UNIVERSIDAD
DE GRANADA

Trabajo Fin de Máster

Máster en Matemáticas

Differential geometry methods in machine learning

Métodos de geometría diferencial en aprendizaje automático

Jose Torrente Teruel

Tutores: David Martín de Diego
Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM)

Cedric M. Campos
Departamento de Matemática Aplicada, Ciencia e Ingeniería de los Materiales y Tecnología Electrónica
Universidad Rey Juan Carlos

Julio 2022

MÁSTER EN
MATEMÁTICAS

Presentación del Trabajo Fin de Máster

(UNA VEZ CUMPLIMENTADO, DEBERÁ PRESENTARSE JUNTO CON LA MEMORIA)

Máster en: Matemáticas

Curso académico: 2021-2022.

DATOS DEL ALUMNO

Apellidos: Torrente Teruel	**Nombre: José**
D.N.I.: 45900355Z	**Tlf.: 659408428**
Correo electrónico: josett@correo.ugr.es	

DATOS DEL TRABAJO

TÍTULO: **Métodos de geometría diferencial en aprendizaje automático** **Differential geometry methods in machine learning**
NOMBRE DEL TUTOR/CO-TUTOR: **David Martín de Diego** **Cédric Martínez Campos**

Declaro explícitamente que el trabajo presentado es original, entendido en el sentido de que no he utilizado fuentes sin citarlas debidamente.

En Granada, a 8 de Julio de 2022.

Firmado: Jose Torrente Teruel

Abstract

Optimization theory has played a very important role in recent developments in machine learning. In this work, we focus on the development of the discrete variational geometry formalism that allows the introduction of numerical methods as the dynamical solutions of a discrete mechanical system. This variational description of time-dependent mechanics results in the conservation of the symplectic structure in the fibers of the phase space of a mechanical system. Through the concept of retraction and the left and right actions of the tangent and cotangent spaces on a Lie group, gradient-based optimization methods are introduced as a family of accelerated integrators on Lie groups. The novelty of these lies in the generalization of *Classical Momentum* (CM) and *Nesterov Accelerated Gradient* (NAG) methods on this extended case. Finally, the performance of these is numerically simulated in two examples. One corresponding to a very expensive function to minimize (Rosenbrock function) and a second case where practical applications are found to train a *deep network* that distinguishes between human actions, based on a simplified description of the human body.

Resumen

La teoría de optimización ha jugado un papel muy importante en recientes desarrollos en aprendizaje automático. En este trabajo, nos centramos en el desarrollo del formalismo de geometría variacional discreta que permite la introducción de métodos numéricos como las soluciones dinámicas de un sistema mecánico discreto. Esta descripción variacional de una mecánica dependiente del tiempo induce la conservación de la estructura simpléctica en las fibras del espacio de fases de un sistema mecánico. A través del concepto de retracción y la acción a izquierdas y derechas de los espacios tangente y cotangente sobre un grupo de Lie, se introducen métodos basados en el gradiente de la función a optimizar como una familia de integradores y métodos numéricos de optimización acelerados en grupos de Lie. La novedad de estos radica en la generalización de los métodos clásicos *Classical Momentum* (CM) y *Nesterov Accelerated Gradient* (NAG) en este caso extendido. Finalmente, se simula numéricamente la actuación de los métodos presentados en dos ejemplos. Uno correspondiente a una función muy costosa de minimizar (función de Rosenbrock) y un segundo caso donde se encuentran aplicaciones prácticas para entrenar una *deep network* que distingue entre acciones humanas, en base a una descripción simplificada del cuerpo humano.

Acknowledgments

Tengo que agradecer a mis tutores, David y Cédric, la acogida desde el primer día como a uno a más. Me siento muy afortunado de haber trabajado con vosotros. A Paloma, sin tu apoyo esto no sería posible. A mi familia, por la comprensión y consejos en una etapa de decisiones difíciles. A mis compañeros de máster, por los buenos momentos que me habéis dado en este corto tiempo. Por último, a mis compañeros de piso, por haber formado una familia durante este año.

Contents

Introduction

In recent years, Machine Learning and data analysis have motivated some new developments in optimization theory due its use in deep network training and other areas. Important examples has to be with the interaction between humans and computers, such as computer vision, psychology, artificial intelligence, etc (Sebe et al., 2005). Here we will focus on a human classification problem, where the actions will be characterized through the information of a set of rotations that represent the relative position of the skeleton joints in a sequence of images (Huang et al., 2017). This will be discussed at the end of the work.

In general, those examples involve usually large data sets and it is crucial to use very computationally efficient methods to make it possible to implement the tools. Therefore, optimization theory plays an important role, providing not only methods and convergence results, but also fundamental knowledge about them.

A typical problem in optimization is usually the minimization of a certain function $\phi\colon Q \to \mathbb{R}$ on some space Q. It is usual to impose ϕ to be continuously differentiable and convex with Lipschitzian gradient, for linear spaces, or geodesically convex in the case Q is a Riemannian manifold. First-order methods based on its gradient are widely used to solve this problem because of their reduced computational cost. The gradient descent (GD) is usually slow (order $\mathcal{O}(1/k)$ with respect to the step k), so the first accelerated gradient methods on linear spaces were developed years ago, through the so-called *Polyak's Heavy Ball* or *Classical Momentum* (CM) (Polyak, 1964) and *Nesterov Accelerated Gradient* (NAG) (Nesterov, 1983). They are called accelerated methods in the sense they are above first-order convergence. In fact, NAG achieves the convergence lower bound (order $\mathcal{O}(1/k^2)$ with respect to the step k) for gradient-based optimization (Nemirovskii and Yudin, 1983).

