX^\top)V. \tag{415}$$

Calculations to obtain the full Christoffel form of the second kind lead to the expression

$$\bar{\bar{\Gamma}}_X(V,W) = \tfrac{1}{2}\left\{(VW^\top + WV^\top)X + XW^\top(I_n - XX^\top)V + XV^\top(I_n - XX^\top)W\right\} \tag{416}$$

through the polarization formula. ∎

To conclude this subsection, let us prove an interesting result that concerns the parallel transport of a manifold logarithm, which will be invoked later in connection to non-linear control.

Theorem 6. *Given two points* x, $y \in \mathbb{M}$, *provided the manifold logarithms* $\log_x y$ *and* $\log_y x$ *exist, it holds that*

$$\mathbb{P}^{x\to y}(\log_x y) = -\log_y x. \tag{417}$$

Proof. Let $\gamma_x^y : [0,1] \to \mathbb{M}$ denote the geodesic arc connecting x to y and let $\tilde{\gamma}_y^x(t) := \gamma_x^y(1-t)$ denote the associated reversed geodesic. By definition of the logarithmic map, it holds that $\dot{\gamma}_x^y(0) = \log_x y$. Since $\tilde{\gamma}_y^x$ is a geodesic, it also holds that $\dot{\tilde{\gamma}}_y^x(0) = \log_y x$. In addition, by the definition of a reversed geodesic, it holds that $\dot{\tilde{\gamma}}_y^x(0) = -\dot{\gamma}_x^y(1)$. By the self-parallel transport property of any geodesic, it also follows that $\dot{\gamma}_x^y(1) = \mathbb{P}^{x\to y}(\dot{\gamma}_x^y(0))$. The assertion is hence proven. □

10.3. Coordinate-Prone Derivation of Parallel Transport*

The basic requirement of parallel transport along a curve is that it ensures the transported vector to be tangent to the curve at any given point. Clearly, this condition by itself is too weak to give rise to a unique solution to the problem of parallel transport. In Riemannian manifold calculus, it is additionally required that parallel transport along a geodesic preserves the inner product between the transported tangent vector and the tangent to the geodesic curve.

Under such a proviso, a parallel-transport rule may be constructed as follows. Let $\gamma : [0, 1] \to \mathbb{M}$ denote a geodesic curve on a p-dimensional Riemannian manifold $\mathbb{M}$ endowed with a metric tensor of components g_{ij}. Upon taking a vector $v \in T_{\gamma(0)}\mathbb{M}$, define a vector field $v^\gamma \in \Gamma(\mathbb{M})$ (which might be well-defined only along the curve γ and not necessarily outside of it) that denotes the transport of v along the curve γ. The tangency requirement is expressed by $v^\gamma(t) \in T_{\gamma(t)}\mathbb{M}$, for every $t \in [0, 1]$. The inner-product preservation property of parallel transport is expressed by requiring that the sought vector field v^γ meets the formal condition:

$$\langle v^\gamma(t), \dot{\gamma}(t) \rangle_{\gamma(t)} = \text{constant over the curve } \gamma(t). \tag{418}$$

Using local coordinates $v^\gamma = (v^1, v^2, \ldots, v^p)$ and $\gamma = (x^1, x^2, \ldots, x^p)$ and differentiating the above equation with respect to the parameter t yields the condition:

$$\frac{\mathrm{d}}{\mathrm{d}t}(g_{ij}v^i\dot{x}^j) = 0, \tag{419}$$

namely:

$$\frac{\partial g_{ij}}{\partial x^k}\dot{x}^k\dot{x}^j v^i + g_{ij}\dot{x}^j\dot{v}^i + g_{ih}\ddot{x}^h v^i = 0. \tag{420}$$

Now, let us exploit the fact that the components x^h describe a geodesic arc (namely, a self-parallel curve) on the Riemannian manifold $\mathbb{M}$. Accordingly, it holds that $\ddot{x}^h = -\Gamma^h_{kj}\dot{x}^k\dot{x}^j$. Making use of such identity in Equation (420) yields:

$$\left(\left(\frac{\partial g_{ij}}{\partial x^k} - g_{ih}\Gamma^h_{kj} \right) \dot{x}^k v^i + g_{ij}\dot{v}^i \right) \dot{x}^j = 0. \tag{421}$$

Calculations with the Christoffel symbols show that the above equation may be rewritten as:

$$g_{hj}(\Gamma^h_{ik}v^i\dot{x}^k + \dot{v}^h)\dot{x}^j = 0. \tag{422}$$

from such relation, one may derive the set of differential equations

$$\dot{v}^h = -\Gamma^h_{ik}v^i\dot{x}^k. \tag{423}$$

which describe the evolution of the components of the transport field. Solving the above set of non-linear equations under appropriate initial conditions yields a rule to perform parallel transport.

11. Manifold Retraction and Vector Transport

Since the computation of an exponential map on a given manifold may be cumbersome, the notion of retraction is sometimes invoked. A *manifold retraction* map $R : T\mathbb{M} \to \mathbb{M}$ is a function that satisfies the following requirements [62] and that is easier to compute than an exponential map:

1. Any restriction R_x is defined in some open ball $B(0, \rho_x)$ of radius $\rho_x > 0$ about $0 \in T_x\mathbb{M}$ and is continuously differentiable;
2. It holds that $R_x(v) = x$ if $v = 0$;
3. Let $v(t) \in B(0, \rho_x) \subset T_x\mathbb{M}$ denote any smooth function on the tangent space $T_x\mathbb{M}$, with $t \in [-\epsilon, \epsilon]$ and $v(0) = 0$. The curve $y(t) = R_x(v(t))$, for $t \in [-\epsilon, \epsilon]$, lies in

a neighborhood of $x = y(0)$. It holds that $\dot{y} = dR_x|_v(\dot{v})$ where $dR_x|_v : T_vT_x\mathbb{M} \to T_{R_x(v)}\mathbb{M}$ denotes a tangent map. Notice that $v(t)$ lies in $T_x\mathbb{M}$ for every t; hence, the derivative $\dot{v}$ is defined simply as

$$\dot{v}(t) := \lim_{h \to 0} \frac{v(t+h) - v(t)}{h}. \tag{424}$$

For $t = 0$, it holds that $dR_x|_0 : T_0T_x\mathbb{M} \to T_{R_x(0)}\mathbb{M}$. Let us identify $T_0T_x\mathbb{M} \cong T_x\mathbb{M}$ and let us recall that $T_{R_x(0)}\mathbb{M} = T_x\mathbb{M}$. In order for R_x to be a retraction, the map $dR_x|_0$ must equate the identity map in $T_x\mathbb{M}$.

In practice, a retraction $R_x(v)$ sends a tangent vector $v \in T_xM$ to a manifold $\mathbb{M} \ni x$ into a neighbor of the point x. Any exponential map of a Riemannian manifold is a retraction. Another class of retractions was surveyed in [63].

Manifold retraction is a computationally convenient replacement of exponential map and, according to the definition given above, it behaves as a manifold exponential up to first order. As manifold logarithm is a local inverse of manifold exponential, one might wonder if a local inverse of retraction exists. Such problems have been studied in a number of papers and a possible solution has been codified under the name of a *manifold lifting map* [64].

Let us examine a few formulas of interest about manifold retraction.

- **Hypercube**: Not surprisingly, the simplest manifold retraction on the hypercube is realized through array addition, namely $R_x(v) := x + v$.
- **Hypersphere**: Let $x \in \mathbb{S}^{p-1}$ and $v \in T_x\mathbb{S}^{p-1}$. A simple retraction map on the unit hypersphere is

$$R_x(v) := \frac{x+v}{\|x+v\|}, \tag{425}$$

  where $\|\cdot\|$ denotes a 2-norm. Let us verify that this map meets the three mentioned requirements to be a retraction:

  - *Requirements 1 and 2*: Easily verified by inspection.
  - *Requirement 3*: Let $y(t) = \frac{x+v(t)}{[(x+v(t))^\top(x+v(t))]^{1/2}}$. Deriving both sides with respect to t yields:

  $$\dot{y} = \frac{\dot{v}\|x+v\| + (x+v)\frac{1}{2}\frac{\dot{v}(x+v)+(x+v)^\top\dot{v}}{\|x+v\|}}{\|x+v\|^2} = \frac{\dot{v}\|x+v\|^2 + (x+v)\dot{v}^\top(x+v)}{\|x+v\|^3}.$$

  Setting $v = 0$ leads to

  $$\dot{y}(0) = \frac{\dot{v}(0)\|x\|^2 + x\dot{v}^\top x}{\|x\|^3}.$$

  Since $\|x\| = 1$ and $\dot{v} \in T_x\mathbb{S}^{p-1}$, it follows that $\dot{v}^\top x = 0$, and hence the result is proven.

- **Stiefel manifold**: There exist a number of retractions on the Stiefel manifold, which are briefly outlined below.

  1. **Retraction based on QR factorization**: In [64], it was shown that one of the retractions that map a tangent vector of $T_X\mathrm{St}(n,p)$ to $\mathrm{St}(n,p)$ is given by:

  $$R_X^{\mathrm{qr}}(V) := \mathrm{qf}(X+V), \tag{426}$$

  where $\mathrm{qf}(\cdot)$ denotes the Q-factor of the thin QR factorization of its $\mathbb{R}^{p \times n}$ matrix argument.

  2. **Retraction based on polar factorization**: Given a point $X \in \mathrm{St}(n,p)$ and a vector $V \in T_X\mathrm{St}(n,p)$, the polar-factorization-based retraction may be written as [64]:

  $$R_X^{\mathrm{pol}}(V) := \mathrm{pf}(X+V), \tag{427}$$

where $\mathrm{pf}(\cdot)$ denotes the polar factor of a given matrix. Such a retraction may be written in closed form. In fact, write $X + V = QS$. From the conditions $Q^\top Q = I_p$ and $S^\top = S$, it follows that:

$$(X + V)^\top (X + V) = S^\top Q^\top QS \Rightarrow X^\top X + X^\top V + V^\top X + V^\top V = S^2.$$

Since it holds that $X^\top X = I_p$ and $X^\top V + V^\top X = 0$, and the matrix $I_p + V^\top V$ is positive-definite, one obtains $S = (I_p + V^\top V)^{\frac{1}{2}}$. From the equality $X + V = QS$, the following closed-form expression for the polar factorization-based retraction is obtained:

$$R_X^{\mathrm{pol}}(V) = (X + V)(I_p + V^\top V)^{-\frac{1}{2}}. \tag{428}$$

3. **Orthographic retraction map**: The paper [65] studies *orthographic retractions* on submanifolds of Euclidean spaces. In the paper [65], it is proven that, given a pair $(X, V) \in T\mathrm{St}(n, p)$, under the proviso that V is *sufficiently close to* $0 \in T_X\mathrm{St}(n, p)$, there exists a normal array $Z \in N_X\mathrm{St}(n, p)$ such that the function

    $$R_X^{\mathrm{orth}}(V) := X + V + Z \tag{429}$$

    is a retraction on $\mathrm{St}(n, p)$. By the structure of the normal spaces, the orthographic retraction map on the Stiefel manifold $\mathrm{St}(n, p)$ reads as:

    $$R_X^{\mathrm{orth}}(V) = X + V + XS, \tag{430}$$

    provided that there exists a $p \times p$ symmetric matrix S such that $V + X(I_p + S) \in \mathrm{St}(n, p)$, namely, such that:

    $$(X + V + XS)^T(X + V + XS) = I_p. \tag{431}$$

    The above equation in the unknown matrix S may be written in plain form as:

    $$-S^2 - S(X^TV + I_p) - (V^TX + I_p)S - V^TV = 0. \tag{432}$$

    The Equation (432) represents an instance of Continuous-time Algebraic Riccati Equation (CARE). The orthographic retraction map (430) may be computed numerically as shown in [64].