After several years, researchers have managed to come closer to an explanation of the "acceleration phenomenon". Firstly, Su et al. (2016) obtained the continuous limit of NAG as a time-dependent second order differential equation. Moreover, a variational perspective for these methods was introduced that year by Wibisono et al. (2016), generating methods from a time-dependent Lagrangian system with order of convergence (with respect to the step k) analogue to the continuous-time rate. It is known that the variational integrators produce automatically symplectic integrators (Marsden and West, 2001), preserving some of the geometric structure of the initial continuous system and then exhibiting long-time energy stability. Induced by this perspective, there have been other works introducing symplectic accelerated integrators in optimization, as

Betancourt et al. (2018) (an introduction to symplectic integration can be seen in Sanz-Serna and Calvo (1994); Hairer et al. (2006); Blanes and Casas (2016)).

The study of accelerated methods in non-linear spaces becomes even more important, due to the increased complexity of the problems. Particularly, optimization problems on Lie groups or manifolds arise within many topics in Machine Learning, applied mathematics and even engineering (Lu and Li, 2020). Structure-preserving methods will be more efficient than those embedding the space Q in a higher-dimensional Euclidean one by means of constraints or local coordinates.

Some new developments use an extended formalism to compute accelerated methods on Lie groups, obtaining an adaptive step-size (Lee et al., 2021). However, this can lead to convergence and computational problems, for example, if the discrete equations produce the step-size to tend to zero. This is the reason why we choose another different approach.

In this work, we follow the results by Campos et al. (2021). For a convex $\mathcal{C}^1$ function $\phi\colon D \subset \mathbb{R}^n \to \mathbb{R}$, where $D \subset \mathbb{R}^n$ is a convex open region, they introduce a discretization of a certain Lagrangian system whose discrete equations are

$$\Delta x_k = \mu_k \Delta \left[x_{k-1} - \varepsilon \eta_{k-1} \nabla \phi(x_{k-1}) \right] - \eta_k \nabla \phi(x_k)\,, \tag{1}$$

where Δ is the forward difference operator, μ_k and η_k are suitable coefficients (the method's strategy), and ε is a boolean coefficient: $\varepsilon = 0$ to obtain CM and $\varepsilon = 1$ for NAG. The terms accompanying ε are associated to a force, hence NAG is in fact CM with forces. Taking $\mu_k = 0$ we obtain gradient descent. This formalism provides a setting that can be easily generalized to Lie groups.

The objective of this work is to introduce the geometric formalism that allows the construction of CM-type accelerated integrators and, if exists, of NAG method on Lie groups. In Chapter 1, we introduce the necessary preliminaries on general manifolds to understand the geometry behind the discretization of a continuous mechanical system, both from Lagrangian and Hamiltonian viewpoint. We will take into account the possible time-dependence and the modification of the discrete equations induced by forces on a particular system. The end of this chapter includes some basis on the so-called retraction maps, a crucial tool to obtain simple discretizations of a continuous Lagrangian. Chapter 2 is devoted to the development of the proposed accelerated methods on Lie groups, using their inner structure to trivialize the tangent and cotangent bundles. Special emphasis is given to the rotations group SO(3). At the end, we will obtain a retraction-dependent family of CM and NAG methods on Lie groups. In the bibliography, we have not found an analog to these methods (and in a few weeks we will publish a preprint of this development). Finally, in Chapter 3, we will test our integrators in two different cases. The first has to be with the so-called Rosenbrock function (Rosenbrock, 1960), and makes possible a good comparison of the methods, and the second one is a practical example of their improved performance to train the deep network introduced by Huang et al. (2017).

Chapter 1

Preliminaries

In this section, we give the necessary preliminaries to understand and develop the theory of variational integration on Lie groups. To do that, we introduce here the classical theory of Lagrangian and Hamiltonian mechanics both continuous and discrete (Abraham and Marsden, 1978; de León and Rodrigues, 1989; Marsden and West, 2001).

1.1 Smooth manifolds and notation

Let Q be a smooth manifold. The **tangent bundle** TQ is equipped with a natural structure of vector bundle (Abraham and Marsden, 1978, Section 1.5). Denote by $\tau_Q : TQ \to Q$ the canonical projection onto Q, *i.e.*, $\tau_Q(v_q) = q$, $\forall v_q \in T_qQ$. Coordinates (q^i) on Q induce natural coordinates $(q^i, v^i = \frac{\partial}{\partial q^i})$ on TQ in such a way $\tau_Q(q^i, v^i) = (q^i)$. Its dual vector bundle is precisely the **cotangent bundle** T^*Q with projection $\pi_Q : T^*Q \to Q$. Coordinates (q^i) on Q induce natural coordinates $(q^i, p_i = \mathrm{d}q_i)$ on T^*Q, where $\mathrm{d}\colon \Omega(Q) \to \Omega(Q)$ is the well-known exterior derivative defined on the exterior algebra $\Omega(Q)$ consisting of the differential forms (Boothby, 2003). The action of one-forms $\alpha_q \in T_q^*Q$ on vectors $v_q \in T_qQ$ will be sometimes denoted by $\langle \alpha_q, v_q \rangle \in \mathbb{R}$.

The exterior product on $\Omega(Q)$ will be denoted by $\wedge\colon \Omega(Q) \times \Omega(Q) \to \Omega(Q)$, with $\wedge(\alpha, \beta) = \alpha \wedge \beta$. For any vector field $X \in \mathfrak{X}(Q)$, the interior product will be $\imath_X\colon \Omega(Q) \to \Omega(Q)$ and the Lie derivative $L_X\colon \Omega(Q) \to \Omega(Q)$. The relation between them is called *Cartan magic formula*, after the author, and states (Cartan, 1945)

$$L_X = \imath_X \,\mathrm{d} + \mathrm{d}\, \imath_X \,. \tag{1.1}$$

As usual, a **Riemannian metric** g on a smooth manifold Q is a smooth positive-definite $g \in \mathcal{T}_{0,2}(Q)$. For a local chart (q^i) on Q, the metric components are defined as

$$g_{ij}(q) := g_q\left(\left.\frac{\partial}{\partial q^i}\right|_q, \left.\frac{\partial}{\partial q^j}\right|_q\right).$$

Einstein notation will be used along this work. For example, if $v_q, w_q \in T_qQ$, their product has a local expression

$$g_q(v_q, w_q) = g_{ij}(q) v^i w^j \,,$$

for $v_q = v^i \frac{\partial}{\partial q^i}\big|_q$ and $w_q = w^i \frac{\partial}{\partial q^i}\big|_q$.

Given an smooth map $F : Q_1 \to Q_2$ between two manifolds Q_1 and Q_2, the **tangent map** $TF \equiv F_* : TQ_1 \to TQ_2$ is defined by

$$TF(v_q) = T_q F(v_q) := \mathrm{d}F_q(v_q) \,.$$

Based on this tangent lift, the **canonical lift** of a curve $\sigma : I \to Q$ is a curve on the tangent bundle TQ, $\dot\sigma \equiv \frac{d\sigma}{dt} : I \to TQ$, such that

$$\dot\sigma(t) \in T_{\sigma(t)} Q \,.$$

In coordinates, if $\sigma(t) = (q^i(t))$ then $\dot\sigma(t) = (q^i(t), \dot q^i(t))$.

Particularly, the tangent map of the canonical projection $\tau_Q : TQ \to Q$, gives a mapping $T\tau_Q : TTQ \to TQ$. Then, if (q^i) are local coordinates on Q and they induce local coordinates $(q^i, v^i, \Gamma^i_q, \Gamma^i_v)$ on TTQ, we find

$$T\tau_Q(q^i, v^i, \Gamma^i_q, \Gamma^i_v) = (q^i, \Gamma^i_q). \tag{1.2}$$

On the other hand, note the canonical projection in this manifold $\tau_{TQ} : TTQ \to TQ$ is locally given by

$$\tau_{TQ}(q^i, v^i, \Gamma^i_q, \Gamma^i_v) = (q^i, v^i). \tag{1.3}$$

The **Liouville vector field** Δ on TQ (Libermann and Marle, 1987, Chapter 2) is defined by

$$\begin{aligned} \Delta\colon TQ &\longrightarrow TTQ \\ u \in T_q Q \mapsto \Delta(u) &:= \frac{d}{dt}\bigg|_{t=0} (u + tu) = (u)^V_u \,, \end{aligned}$$

and $S : TTQ \to TTQ$ the **vertical endomorphism** is

$$S(\Gamma_u) = ((T_u \tau_Q)(\Gamma_u))^V_u \,,$$

where $(\cdot)^V_u : T_q Q \to T_u(T_q Q)$ denotes the **vertical lift** at $u \in T_q Q$. Note that the **horizontal part** of a vector $z \in T_u TQ$ is $T\tau_Q(z)$ (Abraham and Marsden, 1978, Section 3.7). These maps have an immediate interpretation when they are written in coordinates since

$$\Delta = v^i \frac{\partial}{\partial v^i} \quad \text{and} \quad S\Big(\Gamma^i_q \frac{\partial}{\partial q^i} + \Gamma^i_v \frac{\partial}{\partial v^i}\Big) = \Gamma^i_q \frac{\partial}{\partial v^i} \,.$$

An important class of vector fields on the tangent bundle TQ is the so-called **second order differential equation vector fields** (SODEs). These are defined to be sections of both $T\tau_Q$ and τ_{TQ}, $T\tau_Q(\Gamma) = \tau_{TQ}(\Gamma)$, *i.e.*, if Γ is a SODE on Q and (q^i, v^i) are local coordinates on TQ, the local expression takes the form

$$\Gamma(q, v) = v^i \frac{\partial}{\partial q^i} + \Gamma^i_v(q, v) \frac{\partial}{\partial v^i}, \tag{1.4}$$

where $\Gamma^i_v \in C^\infty(TQ)$. Equivalently, Γ is a SODE if and only if $S(\Gamma) = \Delta$.

1.2 Lagrangian and Hamiltonian Mechanics

1.2.1 Lagrangian description

The information of a physical system is usually encoded in the so-called Lagrangian, a function defined on the tangent bundle of the configuration space. In this section we introduce the Lagrangian description of a mechanical system. The space where the system evolves is assumed as a smooth manifold Q, called the **configuration manifold** or **configuration space**. Its tangent bundle TQ is called the **velocity phase space** (Abraham and Marsden, 1978; de León and Rodrigues, 1989). A time-dependent **Lagrangian** is a $\mathcal{C}^2$ function $L\colon \mathbb{R}\times TQ\to\mathbb{R}$ and the pair (Q,L) constitutes a free **Lagrangian system**.