*   **Real symplectic group**: Possible retractions on a real symplectic group $\mathrm{Sp}(2n)$ are

    $$R_Q(V) := Q\operatorname{Exp}(Q^{-1}V), \tag{433}$$

    whose properties were studied in the contributions [66,67], and the one based on the Cayley map, as explained in [2].

*   **Grassmann manifold**: There exist a number of retractions on the Grassmann manifold, which are briefly outlined below.

    1. **Retraction based on QR-factorization**: In [40], one of the retractions that map a tangent vector $V \in T_{[X]}\mathrm{Gr}(n, p)$ onto $\mathrm{Gr}(n, p)$ is given by:

        $$R_X^{\mathrm{qr}}(V) := \mathrm{qf}(X + V), \tag{434}$$

        where $\mathrm{qf}(\cdot)$ denotes again the Q-factor of the thin QR factorization of its $\mathbb{R}^{p \times n}$ matrix argument and $X \in \mathrm{St}(n, p)$ is a Stiefel representative of the subspace $[X]$.

    2. **Retraction based on polar factorization**: Given a subspace $[X] \in \mathrm{Gr}(n, p)$ and a tangent vector $V \in T_{[X]}\mathrm{Gr}(n, p)$, the polar-factorization-based retraction may be written as [40]:

        $$R_X^{\mathrm{pol}}(V) := (X + V)(I_p + V^\top V)^{-\frac{1}{2}}, \tag{435}$$

where $X \in \text{St}(n, p)$ is a Stiefel representative of the subspace $[X]$.

The parallel transport Equation (362) is, in general, difficult to solve, and hence the parallel transport operator might not be available in closed form for manifolds of interest. Approximations of the exact parallel transport are available, as the 'Schild's ladder' construction [68] and the *vector transport* method [62]. In order to define the notion of vector transport, a smooth manifold $\mathbb{M}$ is supposed to be embedded into a (possibly large) linear ambient space $\mathbb{A}$ of appropriate dimension. Upon endowing the linear space $\mathbb{A}$ with an inner product, it is possible to define an orthogonal projection operator $\Pi_x : \mathbb{A} \to T_x\mathbb{M}$, with $x \in \mathbb{M}$. Define the Whitney sum

$$T\mathbb{M} \oplus T\mathbb{M} := \{(x, (v, w)) \mid x \in \mathbb{M}, v \in T_x\mathbb{M}, w \in T_x\mathbb{M}\}. \tag{436}$$

Then, the vector transport operator

$$\text{vt} : (x, (v, u)) \in T\mathbb{M} \oplus T\mathbb{M} \mapsto w \in T_{\exp_x(v)}\mathbb{M} \tag{437}$$

associated to an exponential map $\exp : T\mathbb{M} \to \mathbb{M}$ is defined as:

$$\text{vt}_{x,v}(u) := \Pi_{\exp_x(v)}(u). \tag{438}$$

In practice, instead of parallel-translating the tangent vector $u \in T_x\mathbb{M}$ along a geodesic arc emanating from the point $x \in \mathbb{M}$ in the direction $v \in T_x\mathbb{M}$ as $\mathbb{P}^{0 \to 1}_{\gamma_{x,v}}(u)$, vector transport moves rigidly the vector u within $\mathbb{A}$ up to the point y and then orthogonally projects the vector u over the tangent space $T_y\mathbb{M}$ where $y = \exp_x(v)$.

With a slight abuse of exact terminology, one may refer to vector transport of a vector $u \in T_x\mathbb{M}$ to a tangent space $T_y\mathbb{M}$ by $\Pi_y(u)$ for a fixed $y \in \mathbb{M}$. In practice, vector transport is based on the following procedure:

1. Embed the manifold $\mathbb{M}$ into a metric ambient space $\mathbb{A}$.
2. Translate rigidly the vector $u \in T_x\mathbb{M}$ across the ambient space $\mathbb{A} \supset \mathbb{M}$ to a point $y \in \mathbb{M}$.
3. Project the vector u into the tangent space $T_y\mathbb{M}$ by means of a suitable orthogonal projector $\Pi_y : \mathbb{A} \to T_y\mathbb{M}$.

Vector transport associated with the above procedure is $\Pi_y(u)$, which moves the vector u from $T_x\mathbb{M} \subset A$ to $T_y\mathbb{M}$.

Depending on the manifold structure, vector transport might result to be much less expensive to compute than exact parallel transport. As a drawback, vector transport does not enjoy some of the fundamental properties of parallel transport; for instance, vector transport does *not* realize an isometry nor a conformal transformation. Isometry may be recovered through an appropriate normalization, though.

12. Control Systems on Manifolds and Numerical Implementation

This section presents an instance of error feedback control of first-order systems on manifold and their numerical implementation, with special emphasis to system synchronization. In the present research, we will regard synchronization as a goal to be achieved by non-linear proportional-type control. We shall see that the design of a synchronizing controller may be effected through the classical notion of Lyapunov function and we shall further introduce the notion of control effort to quantify the energy consumption associated with a control action.

12.1. Synchronization of First-Order Dynamical Systems via Feedback Control

Synchronization is meant primarily to make two *identical* (or twin) dynamical systems synchronize their dynamics overt time, provided that their initial states differ from one another and that one of them is able to access the state of the other. In order not to restrict how far such two initial states may lay apart, the state space $\mathbb{M}$ is assumed to be a *geodesically complete path-connected* manifold. In principle, the restriction that the two

systems to synchronize must be identical is not necessary, as long as their mathematical models insist on the same state manifold.

In a system pair, the independent dynamical system will be referred to as the *leader*, while the controlled dynamical system will be referred to as the *follower*. In this section, we shall assume that the leader and the follower differ from one another, whereby the case of identical systems will follow as a special (although most meaningful) case.

We shall cover in this paper only the case that the leader is represented by a first order, non-autonomous dynamical system on manifold, described by the tangent-bundle differential equation

$$\dot{x}^\ell(t) = f^\ell(t, x^\ell(t)), \text{ with } x^\ell(0) = x_0^\ell \in \mathbb{M}, \ t \geq 0, \tag{439}$$

where $x^\ell \in \mathbb{M}$ denotes the leader's state variable and $f^\ell : \mathbb{R} \times \mathbb{M} \to T\mathbb{M}$ denotes a possibly time-dependent state-transition operator. Likewise, the controlled follower is described by

$$\dot{x}(t) = f(t, x(t)) + u(t), \text{ with } x(0) = x_0 \in \mathbb{M}, \ t \geq 0, \tag{440}$$

where $x \in \mathbb{M}$ denotes the follower's state variable, $f : \mathbb{R} \times \mathbb{M} \to T\mathbb{M}$ denotes a state-transition operator of the follower, $u \in T\mathbb{M}$ denotes a tangent-bundle control field (in particular, $u(t) \in T_{x(t)}\mathbb{M}$) and, in general, the initial state $x_0 \in \mathbb{M}$ differs from the initial state $x_0^\ell \in \mathbb{M}$.

The control field that will drive the follower to synchronize to the leader is defined on the basis of an instantaneous (non-delayed) distance-minimization design. Assume the state manifold $\mathbb{M}$ to be endowed with a specific metric and hence a distance function $d : \mathbb{M} \times \mathbb{M} \to \mathbb{R}$ and a logarithmic map $\log : \mathbb{M} \times \mathbb{M} \to T\mathbb{M}$. Let us define the function

$$\mathscr{D}(t) := \tfrac{1}{2}d^2(x^\ell(t), x(t)) = \tfrac{1}{2}\langle \log_{x^\ell(t)}(x(t)), \log_{x^\ell(t)}(x(t)) \rangle_{x^\ell(t)}, \tag{441}$$

which is proportional to the squared Riemannian distance between the state of the leader and the state of the follower. The following result shows how to define a distance minimizing control action.

Theorem 7. *The control field*

$$u := \mathbb{P}^{x^\ell \to x}(f^\ell(t, x^\ell)) - f(t, x) + c \log_x(x^\ell), \tag{442}$$

with $c > 0$, minimizes the function $\mathscr{D}(t)$ asymptotically.

Proof. The derivative of the function $\mathscr{D}$ with respect to the time-parameter t reads

$$\begin{aligned}
\dot{\mathscr{D}} &= \langle -\log_{x^\ell}(x), \dot{x}^\ell \rangle_{x^\ell} + \langle -\log_x(x^\ell), \dot{x} \rangle_x \\
&= -\langle \log_{x^\ell}(x), f^\ell(t, x^\ell) \rangle_{x^\ell} - \langle \log_x(x^\ell), f(t, x) + u \rangle_x.
\end{aligned} \tag{443}$$

The above expression appears as a sum of two terms referring to two different tangent spaces. It is convenient to move calculations to only one tangent space. We shall take, as the tangent space of reference, the one attached to the state of the follower. Assuming that parallel transport realizes a conformal isometry (see Section 10), it holds that

$$\langle \log_{x^\ell}(x), f^\ell(t, x^\ell) \rangle_{x^\ell} = \langle \mathbb{P}^{x^\ell \to x}(\log_{x^\ell}(x)), \mathbb{P}^{x^\ell \to x}(f^\ell(t, x^\ell)) \rangle_x. \tag{444}$$

In addition, by Theorem 6, it holds that $\mathbb{P}^{x^{\ell} \to x}(\log_{x^{\ell}}(x)) = -\log_x(x^{\ell})$ because parallel transport and manifold logarithm are referred to the same geodesic line. Therefore, the expression (443) may be recast as

$$
\begin{aligned}
\dot{\mathscr{D}} &= \langle \log_x(x^{\ell}), \mathbb{P}^{x^{\ell} \to x}(f^{\ell}(t, x^{\ell})) \rangle_x - \langle \log_x(x^{\ell}), f(t, x) + u \rangle_x \\
&= \langle \log_x(x^{\ell}), \mathbb{P}^{x^{\ell} \to x}(f^{\ell}(t, x^{\ell})) - f(t, x) - u \rangle_x,
\end{aligned}
\tag{445}
$$

thanks to the linearity of the inner product. Let us set the control field u so that the sum $\mathbb{P}^{x^{\ell} \to x}(f^{\ell}(t, x^{\ell})) - f(t, x) - u$ in (445) is proportional to $-\log_x(x^{\ell})$, which leads to the expression (442). The choice that led to the control field (442) implies that the function $\mathscr{D}$ satisfies

$$
\dot{\mathscr{D}} = -c \langle \log_x(x^{\ell}), \log_x(x^{\ell}) \rangle_x = -c \| \log_x(x^{\ell}) \|_x^2 = -2c \, \mathscr{D}.
\tag{446}
$$

Since the inner product defines a positive-definite local norm, the above equation entails the inequality $\dot{\mathscr{D}} \leq 0$ at any time, which implies asymptotic synchronization. Equality may hold only when $\log_x(x^{\ell}) = 0$, which happens when the follower and the leader are perfectly synchronized. □

In the expression (442), the constant c takes the meaning of a communication strength between the leader and the follower. In addition, one might notice that the differential Equation (446) may be solved for $\mathscr{D}$ and gives $\mathscr{D}(t) \propto e^{-2ct}$; hence, synchronization happens with exponential time speed (fast at the beginning, then slower). Such a result is completely independent of the metric that the state manifold $\mathbb{M}$ was endowed with and the speed of synchronization depends only on the constant c.