In general, we assume the Lagrangian to depend explicitly on time and this will be relevant in the next chapters. It will allow to derive accelerated methods in particular examples. Nevertheless, time-independent systems, called **autonomous systems**, are related to the invariance in time of the laws of nature and suffice to solve many problems. In the case of simple mechanical systems, the Lagrangian can be decomposed into two main terms $L = T - V$, where T and V are the **kinetic** and **potential energy**, respectively. The kinetic part is defined to be $T(v_q) := \frac{1}{2}g_q(v_q, v_q)$ and the potential V usually depends only on q.

Now, let consider the set of $\mathcal{C}^2$ curves that connect two fixed points $q_a, q_b \in Q$,

$$\mathcal{C}^2(q_a,q_b,[t_a,t_b]) = \{c\colon [t_a,t_b]\to Q \,:\, c\in C^2([t_a,t_b]), c(t_a)=q_a, c(t_b)=q_b\}.$$

This has a structure of infinite-dimensional smooth manifold (see for example Lang (1995)) and its tangent space is given by (Marsden and Ratiu, 1999, Proposition 8.1.2)

$$\begin{aligned} &T_c\mathcal{C}^2(q_a,q_b,[t_a,t_b]) = \\ &\qquad = \{\delta c\colon [t_a,t_b]\to TQ \,:\, \delta c\in C^1([t_a,t_b]), \tau_Q\circ\delta c = c \text{ and } \delta c(t_a)=\delta c(t_b)=0\}. \end{aligned}$$

The **action functional** is defined as

$$\begin{aligned} \mathbb{S}\colon C^2(q_a,q_b,[t_a,t_b]) &\longrightarrow \mathbb{R} \\ c &\longmapsto \int_{t_a}^{t_b} L(t,c(t),\dot c(t))dt \end{aligned}$$

The next definition relates the equations of motion of a mechanical system to the critical points of the action.

Definition 1.1 (Hamilton principle)**.** *The **physical trajectory** of a Lagrangian system (Q,L) is a curve $c\in\mathcal{C}^2(q_a,q_b,[t_a,t_b])$ such that c is a critical point of the functional $\mathbb{S}$, i.e., $\mathrm{d}\mathbb{S}_c(\delta c)\equiv\delta\mathbb{S}(c)=0$, $\forall\delta c\in T_c\mathcal{C}^2(q_a,q_b,[t_a,t_b])$.*

For simplicity, it is usual to denote $\delta\mathbb{S}=0$, implicitly taken into account that a certain curve $c\in\mathcal{C}^2(q_a,q_b,[t_a,t_b])$ is realizing the expression. This is also called the **principle of stationary action**. Immediately, the curve satisfying Hamilton principle is characterized from a second-order differential equation.

Proposition 1.2 (Euler-Lagrange equations)**.** *Every physical trajectory of a Lagrangian system* (Q, L) *is characterized by the following second order differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}t} D_3L(t, c(t), \dot{c}(t)) - D_2L(t, c(t), \dot{c}(t)) = 0. \tag{1.5}$$

These are the well-known ***Euler-Lagrange equations****.*

Proof. Consider $\mathbb{S}(c) = \int_{t_a}^{t_b} L(t, c(t), \dot{c}(t))dt$ for $c \in \mathcal{C}^2(q_a, q_b, [t_a, t_b])$. Then, for any $\delta c \in T_c\mathcal{C}^2(q_a, q_b, [t_a, t_b])$, by means of the chain rule, the variation of this functional gives

$$\begin{aligned}\delta\mathbb{S}(c) &= \int_{t_a}^{t_b} [\langle D_2L(t, c(t), \dot{c}(t)), \delta c(t)\rangle + \langle D_3L(t, c(t), \dot{c}(t)), \delta\dot{c}(t)\rangle]\,\mathrm{d}t \\ &= \int_{t_a}^{t_b} \langle D_2L(t, c(t), \dot{c}(t)) - \tfrac{\mathrm{d}}{\mathrm{d}t} D_3L(t, c(t), \dot{c}(t)), \delta c(t)\rangle \mathrm{d}t,\end{aligned}$$

where we have used that $\delta\dot{c} = \frac{d}{dt}\delta c$, integration by parts in the last term and the fact that the boundary term $\langle D_3L(t, c(t), \dot{c}(t)), \delta c(t)\rangle|_{t_a}^{t_b}$ vanishes due to $\delta c(t_a) = \delta c(t_b) = 0$. Then, a critical point of $\mathbb{S}$, *i.e.*, the physical trajectory of the system, is given by $\delta\mathbb{S}[c](\delta c) = 0, \forall\delta c \in T_c\mathcal{C}^2(q_a, q_b, [t_a, t_b])$. Since $T_c\mathcal{C}^2(q_a, q_b, [t_a, t_b])$ contains $\mathcal{C}^1_0([t_a, t_b]) = \{\delta c\colon [t_a, t_b] \to TQ \,:\, \delta c \in C^1([t_a, t_b]), supp(\delta c) \text{ is compact}\}$ as a dense subspace, using the fundamental lemma of calculus of variations (Gelfand and Fomin, 1963; Giaquinta and Hildebrandt, 2004) we conclude the proposed Euler-Lagrange equations (1.5).