Let us consider, as an example, a simpler and more familiar case.

Example 40. *As a special case, when $\mathbb{M} = \mathbb{R}^3$ endowed with the standard Euclidean metric, the parallel transport operator is simply the identity map, while $\log_x y = y - x$. Therefore, the criterion function $\mathscr{D}$ takes the shape of a quadratic error, namely $\mathscr{D} = \frac{1}{2} e^{\top} e$, where $e := x - x^{\ell}$, discussed in [69], and the corresponding control field takes the expression $u = f^{\ell}(t, x^{\ell}) - f(t, x) + c(x^{\ell} - x)$, which coincides with the control field discussed in the paper [69].*

The freedom in the choice of an inner product accounts for the generalization considered in [69] that consists of considering a Lyapunov function $\mathscr{D} := \frac{1}{2} e^{\top} P e$ with P being a symmetric, positive-definite weighting matrix. Such a weighting matrix seems to be absent from our formulation of a quadratic error (441), yet it is 'hidden' in such an expression. In fact, invoking once again an ambient space $\mathbb{A}$ that the state space is embedded within and a metric $\langle \cdot, \cdot \rangle^{\mathbb{A}}$, we may rewrite the quadratic error (441) as

$$
\mathscr{D} = \frac{1}{2} \langle \log_x(x^{\ell}), G_x(\log_x(x^{\ell})) \rangle^{\mathbb{A}},
\tag{447}
$$

and hence, by comparison, we see that the synchronization error reads $e := \log_x(x^{\ell})$ and that the role of the weighting matrix P is played by the metric kernel G. ∎

In control theory, it is customary to associate a scalar index, the *control effort*, to the control field [70]. In the context of systems on manifold, we define the control effort $\sigma : \mathbb{R} \to \mathbb{R}$ as

$$
\sigma(t) := \langle u(t), u(t) \rangle_{x(t)} = \| u(t) \|_{x(t)}^2.
\tag{448}
$$

Whenever the follower system and the leader system are twins, namely their state-transition functions are identical $f^{\ell} \equiv f$, under mild continuity conditions on such state-transition functions *the control effort vanishes asymptotically to zero*. In fact, by the continuity of the state-transition function, it follows that $\mathbb{P}^{x^{\ell} \to x}(f(t, x^{\ell})) - f(t, x)$ approaches 0 as $x(t)$ approaches $x^{\ell}(t)$. *Such a conclusion holds no longer true if the follower and leader systems' state-transition functions differ from one another* (namely, whenever $f^{\ell} \neq f$). It is worth underlining that the state-transition functions of the leader and of the follower systems, namely f^{ℓ} and f, may differ to one another even when these systems are identical since the information on the leader's state as acquired by the follower may be affected by

measurement noise, namely, it might hold that $f(t, \cdot) = f^{\ell}(t, \cdot) + \nu(t)$, where $\nu(t)$ denotes a measurement disturbance.

Another interesting observation concerns the speed of synchronization in connection to control effort saving. The constant c in the expression (442) influences the speed of synchronization, namely, the larger c, the speedier the synchronization. However, it is easy to see from the definition (448) that the constant c also affects the control effort, namely, the larger c, the more expensive the control action. Clearly, the proportional control constant c should be chosen as a tradeoff between speed and cost.

12.2. Numerical Methods to Simulate First-Order Systems

In the present subsection we shall recall from the specialized literature the notions of *forward Euler method* and of *explicit* 4th-*order Runge–Kutta method* to simulate numerically classical first-order, non-linear, non-autonomous systems. In addition, in the body of the present section we shall lay out extensions of such numerical methods to implement on a computing platform those mathematical dynamical systems insisting on curved manifolds. The main ideas to achieve such an extension may be summarized as follows:

Extension of straight stepping: Classical numerical methods tailored to $\mathbb{R}^n$ advance the solution from one step to the next by moving a system's state along straight segments. On curved state manifolds, the notion of 'straight segment' needs to be replaced by the notion of 'geodesic arc', and henceforth additive stepping needs to be replaced by exponential-map-based stepping.

Extension of linear stages: High-order methods in $\mathbb{R}^n$ select a stepping direction as a linear combination of a number of estimations of a vector field (in the Runge–Kutta methods, these estimations are called 'stages'). On curved spaces, moving directions are tangent vectors that cannot be combined together directly because they belong to different tangent spaces. Such tangent directions need to be 'aligned' together in a given tangent space by means of parallel transport and then combined together.

The forward Euler scheme (*fEul*) is perhaps the simplest numerical scheme known in the scientific literature to tackle an initial value problem. On a Euclidean space $\mathbb{R}^n$, the forward Euler method to simulate numerically a dynamical system $\dot{x}(t) = f(t, x(t))$ reads as

$$x_{k+1} = x_k + Tf(kT, x_k), \ k = 0, \ 1, \ 2, \ 3, \ \ldots, \ K, \tag{449}$$

where k denotes a step counter ranging from 0 to a given integer $K > 0$, T denotes a time-discretization interval (generally, $T \ll 1$) and x_k denotes an approximation to the true state $x(kT)$. The accuracy and numerical stability of this method turn out to be reasonable as long as the state-transition function meets certain conditions [71].

This Euler method moves forward the current state x_k to the next state x_{k+1} along a straight line directed toward $f(kT, x_k)$ of a fraction specified by T. Since curved manifolds admits no straight lines in the sense of Euclidean geometry, a plain forward Euler method is inherently unsuitable to cope with a tangent-bundle differential equation, as exemplified in the following.

Example 41. *To tackle the problem that arises about the numerical implementation of dynamical systems on manifolds, it is worth examining an explanatory example based on the low-dimensional manifold $\mathbb{SO}(2)$. As already outlined, by embedding the space $\mathbb{SO}(2)$ into the space $A := \mathbb{R}^{2 \times 2}$, any element of $\mathbb{SO}(2)$ may be regarded as a 2-by-2 real-valued matrix $X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$ whose entries must satisfy the constrains: $X_{11}^2 + X_{21}^2 = 1$, $X_{22}^2 + X_{12}^2 = 1$, $X_{11}X_{12} + X_{21}X_{22} = 0$ and $X_{11}X_{22} - X_{12}X_{21} = 1$.*

Let us consider the first-order dynamical system $\dot{X} = f(X)$ on the manifold $\mathbb{SO}(2)$, with $f : X \in \mathbb{SO}(2) \mapsto f(X) \in T_X\mathbb{SO}(2)$. Such a dynamical system may be written as a set of four differential equations of the type $\dot{X}_{ij}(t) = f_{ij}(X_{11}(t), X_{12}(t), X_{21}(t), X_{22}(t))$. In the present case, it holds that $f(X) = X\Omega$ with $\Omega = \begin{bmatrix} 0 & \omega \\ -\omega & 0 \end{bmatrix}$, where $\omega = \omega(X(t)) \in \mathbb{R}$. The forward Euler stepping technique of numerical calculus to solve the above system of differential equations

would read as $X_{k+1} = X_k + Tf(X_k)$, with $T > 0$ denoting a step size and $k \geq 0$ a step counter. Such a numerical stepping method does not take into account the constraints on the entries of matrix X, namely, it generates a trajectory $k \mapsto X_k$ in the ambient space $\mathbb{R}^{2 \times 2}$ rather than in the feasible manifold $\mathbb{SO}(2)$. Namely, starting from a point $X_k \in \mathbb{SO}(2)$, it would yield a new point $X_{k+1} \notin \mathbb{SO}(2)$. The reason of such behavior is that Euler techniques insist on flat spaces and do not cope automatically with curved manifolds.

It is instructive to investigate in detail the effect of an Euler stepping method in the solution of the differential equation $\dot{X} = X\Omega$. In such a context, the Euler stepping equation reads:

$$X_{k+1} = X_k(I_2 + T\Omega_k), \text{ with } \Omega_k = \begin{bmatrix} 0 & \omega_k \\ -\omega_k & 0 \end{bmatrix}, k \geq 0. \tag{450}$$

As the starting point, X_0 satisfies $X_0^\top X_0 = I_2$, and it holds that:

$$X_1^\top X_1 = (I_2 + T\Omega_0)^\top X_0^\top X_0 (I_2 + T\Omega_0) = I_2 - T^2 \Omega_0^2. \tag{451}$$

Notice that $\Omega_k^2 = -\omega_k^2 I_2$, and hence $X_1^\top X_1 = (1 + T^2 \omega_0^2) I_2$. The result of the first step X_1 already lost normality of its columns of an additive amount $T^2 \omega_0^2$ and changed its determinant from 1 to $\sqrt{1 + T^2 \omega_0^2}$. Since T is generally far smaller than 1, and since the deviation is proportional to T^2, such a deviation may not be apparent from the first steps, yet progressively detrimental. However, the first step keeps the orthogonality of the columns of the matrix X (such a result is peculiar of the case $\mathbb{SO}(2)$ only and does not copy to the general case $\mathbb{SO}(n)$ with $n > 2$). For the next step, it holds that:

$$\begin{aligned} X_2^\top X_2 &= (I_2 + T\Omega_1)^\top X_1^\top X_1 (I_2 + T\Omega_1) \\ &= (1 + T^2 \omega_0^2)(I_2 - T^2 \Omega_1^2) \\ &= (1 + T^2 \omega_0^2)(1 + T^2 \omega_1^2) I_2. \end{aligned} \tag{452}$$

By induction, it is readily verified that the matrix X_k keeps monotonically loosing the normality of its two columns of an identical amount. ∎

The forward Euler method (449) can be extended to a smooth manifold by replacing the notion of *straight line* with the notion of *geodesic line*, to give

$$x_{k+1} = \exp_{x_k}(Tf(kT, x_k)), k = 0, 1, 2, 3, \ldots, K. \tag{453}$$

The generalization from (449) to (453) is conceived as follows. On a curved state manifold $\mathbb{M}$, each point x_k belongs to $\mathbb{M}$ while each quantity $f(kT, x_k)$ is a tangent vector in $T_{x_k}\mathbb{M}$: these two quantities cannot be combined in an additive way, because $x_k + Tf(kT, x_k) \notin \mathbb{M}$, rather, such quantities are combined by the help of the exponential map, which describes a geodesic arc departing from the state x_k in the tangent direction $f(kT, x_k)$. The above stepping method allows one to compute the first K discrete points of the trajectory generated by the dynamical system $\dot{x} = f(t, x)$, with $x \in \mathbb{M}$, given the initial state $x_0 \in \mathbb{M}$. For a Euclidean state space $\mathbb{M} = \mathbb{R}^n$, it holds that $\exp_x(v) = x + v$; therefore, the scheme (453) is apparently a generalization of the well-known Euler scheme (449) from Euclidean to curved spaces.

The set of equations to simulate on a computing platform a leader system and a controlled follower system by the fEul method on a manifold are hence laid out as follows

$$\begin{cases} x_{k+1}^\ell = \exp_{x_k^\ell}(Tf^\ell(kT, x_k^\ell)), & \textbf{Free leader} \\ u_k = \mathbb{P}^{x_k^\ell \to x_k}(f^\ell(kT, x_k^\ell)) - f(kT, x_k) + c\log_{x_k}(x_k^\ell), & \textbf{Controller} \\ x_{k+1} = \exp_{x_k}(T(f(kT, x_k) + u_k)), & \textbf{Controlled follower} \\ k = 0, 1, 2, 3, \ldots, K. \end{cases} \tag{454}$$

For the sake of comparison, we notice that, in the case of a leader-follower pair evolving on a Euclidean state space $\mathbb{M} = \mathbb{R}^n$, the above Equation (454) would simplify into

$$
\begin{cases}
x_{k+1}^\ell = x_k^\ell + T f^\ell(kT, x_k^\ell), & \textbf{Free leader} \\
u_k = f^\ell(kT, x_k^\ell) - f(kT, x_k) + c\,(x_k^\ell - x_k), & \textbf{Controller} \\
x_{k+1} = x_k + T\,(f(kT, x_k) + u_k), & \textbf{Controlled follower} \\
k = 0,\ 1,\ 2,\ 3,\ \ldots,\ K,
\end{cases}
\tag{455}
$$

because $x_k^\ell, x_k \in \mathbb{R}^n$ and $f, f^\ell : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$.

A second class of methods that we would like to recall is the explicit, fourth-order Runge–Kutta algorithm. The family of Runge–Kutta numerical integration methods on Euclidean spaces $\mathbb{R}^n$ was conceived in order to increase the precision of lower-order methods, such as the Euler method [72]. The explicit 4th-order Runge–Kutta method (*eRK4*), is based on four partial increments (stages) that, combined together, lead to a complete step:

$$
\begin{cases}
h_{1,k} := f(kT, x_k), \\
h_{2,k} := f(kT + \frac{1}{2}T, x_k + \frac{1}{2}T h_{1,k}), \\
h_{3,k} := f(kT + \frac{1}{2}T, x_k + \frac{1}{2}T h_{2,k}), \\
h_{4,k} := f((k+1)T, x_k + T h_{3,k}), \\
x_{k+1} = x_k + T\,(\frac{h_{1,k}}{6} + \frac{h_{2,k}+h_{3,k}}{3} + \frac{h_{4,k}}{6}), \\
k = 0,\ 1,\ 2,\ 3,\ \ldots,\ K.
\end{cases}
\tag{456}
$$

Similar to the Euler method, the eRK4 method moves forward the current state x_k to the next state x_{k+1} along a straight line toward a specific direction, except that such a direction is computed in a more convoluted way.

The eRK4 method may be extended to a curved manifold by appropriately converting each of the equations in (456). Such a conversion needs to take into account that the state space $\mathbb{M}$ is now a curved manifold:

$$
\begin{cases}
h_{1,k} := f(kT, x_k), \\
x_{k+1/4} = \exp_{x_k}(\frac{1}{2}T h_{1,k}), \\
h_{2,k} := f(kT + \frac{1}{2}T, x_{k+1/4}), \quad \tilde{h}_{2,k} = \mathbb{P}^{x_{k+1/4} \to x_k}(h_{2,k}), \\
x_{k+1/2} = \exp_{x_k}(\frac{1}{2}T\tilde{h}_{2,k}), \\
h_{3,k} := f(kT + \frac{1}{2}T, x_{k+1/2}), \quad \tilde{h}_{3,k} = \mathbb{P}^{x_{k+1/2} \to x_k}(h_{3,k}), \\
x_{k+3/4} = \exp_{x_k}(T\tilde{h}_{3,k}), \\
h_{4,k} := f((k+1)T, x_{k+3/4}), \quad \tilde{h}_{4,k} = \mathbb{P}^{x_{k+3/4} \to x_k}(h_{4,k}), \\
x_{k+1} = \exp_{x_k}\left(\frac{1}{6}T h_{1,k} + \frac{1}{3}T\,(\tilde{h}_{2,k} + \tilde{h}_{3,k}) + \frac{1}{6}T\tilde{h}_{4,k}\right), \\
k = 0,\ 1,\ 2,\ 3,\ \ldots,\ K.
\end{cases}
\tag{457}
$$

(The notation $x_{k+1/4}$, $x_{k+1/2}$, ..., used to indicate intermediate steps, is standard in numerical calculus.) The reasoning that led to this numerical scheme is outlined as follows: Once a direction $h_{1,k}$ is obtained in the first stage, it is used to determine the direction $h_{2,k}$ in the second stage; the formula to compute $h_{2,k}$ in (456) prescribes to evaluate the vector field $f(\cdot, x)$ in a point $x_k + \frac{1}{2}T h_{1,k}$; this last expression, however, cannot be applied directly on a curved manifold and needs to be replaced by $\exp_{x_k}(\frac{1}{2}T h_{1,k})$, which gives the expression $h_{2,k}$ in (457); even so, the obtained result cannot be used directly, since the update rule in (456), which prescribes to advance the value x_k by $x_{k+1} = x_k + T(\frac{1}{6}h_{1,k} + \frac{1}{3}h_{2,k} + \ldots)$, needs to be translated into $x_{k+1} = \exp_{x_k}\left(\frac{1}{6}T h_{1,k} + \frac{1}{3}T h_{2,k} + \ldots\right)$ which, in turn, requires the vector $h_{2,k}$ to belong to $T_{x_k}\mathbb{M}$, while it belongs to $T_{x_{k+1/4}}\mathbb{M}$; for this reason, it is necessary to make a further modification and to parallel transport the vector $h_{2,k}$ to $T_{x_k}\mathbb{M}$, namely, to compute $\tilde{h}_{2,k} = \mathbb{P}^{x_{k+1/4} \to x_k}(h_{2,k})$; the same holds for the subsequent stages.

The complete set of equations required to implement numerically a leader and a controlled follower by the eRK4 method on a manifold are

$$
\begin{cases}
h^\ell_{1,k} := f^\ell(kT, x^\ell_k), \\
x^\ell_{k+1/4} = \exp_{x^\ell_k}\left(\tfrac{1}{2}Th^\ell_{1,k}\right), \\
h^\ell_{2,k} := f^\ell(kT + \tfrac{1}{2}T, x^\ell_{k+1/4}), \quad \tilde{h}^\ell_{2,k} = \mathbb{P}^{x^\ell_{k+1/4} \to x^\ell_k}(h^\ell_{2,k}), \\
x^\ell_{k+1/2} = \exp_{x^\ell_k}\left(\tfrac{1}{2}T\tilde{h}^\ell_{2,k}\right), \\
h^\ell_{3,k} := f^\ell(kT + \tfrac{1}{2}T, x^\ell_{k+1/2}), \quad \tilde{h}^\ell_{3,k} = \mathbb{P}^{x^\ell_{k+1/2} \to x^\ell_k}(h^\ell_{3,k}), \\
x^\ell_{k+3/4} = \exp_{x^\ell_k}\left(T\tilde{h}^\ell_{3,k}\right), \\
h^\ell_{4,k} := f^\ell((k+1)T, x^\ell_{k+3/4}), \quad \tilde{h}^\ell_{4,k} = \mathbb{P}^{x^\ell_{k+3/4} \to x^\ell_k}(h^\ell_{4,k}), \\
x^\ell_{k+1} = \exp_{x^\ell_k}\left(\tfrac{1}{6}Th^\ell_{1,k} + \tfrac{1}{3}T(\tilde{h}^\ell_{2,k} + \tilde{h}^\ell_{3,k}) + \tfrac{1}{6}T\tilde{h}^\ell_{4,k}\right) \\
u_k = \mathbb{P}^{x^\ell_k \to x_k}(f^\ell(kT, x^\ell_k)) - f(kT, x_k) + c\log_{x_k}(x^\ell_k), \\
h_{1,k} := f(kT, x_k), \\
x_{k+1/4} = \exp_{x_k}\left(\tfrac{1}{2}Th_{1,k}\right), \\
h_{2,k} := f(kT + \tfrac{1}{2}T, x_{k+1/4}), \quad \tilde{h}_{2,k} = \mathbb{P}^{x_{k+1/4} \to x_k}(h_{2,k}), \\
x_{k+1/2} = \exp_{x_k}\left(\tfrac{1}{2}T\tilde{h}_{2,k}\right), \\
h_{3,k} := f(kT + \tfrac{1}{2}T, x_{k+1/2}), \quad \tilde{h}_{3,k} = \mathbb{P}^{x_{k+1/2} \to x_k}(h_{3,k}), \\
x_{k+3/4} = \exp_{x_k}\left(T\tilde{h}_{3,k}\right), \\
h_{4,k} := f((k+1)T, x_{k+3/4}), \quad \tilde{h}_{4,k} = \mathbb{P}^{x_{k+3/4} \to x_k}(h_{4,k}), \\
x_{k+1} = \exp_{x_k}\left(\tfrac{1}{6}Th_{1,k} + \tfrac{1}{3}T(\tilde{h}_{2,k} + \tilde{h}_{3,k}) + \tfrac{1}{6}T\tilde{h}_{4,k} + Tu_k\right), \\
k = 0, 1, 2, 3, \ldots, K.
\end{cases}
\tag{458}
$$

Leader — Controller — Follower

The implementation of the above equations on a computing platform does not pose serious concerns as long as the chosen development language possesses adequate commands to deal seamlessly with arrays and array-type functions.

13. Riemannian Hessian of a Manifold-to-Scalar Function

The next step in Taylor series approximation to a manifold-to-scalar function beyond the first-order term (through gradient) is the second-order termed based on the notion of 'Hessian'.