∎

Note that in local coordinates (q^i, v^i) on TQ it turns out

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v^i}\right) - \frac{\partial L}{\partial q^i} = 0\,, \tag{1.6}$$

the usual expression of the Euler-Lagrange equations.

1.2.2 Intrinsic version of the Euler-Lagrange equations

For the time being and until otherwise specified, we assume the Lagrangian L is autonomous, that is $L\colon TQ \to \mathbb{R}$. We say that L is **regular** if the matrix

$$\mathrm{Hess}(L) := \left(\frac{\partial^2 L}{\partial v^i \partial v^j}\right)$$

is non-singular for every coordinate chart.

Define the **Lagrangian energy** $E_L = \Delta(L) - L$. In this case the Euler-Lagrange equations may be written as a system of second-order differential equations obtained by computing the integral curves of the unique vector field X_L satisfying

$$\imath_{X_L}\omega_L = \mathrm{d}E_L\,, \tag{1.7}$$

where $\omega_L = -\mathrm{d}(S^*\mathrm{d}L)$ is the **Poincaré-Cartan two-form**. The one-form $\theta_L = S^*\mathrm{d}L$ is the **Poincaré-Cartan one-form**. The vector field X_L is a SODE vector field on TQ, that is, it satisfies

$$SX_L = \Delta\,,$$

and it is called the **Lagrangian vector field**. Regularity of L is equivalent to ω_L being symplectic and therefore to uniqueness of solution for equation (1.7). In fact, their local expressions are

$$E_L = v^i \frac{\partial L}{\partial v^i} - L\,, \qquad \theta_L = \frac{\partial L}{\partial v^i}\mathrm{d}q^i\,,$$

and

$$\omega_L = \frac{\partial^2 L}{\partial v^i \partial q^j}\mathrm{d}q^i \wedge \mathrm{d}q^j + \frac{\partial^2 L}{\partial v^i \partial v^j}\mathrm{d}q^i \wedge \mathrm{d}v^j\,,$$

which implies that ω_L is symplectic if and only if $\mathrm{Hess}(L)$ is non-singular (for more details see Abraham and Marsden (1978); de León and Rodrigues (1989)).

1.2.3 Forced Lagrangian mechanics

Consider (Q, L) a free Lagrangian system. An **external force** is expressed as a fibre-preserving mapping $F\colon TQ \to T^*Q$, *i.e.*, $\pi_Q \circ F = \tau_Q$.

$$\begin{tikzcd} TQ \arrow[rr, "F"] \arrow[dr, "\tau_Q"'] & & T^*Q \arrow[dl, "\pi_Q"] \\ & Q & \end{tikzcd}$$

In coordinates $F(q^i, v^i) = (q^i, F_i(q, v^i))$.

Now, given a force we can construct a semi-basic 1-form (Libermann and Marle, 1987, Chapter 2) $\mu_F \in \Lambda^1(TQ)$ as $\mu_F := (T\tau_Q)^* F$, *i.e.*, $\langle \mu_F(u_q), W_{u_q}\rangle = \langle F(u_q), T\tau_Q(W_{u_q})\rangle$, for all $W_{u_q} \in T_{u_q}TQ$. In coordinates

$$\mu_F = F_i(q^i, v^i)\,\mathrm{d}q^i\ .$$

Definition 1.3 (Lagrange-d'Alembert principle)**.** *Given a Lagrangian system (Q,L) and a force $F\colon TQ \longrightarrow T^*Q$, the **physical trajectory** of the forced system is a curve $c \in \mathcal{C}^2(q_a, q_b, [t_a, t_b])$ such that c is a critical point of the so-called total virtual work, that is,*

$$\delta \int_{t_a}^{t_b} L(c, \dot{c})\mathrm{d}t + \int_{t_a}^{t_b} \langle F(c, \dot{c}), \delta c\rangle \mathrm{d}t = 0\,, \tag{1.8}$$

for all $\delta c \in T_c\mathcal{C}^2(q_a, q_b, [t_a, t_b])$. See Sato Martín de Almagro (2019), Section 4.1, and references therein.

Using integration by parts, we get the **forced Euler-Lagrange equations**. Locally,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial v^i}\right) - \frac{\partial L}{\partial q^i} = F_i\ . \tag{1.9}$$

The global picture is induced by the vertical vector field Z_F^V given by $\imath_{Z_F^V}\omega_L = -\mu_F$. In coordinates

$$Z_F^V = W^{ij}F_j\frac{\partial}{\partial v^i}\,,$$

where W^{ij} denote the components of the inverse of the Hessian matrix $W = Hess(L)$. The dynamics of the forced Lagrangian system is determined by the vector field $X_L + Z_F^V$:

$$\imath_{X_L+Z_F^V}\omega_L = \mathrm{d}E_L - \mu_F\;.$$

Indeed, the integral curves of $X_L + Z_F^V$ are locally the solutions of the forced Euler-Lagrange equations (1.9).