13.1. Definition, Properties and Coordinate-Free Calculation of Riemannian Hessian

Given a differentiable function $f : \mathbb{M} \to \mathbb{R}$, its *Riemannian Hessian* at a point $x \in \mathbb{M}$ is a linear operator $H_x f : T_x\mathbb{M} \to T_x\mathbb{M}$ that appears in the quadratic term of the Taylor approximation

$$
f(\gamma_{x,v}(t)) = f(x) + \langle \mathrm{grad}_x f, v \rangle_x t + \tfrac{1}{2}\langle H_x f(v), v \rangle_x t^2 + \mathcal{O}(t^3),
\tag{459}
$$

where $\gamma_{x,v} : [-\epsilon, \epsilon] \to \mathbb{M}$ denotes, in principle, any smooth curve such that $\gamma_{x,v}(0) = x$ and $\dot{\gamma}_{x,v}(0) = v$. A major problem with such liberality in the choice of a curve is that, since clearly

$$
\langle H_x(v), v \rangle_x = \frac{\mathrm{d}^2 f(\gamma_{x,v}(t))}{\mathrm{d}t^2}\bigg|_{t=0},
\tag{460}
$$

the Hessian operator would not just depend on the pair (x, v) but also on $\ddot{\gamma}_{x,v}(0)$, which is not prescribed. A Hessian that depends on the shape of the curve γ which, after all, should be instrumental—not determinant—is hardly acceptable in practice, which is why the notion of a Hessian operator is generally defined with respect to a *geodesic* $\gamma_{x,v}$. It is worth pointing out that, according to the 'algebraic' argument recalled in Observation 1, since the

Hessian stems from a quadratic form, only its self-adjoint part plays a role. Indeed, since its very definition stems from a quadratic form, *only its self-adjoint component may be determined*.

Not surprisingly, the Riemannian Hessian on a manifold $\mathbb{M}$ embedded in an ambient space $\mathbb{A}$ is related to the ambient Hessian in $\mathbb{A}$. Still, there is a perhaps surprising outcome in its calculation, namely, it depends on the ambient gradient of f (while the standard Hessian does not). Let us recall what is meant by ambient Hessian $\partial_x^2 f$ of a function $f : \mathbb{A} \to \mathbb{R}$, that is

$$\partial_x^2 f(v) := \lim_{h \to 0} \frac{\partial_{\rho_{x,v}(h)} f - \partial_x f}{h} = \left. \frac{\mathrm{d}}{\mathrm{d}t} \partial_{\rho_{x,v}(t)} f \right|_{t=0}, \tag{461}$$

where $\rho_{x,v} : [-\epsilon, \epsilon] \to \mathbb{M}$ denotes any smooth curve such that $\rho_{x,v}(0) = x \in \mathbb{M}$ and $\dot{\rho}_{x,v}(0) = v \in T_x\mathbb{M}$—for instance, if $f : \mathbb{R}^p \to \mathbb{R}$, then $\partial_x^2 f$ takes the form of a $p \times p$ matrix of partial derivatives and $\partial_x^2 f(v)$ just denotes the matrix-to-column-array product $(\partial_x^2 f)v$.

To carry out the calculation of the Riemannian Hessian, let us recall the relationship (301) that may be written as

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma_{x,v}(t)) = \langle \mathrm{grad}_{\gamma_{x,v}(t)} f, \dot{\gamma}_{x,v}(t) \rangle_{\gamma_{x,v}(t)}. \tag{462}$$

Introducing the metric kernel and the inner product on the ambient space gives:

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma_{x,v}(t)) = \langle G_{\gamma_{x,v}(t)}(\mathrm{grad}_{\gamma_{x,v}(t)} f), \dot{\gamma}_{x,v}(t) \rangle^{\mathbb{A}}. \tag{463}$$

Recalling the explicit expression (284) of the Riemannian gradient based on an orthogonal projector Π gives

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma_{x,v}(t)) = \langle \Pi_{\gamma_{x,v}(t)}(\partial_{\gamma_{x,v}(t)} f), \dot{\gamma}_{x,v}(t) \rangle^{\mathbb{A}}. \tag{464}$$

The right-hand side in the above expression has been written in a way that makes it easier to take a second derivative with respect to the parameter t, which reads as

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\gamma_{x,v}(t)) = \langle \tfrac{\mathrm{d}}{\mathrm{d}t} \Pi_{\gamma_{x,v}(t)}(\partial_{\gamma_{x,v}(t)} f), \dot{\gamma}_{x,v}(t) \rangle^{\mathbb{A}} + \langle \Pi_{\gamma_{x,v}(t)}(\partial_{\gamma_{x,v}(t)} f), \ddot{\gamma}_{x,v}(t) \rangle^{\mathbb{A}}. \tag{465}$$

The first expression on the right-hand side may be written explicitly by introducing a linear operator $\Pi^\bullet$ that represents a derivative of the orthogonal projector, while the second term on the right-hand side may be made more explicit by recalling, from the relationships in (356), the relation between the naïve acceleration over a geodesic and the Christoffel form of second kind. Let us define the linear operator

$$\Pi_x^\bullet(a, v) := \left. \frac{\mathrm{d}}{\mathrm{d}t} \Pi_{\rho_{x,v}(t)}(a) \right|_{t=0}, \tag{466}$$

where $x \in \mathbb{M}$ denotes a point where projection is carried out, $a \in \mathbb{A}$ is an element of the ambient space whose projection is sought, $v \in T_x\mathbb{M}$ is a tangent direction along which a variation of the projected vector $\Pi_x(a)$ is sought to be estimated and $\rho_{x,v} : [-\epsilon, \epsilon] \to \mathbb{M}$ is any smooth curve such that $\rho_{x,v}(0) = x$ and $\dot{\rho}_{x,v}(0) = v$.

To clarify the above definition, let us examine a useful computation rule.

Example 42. *Let us define $g : \mathbb{M} \to \mathbb{A}$ as an ambient-valued function of a manifold-valued argument and a smooth curve $\gamma : \mathbb{R} \to \mathbb{M}$. Let us now consider the composition $\Pi_{\gamma(t)}(g(\gamma(t)))$. We aim at calculating its derivative with respect to the parameter t. It holds that*

$$\frac{\mathrm{d}}{\mathrm{d}t} \Pi_\gamma(g(\gamma)) = \Pi_\gamma^\bullet(g(\gamma), \dot{\gamma}) + \Pi_\gamma(\partial_\gamma g(\dot{\gamma})), \tag{467}$$

which is a linear function of $\dot\gamma$. ∎

On the basis of the above calculation rule, of interest in the following developments, we can prove a collateral property of the map $\Pi^\bullet$.

Example 43. *The map $\Pi^\bullet$ applied to* two *tangent vectors returns a normal element. In fact, let us recall that $\Pi_x(\Pi_x(a)) = \Pi_x(a)$ for every $x \in \mathbb{M}$ and $a \in \mathbb{A}$. On any smooth curve $\gamma : [-\epsilon, \epsilon] \to \mathbb{M}$ such that $\gamma(0) = x$ and $\dot\gamma(0) = v$ it holds that*

$$\Pi_\gamma(\Pi_\gamma(a)) = \Pi_\gamma(a). \tag{468}$$

Deriving with respect to the parameter t and applying the computation rule (467), yields

$$\Pi^\bullet_\gamma(\Pi_\gamma(a), \dot\gamma) + \Pi_\gamma(\Pi^\bullet_\gamma(a, \dot\gamma)) = \Pi^\bullet_\gamma(a, \dot\gamma). \tag{469}$$

Setting $t = 0$ yields:

$$\Pi^\bullet_x(\Pi_x(a), v) + \Pi_x(\Pi^\bullet_x(a, v)) = \Pi^\bullet_x(a, v). \tag{470}$$

Now we can set $w := \Pi_x(a) \in T_x\mathbb{M}$ and write

$$\Pi^\bullet_x(w, v) = \Pi^\bullet_x(a, v) - \Pi_x(\Pi^\bullet_x(a, v)). \tag{471}$$

The term on the right-hand side represents the normal component of $\Pi^\bullet_x(a, v)$, and hence $\Pi_x(\Pi^\bullet_x(w, v)) = 0$. ∎

On the basis of the definition (337), the second derivative (465) may be rewritten as

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\gamma_{x,v}(t)) &= \langle \Pi^\bullet_{\gamma_{x,v}(t)}(\partial_{\gamma_{x,v}(t)} f, \dot\gamma_{x,v}(t)), \dot\gamma_{x,v}(t) \rangle^\mathbb{A} \\
&+ \langle \Pi_{\gamma_{x,v}(t)}(\tfrac{\mathrm{d}}{\mathrm{d}t} \partial_{\gamma_{x,v}(t)} f), \dot\gamma_{x,v}(t) \rangle^\mathbb{A} \\
&- \langle \Pi_{\gamma_{x,v}(t)}(\partial_{\gamma_{x,v}(t)} f), \bar{\bar{\Gamma}}_{\gamma_{x,v}(t)}(\dot\gamma_{x,v}(t), \dot\gamma_{x,v}(t)) \rangle^\mathbb{A}.
\end{aligned} \tag{472}$$

Setting $t = 0$ leads to

$$\langle H_x f(v), v \rangle_x = \langle \Pi^\bullet_x(\partial_x f, v), v \rangle^\mathbb{A} + \langle \Pi_x(\partial^2_x f(v)), v \rangle^\mathbb{A} - \langle \Pi_x(\partial_x f), \bar{\bar{\Gamma}}_x(v, v) \rangle^\mathbb{A}. \tag{473}$$

As a last step, recall that, in the rightmost term, it is possible to swap one instance of v with the term $\Pi_x(\partial_x f)$; hence, we write

$$\langle G_x(H_x f(v)), v \rangle^\mathbb{A} = \langle \Pi^\bullet_x(\partial_x f, v), v \rangle^\mathbb{A} + \langle \Pi_x(\partial^2_x f(v)), v \rangle^\mathbb{A} - \langle \bar{\bar{\Gamma}}_x(\Pi_x(\partial_x f), v), v \rangle^\mathbb{A}, \tag{474}$$

which must hold for every $v \in T_x\mathbb{M}$. The last expression is equivalent to

$$G_x(H_x f(v)) - \Pi^\bullet_x(\partial_x f, v) - \Pi_x(\partial^2_x f(v)) + \bar{\bar{\Gamma}}_x(\Pi_x(\partial_x f), v) \in N_x\mathbb{M}. \tag{475}$$

An explicit expression of the Riemannian Hessian may now be obtained by applying the orthogonal projector Π_x to both sides and then the inverse metric kernel to the result, which ultimately gives

$$H_x f(v) = G^{-1}_x(\Pi_x(\partial^2_x f(v))) + G^{-1}_x(\Pi_x(\Pi^\bullet_x(\partial_x f, v))) - G^{-1}_x(\Pi_x(\bar{\bar{\Gamma}}_x(\Pi_x(\partial_x f), v))). \tag{476}$$

Notice that, unless the difference between the last two terms is zero, the Riemannian Hessian depends on the ambient gradient $\partial_x f$.

It is worth examining separately a few special cases in which the expression of the Riemannian Hessian simplifies:

- **Case that Π_x is independent of x**: This case occurs when tangent spaces coincide to one another. In this case, the projection operator may simply be denoted as Π and it holds that $\Pi^\bullet \equiv 0$; therefore,

$$H_x f(v) = G_x^{-1}(\Pi(\partial_x^2 f(v))) - G_x^{-1}(\Pi(\bar{\bar{\Gamma}}_x(\Pi(\partial_x f), v))). \tag{477}$$

- **Case that the metric kernel coincides with the identity**: This is the case most generally covered in the literature that occurs when the ambient space $\mathbb{A}$ is a Euclidean space. In this case

$$H_x f(v) = \Pi_x(\Pi_x^\bullet(\partial_x f, v)) + \Pi_x(\partial_x^2 f(v)) - \Pi_x(\bar{\Gamma}_x(\Pi_x(\partial_x f), v)). \tag{478}$$

This case was explicitly covered in [73], in which the expression of the Hessian was given in terms of *Weingarten map*. The importance of the Weingarten map in applied sciences was further highlighted in [74].

- **Case that the gradient of the function f is null**: In a point $x \in \mathbb{M}$ where $\mathrm{grad}_x f = 0$ the expression of the Hessian simplifies noticeably. In fact the last two terms in (476) vanish to zero; therefore, it holds that

$$H_x f(v) = (G_x^{-1} \circ \Pi_x)(\partial_x^2 f(v)). \tag{479}$$

This is in fact the only case in which the Hessian does not depend on the Christoffel form.

- **Case of the manifold $\mathbb{R}^p$ endowed with a Euclidean metric**: This is the reference case that we may look at for familiarity. In this case, $G_x = \mathrm{id}_x$, $\Pi_x = \mathrm{id}_x$, $\bar{\Gamma}_x = 0$; hence,

$$H_x f(v) = \partial_x^2 f(v), \tag{480}$$

which looks exactly as one expects.

To what concerns the operator $\Pi^\bullet$, we are going to assume that the following commutation property with the ambient metric holds: For every $x \in \mathbb{M}$, $a \in \mathbb{A}$, $u, v \in T_x\mathbb{M}$ it is true that

$$\langle \Pi_x^\bullet(a, v), w \rangle^{\mathbb{A}} = \langle \Pi_x^\bullet(a, w), v \rangle^{\mathbb{A}}. \tag{481}$$

Let us go into some detail about the practical computation of the operator $\Pi^\bullet$ by surveying an example related to the Stiefel manifold.

Example 44. *The Stiefel manifold embedded in its canonical ambient space endowed with the Euclidean metric has an orthogonal projector defined by the relation (286), namely [33]:*

$$\Pi_X(A) := A - \tfrac{1}{2}X(X^\top A + A^\top X). \tag{482}$$

In order to apply the definition (466), let us first take any smooth curve $\rho_{X,V}(t)$, with $X \in \mathrm{St}(n,p)$ and $V \in T_X\mathrm{St}(n,p)$, and write

$$\Pi_{\rho_{X,V}(t)}(A) := A - \tfrac{1}{2}\rho_{X,V}(t)(\rho_{X,V}^\top(t)A + A^\top \rho_{X,V}(t)). \tag{483}$$

Taking the first derivative, side by side, with respect to the parameter t gives:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\Pi_{\rho_{X,V}(t)}(A) := &-\tfrac{1}{2}\dot{\rho}_{X,V}(t)(\rho_{X,V}^\top(t)A + A^\top \rho_{X,V}(t)) \\
&-\tfrac{1}{2}\rho_{X,V}(t)(\dot{\rho}_{X,V}^\top(t)A + A^\top \dot{\rho}_{X,V}(t)).
\end{aligned} \tag{484}$$

Now, setting $t = 0$ and recalling that $\rho_{X,V}(0) = X$ and $\dot{\rho}_{X,V}(0) = V$ yields

$$\Pi_X^\bullet(A, V) = -\tfrac{1}{2}X(V^\top A + A^\top V) - \tfrac{1}{2}V(X^\top A + A^\top X). \tag{485}$$

Let us verify the commutativity property (481). *We have*

$$
\begin{aligned}
\langle \Pi_X^\bullet(A,V), W \rangle^{\mathbb{A}} &= -\tfrac{1}{2}\mathrm{tr}(W^\top X(V^\top A + A^\top V)) - \tfrac{1}{2}\mathrm{tr}(W^\top V(X^\top A + A^\top X)), \\
\langle \Pi_X^\bullet(A,W), V \rangle^{\mathbb{A}} &= -\tfrac{1}{2}\mathrm{tr}(V^\top X(W^\top A + A^\top W)) - \tfrac{1}{2}\mathrm{tr}(V^\top W(X^\top A + A^\top X)).
\end{aligned}
\tag{486}
$$

Notice that the first terms on the right-hand sides of both inner products are null, because $W^\top X$ and $V^\top X$ are skew-symmetric, while the remaining terms may be proven equal by the trace cyclic permutation property and by symmetry. ∎

The case of the special orthogonal group is relevant and instructive as well.

Example 45. *The special orthogonal group $\mathbb{SO}(n)$ embedded into the ambient space $\mathbb{A} := \mathbb{R}^{n \times n}$, endowed with the Euclidean metric, admits the orthogonal projection*

$$
\Pi_R(A) := \tfrac{1}{2}(A - RA^\top R).
\tag{487}
$$

Let us verify that Π_R maps any matrix from $\mathbb{A}$ to the tangent space $T_R\mathbb{SO}(n)$:

$$
R^\top \Pi_R(A) + \Pi_R^\top(A)R = \tfrac{1}{2}R^\top(A - RA^\top R) + \tfrac{1}{2}(A - RA^\top R)^\top R = 0,
\tag{488}
$$

by easy matrix calculations. In addition, let us verify that $A - \Pi_R(A) \in N_R\mathbb{SO}(n)$:

$$
\begin{aligned}
\langle A - \Pi_R(A), V \rangle^{\mathbb{A}} &= \mathrm{tr}\left(V^\top \left(A - \tfrac{1}{2}(A - RA^\top R) \right) \right) \\
&= \tfrac{1}{2}\mathrm{tr}\left(V^\top A + (V^\top R)A^\top R \right) \\
&= \tfrac{1}{2}\mathrm{tr}\left(V^\top A + (-R^\top V)A^\top R \right) \\
&= \tfrac{1}{2}\mathrm{tr}\left(V^\top A - A^\top (RR^\top)V) \right) \\
&= \tfrac{1}{2}\mathrm{tr}(V^\top A - A^\top V) \\
&= 0.
\end{aligned}
\tag{489}
$$

To apply the definition (466), *let us first take a smooth curve $\rho_{R,V}(t)$, with $R \in \mathbb{SO}(n)$ and $V \in T_X\mathbb{SO}(n)$, and evaluate*

$$
\Pi_{\rho_{R,V}(t)}(A) := \tfrac{1}{2}(A - \rho_{R,V}(t)A^\top \rho_{R,V}(t)).
\tag{490}
$$

Taking the first derivative, side by side, with respect to the parameter t gives:

$$
\frac{\mathrm{d}}{\mathrm{d}t}\Pi_{\rho_{R,V}(t)}(A) := -\tfrac{1}{2}\dot{\rho}_{R,V}(t)A^\top \rho_{R,V}(t) - \tfrac{1}{2}\rho_{R,V}(t)A^\top \dot{\rho}_{R,V}(t).
\tag{491}
$$

Setting $t = 0$ and recalling that $\rho_{R,V}(0) = R$ and $\dot{\rho}_{R,V}(0) = R$ yields

$$
\Pi_R^\bullet(A,V) = -\tfrac{1}{2}\left(VA^\top R + RA^\top V \right).
\tag{492}
$$

To end this example, let us verify the commutativity property (481). *Let us observe that*

$$
\begin{aligned}
\langle \Pi_R^\bullet(A,V), W \rangle^{\mathbb{A}} &= -\tfrac{1}{2}\mathrm{tr}(W^\top VA^\top R) - \tfrac{1}{2}\mathrm{tr}(W^\top RA^\top V), \\
\langle \Pi_R^\bullet(A,W), V \rangle^{\mathbb{A}} &= -\tfrac{1}{2}\mathrm{tr}(V^\top WA^\top R) - \tfrac{1}{2}\mathrm{tr}(V^\top RA^\top W).
\end{aligned}
\tag{493}
$$

The above expressions may be proven equal by some clever matrix manipulation. For example, by noticing that $W^\top = -R^\top W R^\top$ and $V = -RV^\top R$, one may show that

$$
\begin{aligned}
\operatorname{tr}(W^\top R A^\top V) &= \operatorname{tr}((-R^\top W R^\top) R A^\top (-R V^\top R)) \\
&= \operatorname{tr}(W A^\top R V^\top) \\
&= \operatorname{tr}(V^\top W A^\top R),
\end{aligned}
\tag{494}
$$

by repeated application of cyclic permutation invariance.

Let us complete the calculation of the Riemannian Hessian for the special orthogonal group. According to the Hessian formula (478), we preliminarily need to evaluate

$$
\begin{aligned}
\Pi_R(\Pi_R^\bullet(A, V)) &= \tfrac{1}{4}\Big\{(AV^\top - VA^\top)R - R(A^\top V - V^\top A)\Big\}, \\
\Pi_R(\bar{\bar{\Gamma}}_R(V, W)) &= 0,
\end{aligned}
\tag{495}
$$

therefore

$$
\begin{aligned}
H_R f(V) ={}& \tfrac{1}{2}(\partial_R^2 f(V) - R\,(\partial_R^2 f(V))^\top R) \\
&- \tfrac{1}{4}\Big\{(\partial_R f V^\top - V\partial_R^\top f)R - R(\partial_R^\top f V - V^\top \partial_R f)\Big\}.
\end{aligned}
\tag{496}
$$

It is easy to verify, by direct calculation, that $H_R(V) \in T_R\mathbb{SO}(n)$. ∎

Let us survey the computation of the Riemannian Hessian on the hypersphere.

Example 46. *Let us compute the Riemannian Hessian associated with the hypersphere $\mathbb{M} := \mathbb{S}^{p-1}$ endowed with the canonical metric $\langle u, v\rangle_x := u^\top v$, embedded in the ambient space $\mathbb{A} := \mathbb{R}^p$ endowed with the metric $\langle u, v\rangle^{\mathbb{A}} := u^\top v$. Recall that it holds $\Pi_x(a) := (I_p - xx^\top)a$ and $\bar{\Gamma}_x(u, v) = -x(u^\top v)$.*

It is just necessary to compute the derivative $\Pi_x^\bullet(a, v)$ and then to make use of the relationship (478). By definition, we have

$$
\Pi_x^\bullet(a, v) := \left.\frac{\mathrm{d}}{\mathrm{d}t}\Pi_{\rho_{x,v}(t)}(a)\right|_{t=0} = -(xv^\top + vx^\top)a.
\tag{497}
$$

Ultimately, we obtain

$$
H_x f(v) = (I_p - xx^\top)(\partial_x^2 f)v - (x^\top \partial_x f)v,
\tag{498}
$$

which represents the Riemannian Hessian utilized in [61].