1.2.4 Legendre transformation

There is an alternative description to the Lagrangian mechanics in terms of the Hamiltonian function H, which is defined on the cotangent bundle T^*Q. Given a Lagrangian L, we defined above the energy function E_L. The tool that enables us to define the objects on the Hamiltonian side is the so-called **Legendre transformation** or **fibre derivative** $\mathcal{F}L\colon TQ \to T^*Q$ defined by

$$\langle \mathcal{F}L(v_q), w_q\rangle = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} L(v_q + sw_q),\ \forall v_q, w_q \in T_qQ.$$

Remember that now we are considering an autonomous system. Taking local coordinates (q^i) in Q, it is easy to obtain

$$\mathcal{F}L(q,v) = \left(q^i, \frac{\partial L}{\partial v^i}(q,v)\right).$$

The Legendre transformation can be defined in the general context of mapping between two vector bundles over a manifold, but we do not deal with it because it is not important for our task here (Abraham and Marsden, 1978, Section 3.5).

As a matter of fact, we can characterize the regularity of $L\colon TQ \to \mathbb{R}$ by $\mathcal{F}L$ being a local diffeomorphism. In local coordinates, it is equivalent to the regularity of the Hessian matrix $\mathrm{Hess}(L)$. For instance, if we have a simple mechanical system, in which $L(q,v) = T(q,v) - V(q)$ with a kinetic energy $T(q,v) = \frac{1}{2}g_q(v,v)$, then the Hessian matrix above coincide with the one of the metric tensor g and then the Legendre transformation is always regular.

1.2.5 Hamiltonian description

If the Lagrangian is regular then we can locally express $\dot{q}^i = \dot{q}^i(q,p)$ and we can relate the coordinates on TQ and T^*Q as follows,

$$\mathcal{F}L(q^i, v^i) = (q^i, p_i)\,,$$

where $(q^i, p_i := \frac{\partial L}{\partial v^i})$ are now local coordinates on T^*Q. Then we can locally express $v^i = v^i(q,p)$ and introduce a function

$$H(q^i, p_i) = p_j v^j(q,p) - L(q^i, v^i(q,p))\;.$$

Straightforward computations from Euler-Lagrange equations result in what will be called *local Hamilton equations*,

$$
\begin{aligned}
\frac{\mathrm{d}q^i}{\mathrm{d}t} &= \frac{\partial H}{\partial p_i} \\
\frac{\mathrm{d}p_i}{\mathrm{d}t} &= -\frac{\partial H}{\partial q^i}
\end{aligned}
\tag{1.10}
$$

with $i = 1, \dots, n$.

In T^*Q the qualitative properties of a mechanical system emerge more naturally than in the Lagrangian description. The Lagrangian L is called **hyperregular** if $\mathcal{F}L$ is a (global) diffeomorphism. In that case we can define a function $H := E_L \circ (\mathcal{F}L)^{-1} \colon T^*Q \longrightarrow \mathbb{R}$ called a **Hamiltonian**. The pair (Q, H) is called a **Hamiltonian system**.

$$
\begin{tikzcd}
TQ \arrow[rr, "\mathcal{F}L"] \arrow[dr, "E_L"'] & & T^*Q \arrow[dl, "H"] \\
& \mathbb{R} &
\end{tikzcd}
$$

Remark the above local expression coincide with this definition.

Define the **Liouville 1-form** $\theta_Q \in \Lambda^1(T^*Q)$ by

$$\theta_Q(\mu_q)(X_{\mu_q}) = \langle \mu_q, T_{\mu_q}\pi_Q(X_{\mu_q}) \rangle$$

where $\mu_q \in T_q^*Q$ and $X_{\mu_q} \in T_{\mu_q}T^*Q$. In coordinates,

$$\theta_Q = p_i \, \mathrm{d}q^i \, .$$

The **canonical symplectic 2-form** ω_Q is now $\omega_Q = -d\theta_Q$ and, in coordinates,

$$\omega_Q = \mathrm{d}q^i \wedge \mathrm{d}p_i \, .$$

This just define the canonical structure of symplectic manifold in T^*Q. The **Hamiltonian vector field** $X_H \in \mathfrak{X}(T^*Q)$ is the unique one satisfying Hamilton equations

$$\imath_{X_H}\omega_Q = \mathrm{d}H \, . \tag{1.11}$$

The integral curves of X_H are the solutions of the local Hamiltonian equations (1.10). In fact, it can be written locally as

$$X_H = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i} \, .$$

Using equation (1.11) and Cartan's magic formula (1.1), it is easily derived $L_{X_H}\omega_Q = 0$. Therefore, the next result holds.

Theorem 1.4. *The flow $\Phi \colon D \subseteq \mathbb{R} \times T^*Q \to T^*Q$ of the Hamiltonian vector field X_H is a symplectomorphism for every time t, i.e.,*

$$\Phi_t^*\omega_Q = \omega_Q \, .$$

The Hamiltonian is also preserved since

$$0 = \omega_Q(X_H, X_H) = X_H(H)\,,$$

and then

$$H \circ \Phi_t = H\,.$$

Given a Hamiltonian $H\colon T^*Q \to \mathbb{R}$, as in the case of the Legendre transformation, we may construct the fibre derivative $\mathcal{F}H\colon T^*Q \to TQ$ as

$$\langle \mathcal{F}H(\mu_q), \beta_q\rangle = \tfrac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} H(\mu_q + t\beta_q).$$

In coordinates, $\mathcal{F}H(q^i, p_i) = (q^i, \frac{\partial H}{\partial p_i}(q,p))$. We say that the Hamiltonian is **regular** if $\mathcal{F}H$ is a local diffeomorphism, which in local coordinates is equivalent to the regularity of the Hessian matrix

$$M^{ij} = \frac{\partial^2 H}{\partial p_i \partial p_j}\,.$$

Proposition 1.5. *Let (Q, L) be a hyperregular Lagrangian system. If $H := E_L \circ (\mathcal{F}L)^{-1}$ is the associated Hamiltonian, then $\theta_L = (\mathcal{F}L)^*\,\theta_Q$ and $\omega_L = (\mathcal{F}L)^*\,\omega_Q$. Moreover, $\mathcal{F}H = \mathcal{F}L^{-1}$ and $X_L = (\mathcal{F}H)_* X_H$.*

Proof. Given $w_u \in T_u TQ$ we have

$$\begin{aligned}(\mathcal{F}L)^*\theta_Q(w_u) &= \theta_Q\big(T(\mathcal{F}L)w_u\big)\\ &= \langle \mathcal{F}L(u), T\pi_Q\big(T(\mathcal{F}L)w_u\big)\rangle\,.\end{aligned}$$

Then, from $\pi_Q \circ \mathcal{F}L = \tau_Q$, we get

$$\begin{aligned}(\mathcal{F}L)^*\theta_Q(w_u) &= \langle \mathcal{F}L(u), T\big(\pi_Q \circ \mathcal{F}L\big)w_u\big)\rangle\\ &= \langle \mathcal{F}L(u), T\tau_Q(w_u)\rangle\\ &= \theta_L(w)\,.\end{aligned}$$

Due to the fact that the pull-back commutes with the exterior derivative, we get $\omega_L = (\mathcal{F}L)^*\omega_Q$.

On the other hand, we know that locally we have $p_i = \frac{\partial L}{\partial v^i}$ and then $H \circ \mathcal{F}L(q^i, v^i) = p_j v^j - L(q^i, v^i)$. From the chain rule, we get

$$\frac{\partial H}{\partial p_i} = v^i + p_j \frac{\partial v^j}{\partial p_i} - \frac{\partial v^k}{\partial p_i}\frac{\partial L}{\partial v^k} = v^i\,.$$

Hence $\mathcal{F}H \circ \mathcal{F}L = Id_{TQ}$, so as $\mathcal{F}L$ is a diffeomorphism, $\mathcal{F}H = \mathcal{F}L^{-1}$.

Finally, from uniqueness of X_L and equations (1.7) and (1.11) we get $X_L = (\mathcal{F}H)_* X_H$. ■

Consider now an external force $F\colon TQ \longrightarrow T^*Q$ and denote by F^H to the pullback of the force by $\mathcal{F}H$, *i.e.*, $F^H = (\mathcal{F}H)^*F\colon T^*Q \to T^*Q$.

$$\begin{tikzcd} T^*Q \arrow[rr, "F^H"] \arrow[dr, "\pi_Q"'] & & T^*Q \arrow[dl, "\pi_Q"] \\ & Q & \end{tikzcd}$$

Analogously to the Lagrangian case, it is possible to modify the Hamiltonian vector field X_H to obtain the so-called **forced Hamilton equations** as the integral curves of the vector field $X_H + Y_F^V$ where $Y_F^V \in \mathfrak{X}(T^*Q)$ is defined by

$$Y_F^V(\alpha_q) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}(\alpha_q + tF^H(\alpha_q)) \, .$$

They satisfy

$$\imath_{X_H+Y_F^V}\omega_Q = \mathrm{d}H - \mu_F^H \, , \tag{1.12}$$

where $\mu_F^H := (T\pi_Q)^*F$ is the (horizontal) one-form induced by F. We will say the the **forced Hamiltonian system** is determined by the pair (H, F^H). Locally,

$$Y_F^V = F_i\left(q^j, \frac{\partial H}{\partial p_j}(q,p)\right)\frac{\partial}{\partial p_i} = F_i^H(q,p)\frac{\partial}{\partial p_i} \, ,$$

modifying Hamilton equations as follows

$$\begin{aligned} \frac{\mathrm{d}q^i}{\mathrm{d}t} &= \frac{\partial H}{\partial p_i}(q,p) \, , && (1.13) \\ \frac{\mathrm{d}p_i}{\mathrm{d}t} &= -\frac{\partial H}{\partial q^i}(q,p) + F_i^H(q,p) \, . && (1.14) \end{aligned}$$

1.2.6 Extension to the time-dependent case

Let Q be a manifold and TQ its tangent bundle. As usual, we have natural coordinates (t, q^i, v^i) on $\mathbb{R} \times TQ$ which is the appropriate velocity phase space for time-dependent systems. In that case, it is possible to check that the energy $E_L\colon \mathbb{R} \times TQ \to \mathbb{R}$ is not preserved in general since

$$\frac{\mathrm{d}E_L}{\mathrm{d}t} = \frac{\partial L}{\partial t} \, .$$

We now pass to the Hamiltonian formalism using the (time-dependent) Legendre transformation

$$\mathcal{F}L\colon \mathbb{R} \times TQ \longrightarrow \mathbb{R} \times T^*Q$$

such that, for all $t \in \mathbb{R}$ and $v_q, w_q \in TQ$,

$$\langle \mathcal{F}L(t, v_q), (t, w_q)\rangle := \frac{\partial}{\partial s}\Big|_{s=0} L(t, v_q + sw_q) = D_3 L_{v_q}(w_q) \, .$$