It is instrumental to notice that, for $a \in \mathbb{R}^p$ and $v \in T_x\mathbb{S}^{p-1}$, the quantity $\Pi_x^\bullet(a, v)$ possesses a radial (normal) component $x(v^\top a)$ as well as a tangential component $v(x^\top a)$. Let us verify the property (481). We have

$$
\begin{aligned}
\langle \Pi_x^\bullet(a, v), w\rangle^{\mathbb{A}} &= -(w^\top x)(v^\top a) + (w^\top v)(x^\top a), \\
\langle \Pi_x^\bullet(a, w), v\rangle^{\mathbb{A}} &= -(v^\top x)(w^\top a) + (v^\top w)(x^\top a).
\end{aligned}
\tag{499}
$$

The first terms on the right-hand sides are null because $v, w \in T_x\mathbb{S}^{p-1}$, while the remaining terms are to one another equal. ∎

The space of symmetric, positive-definite matrices is interesting to what concerns Riemannian Hessian.

Example 47. *Let us recall that, for the space $\mathbb{S}^+(n)$, we have chosen $G_P(W) := P^{-1}WP^{-1}$, also it holds that $\bar{\bar{\Gamma}}_P(V, W) := -\tfrac{1}{2}VP^{-1}W - \tfrac{1}{2}WP^{-1}V$, $\Pi_P(A) = \tfrac{1}{2}(A + A^\top)$ for any $A \in \mathbb{GL}(n)$.*

Since Π_P does not explicitly depend on the point P, we may utilize formula (477) to evaluate the Hessian.   ∎

As a last example, let us compute the Riemannian Hessian of a linear function. In the familiar $\mathbb{R}^n$ case, we should obtain a null Hessian, while in a manifold setting this is easily guessed not to be the case.

Example 48. *Let a smooth manifold $\mathbb{M}$ be embedded in an ambient space $\mathbb{A}$ and let $\ell : \mathbb{M} \to \mathbb{R}$ denote a linear function such that $\partial_x \ell = a \in \mathbb{A}$ constant and $\partial_x^2 \ell = 0$. Not having assumed anything else, the Hessian of the function ℓ should be calculated by means of the expression (476), namely*

$$H_x \ell(v) = G_x^{-1} \left\{ \Pi_x \left(\partial_x^2 \ell(v) + \Pi_x^\bullet (\partial_x \ell, v) - \bar{\bar{\Gamma}}_x (\Pi_x(\partial_x \ell), v) \right) \right\}. \tag{500}$$

In this case, the Riemannian Hessian reads as

$$H_x \ell(v) = G_x^{-1} \left\{ \Pi_x \left(\Pi_x^\bullet (a, v) - \bar{\bar{\Gamma}}_x (\Pi_x(a), v) \right) \right\}, \tag{501}$$

which clearly varies from point to point on $\mathbb{M}$.   ∎

The property (481) implies that in an inner product such as $\langle H_x(v), w \rangle_x$ the tangent vector arguments may be swapped, which implies *the Riemannian Hessian operator is self-adjoint*, namely

$$\langle H_x(v), w \rangle_x = \langle v, H_x(w) \rangle_x. \tag{502}$$

This property may be proven directly by the relationship (476); in fact,

$$\begin{aligned}
\langle H_x(v), w \rangle_x &= \langle G_x^{-1}(\Pi_x(\partial_x^2 f(v))), G_x(w) \rangle^{\mathbb{A}} \\
&+ \langle G_x^{-1}(\Pi_x(\Pi_x^\bullet(\partial_x f, v))), G_x(w) \rangle^{\mathbb{A}} \\
&- \langle G_x^{-1}(\Pi_x(\bar{\bar{\Gamma}}_x(\Pi_x(\partial_x f), v))), G_x(w) \rangle^{\mathbb{A}} \\
&= \langle \Pi_x(\partial_x^2 f(v)), w \rangle^{\mathbb{A}} \\
&+ \langle \Pi_x(\Pi_x^\bullet(\partial_x f, v)), w \rangle^{\mathbb{A}} \\
&- \langle \Pi_x(\bar{\bar{\Gamma}}_x(\Pi_x(\partial_x f), v)), w \rangle^{\mathbb{A}} \\
&= \langle \partial_x^2 f(v), w \rangle^{\mathbb{A}} + \langle \Pi_x^\bullet(\partial_x f, v), w \rangle^{\mathbb{A}} - \langle \bar{\bar{\Gamma}}_x(\Pi_x(\partial_x f), v), w \rangle^{\mathbb{A}}.
\end{aligned} \tag{503}$$

The first inner product is symmetric in v and w because the ambient Hessian is self-adjoint, the second term is self-adjoint by the property (481) and the last term is symmetric in v and w because the Christoffel form commutes with the inner product.

13.2. A Newton-like Optimization Algorithm

Given a function $f : \mathbb{M} \to \mathbb{R}$, we may define a quadratic approximation at a point $x \in \mathbb{M}$ in the direction $v \in T_x\mathbb{M}$ as

$$Q_f(x, v) := f(x) + \langle \text{grad}_x f, v \rangle_x + \tfrac{1}{2} \langle H_x f(v), v \rangle_x. \tag{504}$$

As often recalled, the inner products may be evaluated in terms of ambient metrics and metric kernel. Therefore, the change in the value of the function f may be evaluated as

$$Q_f(x, v) - f(x) = \langle G_x(\text{grad}_x f), v \rangle^{\mathbb{A}} + \tfrac{1}{2} \langle G_x(H_x f(v)), v \rangle^{\mathbb{A}}. \tag{505}$$

In optimization, it is fundamental to find a direction of maximal change from a given point, which may be determined by solving the following problem in v:

$$\frac{\partial}{\partial v}(Q_f(x, v) - f(x)) = 0. \tag{506}$$

A consequence of a property shown in the previous subsection, namely that the Riemannian Hessian is self-adjoint, is that

$$\partial_v \langle H_x f(v), v \rangle_x = 2\, H_x f(v). \tag{507}$$

Recalling that $G_x \circ H_x f$ is linear, the Equation (506) hence reads

$$G_x(\mathrm{grad}_x f) + G_x(H_x f(v)) = 0. \tag{508}$$

Recalling that G_x is invertible, the above relationship simplifies to $\mathrm{grad}_x f + H_x f(v) = 0$, which leads to the optimal direction

$$v = -(H_x f)^{-1}(\mathrm{grad}_x f). \tag{509}$$

Notice that $v \in T_x\mathbb{M}$, and hence the result is consistent to what was expected. It is the case to notice that the inverse of the Hessian may be hard to express in closed form; hence, it is sometimes easier to just set up the (linear) equation $H_x(v) = -\mathrm{grad}_x f$ and solve it by any linear-algebra tool.

The optimal direction may be exploited in a Newton-like optimization algorithm as follows:

$$\begin{cases} v_k = -(H_{x_k} f)^{-1}(\mathrm{grad}_{x_k} f), \\ x_{k+1} = \exp_{x_k}(h\, v_k), \end{cases} \tag{510}$$

where $k \geq 0$ is an integer step-counter, $h > 0$ is a step-size and x_0 denotes an initial guess.

14. Conclusions

The present paper focused on manifold calculus to describe the system-theoretic properties of non-linear systems whose states belong to curved manifolds and was conceived as the first part of a longer tutorial in manifold calculus.

In particular, the present tutorial paper focuses mainly on mathematical definitions and calculation rules, expressed in the language of embedded manifold calculus, that form a knowledge base to develop further concepts and applications. A number of manifolds of interest in applications were covered and a number of examples clarified some collateral, yet interesting, aspects.

A section of the present paper focuses on the design of a manifold-type system synchronization algorithm based on feedback control and on developing numerical methods tailored to curved manifolds to implement such systems and algorithms.

Since the present contribution aimed to lay out the basic concepts in manifold calculus and Lie group theory, it only covers the basis of first-order dynamical systems and proportional type control. Second-order systems and their control require advanced notions, such as covariant derivation. Such advanced notions will be covered in a forthcoming contribution.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Abraham, R.; Marsden, J.; Ratiu, T. *Manifolds, Tensor Analysis, and Applications*; Springer: Berlin/Heidelberg, Germany, 1988.
2. Arnol'd, V.; Givental', A. Symplectic geometry. In *Dynamical Systems IV: Symplectic Geometry & Its Applications*, 2nd ed.; Arnol'd, V., Novikov, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2001; Volume 4, pp. 1–138.
3. Bloch, A. An Introduction to Aspects of Geometric Control Theory. In *Nonholonomic Mechanics and Control*; Interdisciplinary Applied Mathematics; Krishnaprasad, P., Murray, R., Eds.; Springer: New York, NY, USA, 2015; Volume 24.

4.   Bullo, F.; Lewis, A. *Geometric Control of Mechanical Systems: Modeling, Analysis, and Design for Mechanical Control Systems*; Springer: Berlin/Heidelberg, Germany, 2005.

5.   Sastry, S. Geometric Nonlinear Control. In *Nonlinear Systems—Analysis, Stability, and Control*; Sastry, S., Ed.; Springer Science + Business Media: New York, NY, USA, 1999; pp. 510–573.

6.   Hairer, E.; Lubich, C.; Wanner, G. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, 2nd ed.; Springer: Berlin, Germany; New York, NY, USA, 2006.

7.   Ermentrout, G.; Rinzel, J. Beyond a pacemaker's entrainment limit-phase walk-through. *Am. J. Physiol.* **1984**, *246*, R102–R106. [CrossRef] [PubMed]

8.   Schäfer, C.; Rosenblum, M.; Kurths, J.; Abel, H. Heartbeat synchronised with ventilation. *Nature* **1998**, *392*, 239–240. [CrossRef] [PubMed]

9.   Blasius, B.; Huppert, A.; Stone, L. Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature* **1999**, *399*, 354–359. [CrossRef] [PubMed]

10. Castrejón-Pita, A.; Read, P. Synchronization in a pair of thermally coupled rotating baroclinic annuli: Understanding atmospheric teleconnections in the laboratory. *Phys. Rev. Lett.* **2010**, *104*, 204501. [CrossRef]

11. Ermentrout, B.; Wechselberger, M. Canards, clusters, and synchronization in a weakly coupled interneuron model. *SIAM J. Appl. Dyn. Syst.* **2009**, *8*, 253–278. [CrossRef]

12. Kloeden, P. Synchronization of nonautonomous dynamical systems. *Electron. J. Differ. Equ.* **2003**, *1*, 1–10.

13. Pikovsky, A.; Rosenblum, M.; Kurths, J. *Synchronization—A Universal Concept in Nonlinear Sciences*; Cambridge University Press: Cambridge, UK, 2001.

14. Stankovski, T. *Tackling the Inverse Problem for Non-Autonomous Systems: Application to the Life Sciences*; Springer Theses; Springer: Berlin/Heidelberg, Germany, 2014.

15. Blekhman, I.; Landa, P.; Rosenblum, M. Synchronization and chaotization in interacting dynamical systems. *Appl. Mech. Rev.* **1995**, *48*, 733–752. [CrossRef]

16. Luo, A. A theory for synchronization of dynamical systems. *Commun. Nonlinear Sci. Numer. Simul.* **2009**, *14*, 1901–1951. [CrossRef]

17. Stoer, J.; Bulirsch, R. *Introduction to Numerical Analysis*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2002.

18. Golub, G.; Van Loan, C. *Matrix Computations*, 3rd ed.; Johns Hopkins: Baltimore, MD, USA, 1996.

19. Higham, N. Computing the polar decomposition—With applications. *SIAM J. Sci. Stat. Comput.* **1986**, *7*, 1160–1174. [CrossRef]

20. Fiori, S.; Prifti, S. Exact low-order polynomial expressions to compute the Kolmogoroff-Nagumo mean in the affine symplectic group of optical transference matrices. *Linear Multilinear Algebra* **2017**, *65*, 840–856. [CrossRef]

21. Olver, P. *Applications of Lie Groups to Differential Equations*, 2nd ed.; Graduate Texts in Mathematics; Springer: Berlin/Heidelberg, Germany, 2003; Volume 107.

22. Fiori, S. A fast fixed-point neural blind deconvolution algorithm. *IEEE Trans. Neural Netw.* **2004**, *15*, 455–459. [CrossRef]

23. Fiori, S. Geodesic-based and projection-based neural blind deconvolution algorithms. *Signal Process.* **2008**, *88*, 521–538. [CrossRef]

24. Pajunen, P.; Girolami, M. Implementing decisions in binary decision trees using independent component analysis. In Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, 19–22 June 2000; pp. 477–481.

25. Zhu, Y.; Mio, W.; Liu, X. Optimal dimension reduction for image retrieval with correlation metrics. In Proceedings of the International Conference on Neural Networks (IJCNN 2009), Atlanta, GA, USA, 14–19 June 2009; pp. 3565–3570.

26. Yershova, A.; LaValle, S. Deterministic sampling methods for spheres and SO(3). In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2004), New Orleans, LA, USA, 26 April–1 May 2004; Volume 4, pp. 3974–3980.

27. Trendafilov, N.; Lippert, R. The multimode Procrustes problem. *Linear Algebra Its Appl.* **2002**, *349*, 245–264. [CrossRef]

28. Vasconcelos, J.; Elkaim, G.; Silvestre, C.; Oliveira, P.; Cardeira, B. Geometric approach to strapdown magnetometer calibration in sensor frame. *IEEE Trans. Aerosp. Electron. Syst.* **2011**, *47*, 1293–1306. [CrossRef]

29. Viswanathan, S.; Sanyal, A.; Samiei, E. Integrated Guidance and Feedback Control of Underactuated Robotics System in SE(3). *J. Intell. Robot Syst.* **2018**, *89*, 251–263. [CrossRef]

30. Celledoni, E.; Fiori, S. Neural learning by geometric integration of reduced 'rigid-body' equations. *J. Comput. Appl. Math.* **2004**, *172*, 247–269. [CrossRef]

31. Yoo, J.; Choi, S. Orthogonal nonnegative matrix factorization: Multiplicative updates on Stiefel manifolds. In Proceedings of the Intelligent Data Engineering and Automated Learning (IDEAL 2008), Daejeon, Korea, 2–5 November 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 140–147.

32. Amari, S.I. Natural gradient learning for over- and under-complete bases in ICA. *Neural Comput.* **1999**, *11*, 1875–1883. [CrossRef]

33. Edelman, A.; Arias, T.; Smith, S. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* **1998**, *20*, 303–353. [CrossRef]

34. Eldén, L.; Park, H. A Procrustes problem on the Stiefel manifold. *Numer. Math.* **1999**, *82*, 599–619. [CrossRef]

35. Yger, F.; Berar, M.; Gasso, G.; Rakotomamonjy, A. Oblique principal subspace tracking on manifold. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 2429–2432.

36. Moakher, M. A differential geometry approach to the geometric mean of symmetric positive definite matrices. *SIAM J. Matrix Anal. Appl.* **2005**, *26*, 735–747. [CrossRef]

37. Smith, S. Covariance, subspace, and instrinsic Cramér-Rao bounds. *IEEE Trans. Signal Process.* **2005**, *53*, 1610–1630. [CrossRef]

38. Bonnabel, S.; Sepulchre, R. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM J. Matrix Anal. Appl.* **2009**, *31*, 1055–1070. [CrossRef]

39. Fiori, S. Extended Hamiltonian learning on Riemannian manifolds: Numerical aspects. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 7–21. [CrossRef]

40. Fiori, S.; Kaneko, T.; Tanaka, T. Tangent-bundle maps on the Grassmann manifold: Application to empirical arithmetic averaging. *IEEE Trans. Signal Process.* **2015**, *63*, 155–168. [CrossRef]

41. Stratmann, P.; Lakatos, D.; Albu-Schäffer, A. Neuromodulation and synaptic plasticity for the control of fast periodic movement: Energy efficiency in coupled compliant joints via PCA. *Front. Neurorobot.* **2016**, *10*, 2. [CrossRef]

42. Leonard, N.; Krishnaprasad, P. Control of switched electrical networks using averaging on Lie groups. In Proceedings of the 33rd IEEE Conference on Decision and Control, Lake Buena Vista, FL, USA, 14–16 December 1994; pp. 1919–1924.

43. Tapp, K. *Matrix Groups for Undergraduates*, 2nd ed.; Student Mathematical Library; American Mathematical Society: Providence, RI, USA, 2016; Volume 79.

44. Arsigny, V.; Pennec, X.; Ayache, N. Bi-Invariant Means in Lie Groups. Application to Left-Invariant Polyaffine Transformations. 2006. Available online: https://hal.inria.fr/inria-00071383 (accessed on 26 October 2021).

45. Amari, S.I. Natural gradient works efficiently in learning. *Neural Comput.* **1998**, *10*, 251–276. [CrossRef]

46. Khvedelidze, A.; Mladenov, D. Generalized Calogero-Moser-Sutherland models from geodesic motion on $GL^+(n, \mathbb{R})$ group manifold. *Phys. Lett. A* **2002**, *299*, 522–530. [CrossRef]

47. Henderson, H.V.; Searle, S.R. On deriving the inverse of a sum of matrices. *SIAM Rev.* **1981**, *23*, 53–60. [CrossRef]

48. Courant, R.; Hilbert, D. *Methods of Mathematical Physics*, 1st ed.; Interscience Publishers: New York, NY, USA, 1953; Volume I.

49. Del Buono, N.; Lopez, L. Runge-Kutta type methods based on geodesics for systems of ODEs on the Stiefel manifold. *BIT Numer. Math.* **2001**, *41*, 912–923. [CrossRef]

50. Wang, J.; Sun, H.; Fiori, S. Empirical means on pseudo-orthogonal groups. *Mathematics* **2019**, *7*, 940. [CrossRef]

51. McGraw, T.; Vemuri, B.; Yezierski, B.; Mareci, T. Von Mises-Fisher mixture model of the diffusion ODF. In Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI 2006), Arlington, VA, USA, 6–9 April 2006; pp. 65–68.

52. Fréchet, M. Les élements aléatoires de nature quelconque dans un espace distancié. *Ann. De l'Institut Henri PoincarÉ* **1948**, *10*, 215–310.

53. Karcher, H. Riemannian center of mass and mollifier smoothing. *Commun. Pure Appl. Math.* **1977**, *30*, 509–541. [CrossRef]

54. Tyagi, A.; Davis, J. A recursive filter for linear systems on Riemannian manifolds. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8. [CrossRef]

55. Simoncini, V. Computational methods for linear matrix equations. *SIAM Rev.* **2016**, *58*, 377–441. [CrossRef]

56. Oja, E. Simplified neuron model as a principal component analyzer. *J. Math. Biol.* **1982**, *15*, 267–273. [CrossRef]

57. Oja, E. Neural networks, principal components, and subspaces. *Int. J. Neural Syst.* **1989**, *1*, 61–68. [CrossRef]

58. Nedić, A.; Olshevsky, A.; Shi, W. Decentralized consensus optimization and resource allocation. In *Large-Scale and Distributed Optimization*; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 2018; Volume 2227.

59. Nedić, A.; Ozdaglar, A.; Parrilo, P. Constrained consensus and optimization in multi-agent networks. *IEEE Trans. Autom. Control* **2010**, *55*, 922–938. [CrossRef]

60. Ambrose, W.; Singer, I. A theorem on holonomy. *Trans. Am. Math. Soc.* **1953**, *75*, 428–443. [CrossRef]

61. Fiori, S. Blind deconvolution by a Newton method on the non-unitary hypersphere. *Int. J. Adapt. Control Signal Process.* **2013**, *27*, 488–518. [CrossRef]

62. Qi, C.H.; Gallivan, K.; Absil, P.A. Riemannian BFGS Algorithm with Applications. In *Recent Advances in Optimization and Its Applications in Engineering, Part 3*; Diehl, M., Diehl, M., Glineur, F., Jarlebring, E., Michiels, W., Eds.; Springer: Berlin/Heidelberg, Germany, 2010, pp. 183–192.

63. Vandereycken, B.; Vandewalle, S. A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations. *SIAM J. Matrix Anal. Appl.* **2010**, *31*, 2553–2579. [CrossRef]

64. Kaneko, T.; Fiori, S.; Tanaka, T. Empirical Arithmetic Averaging over the Compact Stiefel Manifold. *IEEE Trans. Signal Process.* **2013**, *61*, 883–894. [CrossRef]

65. Absil, P.A.; Malick, J. Projection-like retractions on matrix manifolds. *SIAM J. Optim.* **2012**, *22*, 135–158. [CrossRef]

66. Fiori, S. Averaging over the Lie group of optical systems transference matrices. *Front. Electr. Electron. Eng. China* **2011**, *6*, 137–145. [CrossRef]

67. Harris, W.; Cardoso, J. The exponential-mean-log-transference as a possible representation of the optical character of an average eye. *Ophthalmic Physiol. Opt.* **2006**, *26*, 380–383. [CrossRef]

68. Misner, C.; Thorne, K.; Wheeler, J. *Gravitation*; W.H. Freeman Publisher: New York, NY, USA, 1973.

69. Ding, K.; Han, Q.L. Master-slave synchronization of nonautonomous chaotic systems and its application to rotating pendulums. *Int. J. Bifurc. Chaos* **2012**, *22*, 1250147. [CrossRef]

70. Magdy, M.; Ng, T. Regulation and control effort in self-tuning controllers. *IEE Proc. D—Control Theory Appl.* **1986**, *133*, 289–292. [CrossRef]

71. Courant, R.; Friedrichs, K.; Lewy, H. On the partial difference equations of mathematical physics. *IBM J.* **1967**, *11*, 215–234. [CrossRef]

72. Lambert, J. *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem*; Wiley: New York, NY, USA, 1992.

73. Absil, P.; Mahony, R.; Trumpf, J. An extrinsic look at the Riemannian Hessian. In *Geometric Science of Information. GSI 2013*; Lecture Notes in Computer Science; Nielsen, F., Barbaresco, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8085. [CrossRef]

74. Cao, Y.; Li, D.; Sun, H.; Assadi, A.H.; Zhang, S. Efficient Weingarten map and curvature estimation on manifolds. *Mach. Learn.* **2021**, *110*, 1319–1344. [CrossRef]