Then, natural coordinates on $\mathbb{R} \times T^*Q$ are (t, q^i, p_i). Locally,

$$\mathcal{F}L(t, q^i, v^i) = \left(t, q^i, \frac{\partial L}{\partial \dot{q}^i}\right) .$$

We assume that the Legendre transformation is a diffeomorphism (Lagrangian hyperregular) and then define the Hamiltonian function $H\colon \mathbb{R} \times T^*Q \to \mathbb{R}$ by

$$H = E_L \circ (\mathcal{F}L)^{-1} ,$$

inducing the **canonical cosymplectic structure** (Ω_H, η) on $\mathbb{R} \times T^*Q$ (Cappelletti-Montano et al., 2013) as

$$\eta = \mathrm{pr}_1^* \mathrm{d}t \, , \qquad \Omega_H = -\mathrm{d}(\mathrm{pr}_2^* \theta_Q - H\eta) = \Omega_Q + \mathrm{d}H \wedge \mathrm{d}t \, ,$$

where pr_i, $i = 1, 2$, are the projections to each Cartesian factor in $\mathbb{R} \times T^*Q$ and θ_Q is the Liouville 1-form on T^*Q. We also denote by $\Omega_Q = -\mathrm{d}\mathrm{pr}_2^*\theta_Q$ the pullback of the canonical symplectic 2-form $\omega_Q = -\mathrm{d}\theta_Q$ on T^*Q. Locally, $\Omega_Q = \mathrm{d}q^i \wedge \mathrm{d}p_i$. Note that now Ω_Q is presymplectic since $\ker \Omega_Q = \mathrm{span}\{\partial/\partial t\}$. In induced coordinates (t, q^i, p_i) on $\mathbb{R} \times T^*Q$,

$$\Omega_H = \mathrm{d}q^i \wedge \mathrm{d}p_i + \mathrm{d}H \wedge \mathrm{d}t \, , \qquad \eta = \mathrm{d}t \, .$$

Define the **evolution vector field** $X_H \in \mathfrak{X}(\mathbb{R} \times T^*Q)$ by

$$i_{X_H} \Omega_H = 0 \, , \qquad i_{X_H} \eta = 1 \, . \tag{1.15}$$

Locally,

$$X_H = \frac{\partial}{\partial t} + \frac{\partial H}{\partial p_i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p_i} \, ,$$

and the integral curves of the evolution vector fields are given by

$$\dot{t} = 1 \, , \qquad \dot{q}^i = \frac{\partial H}{\partial p_i} \, , \qquad \dot{p}_i = -\frac{\partial H}{\partial q^i} \, .$$

These are precisely the curves of the form $t \mapsto \mathcal{F}L(t, \sigma'(t))$ where $\sigma\colon I \to Q$ is a solution of the Euler-Lagrange equations for $L\colon \mathbb{R} \times TQ \to \mathbb{R}$.

From equation (1.15) we deduce that the flow of X_H satisfies the following preservation properties

$$\mathcal{L}_{X_H} \Omega_H = 0 \, , \qquad \mathcal{L}_{X_H} \eta = 0 \, . \tag{1.16}$$

Denote by $\Psi_s\colon \mathcal{U} \subset \mathbb{R} \times T^*Q \longrightarrow \mathbb{R} \times T^*Q$ the flow of the evolution vector field X_H, where $\mathcal{U}$ is an open subset of $\mathbb{R} \times T^*Q$. Observe that

$$\Psi_s(t, \alpha_q) = (t + s, \Psi_{t,s}(\alpha_q)), \qquad \alpha_q \in T_q^*Q,$$

where $\Psi_{t,s}(\alpha_q) = \mathrm{pr}_2(\Psi_s(t, \alpha_q))$. Therefore from the flow of X_H we induce a map

$$\Psi_{t,s}\colon \mathcal{U}_t \subseteq T^*Q \to T^*Q$$

where $\mathcal{U}_t = \{\alpha_q \in T^*Q \mid (t, \alpha_q) \in \mathcal{U}\}$. Conversely, if we know $\Psi_{t,s}$ for all t, we can recover the flow Ψ_s of X_H.

From Equations (1.16), it is straightforward that

$$\Psi_s^*(\Omega_H) = \Omega_H, \qquad \Psi_s^*(\eta) = \eta \,. \tag{1.17}$$

The following theorem relates the previous preservation properties with the symplecticity of the map family $\{\Psi_{t,s} \colon \mathcal{U}_t \subset T^*Q \to T^*Q\}$.

Theorem 1.6. *(Campos et al., 2021, Theorem 4.1) $\Psi_{t,s} \colon \mathcal{U}_t \subseteq T^*Q \to T^*Q$ is a symplectomorphism, i.e., $\Psi_{t,s}^*\omega_Q = \omega_Q$.*

Proof. Firstly, using $T_{(t,\alpha_q)}(\mathbb{R} \times T^*Q) \equiv T_t\mathbb{R} \times T_{\alpha_q}T^*Q$, every vector $V_{(t,\alpha_q)} \in T_{(t,\alpha_q)}(\mathbb{R} \times T^*Q)$ can be split in

$$V_{(t,\alpha_q)} = V_t(\alpha_q) + V_{\alpha_q}(t)\,,$$

where $V_t(\alpha_q) \in T_t\mathbb{R}$ and $V_{\alpha_q}(t) \in T_{\alpha_q}T^*Q$. Observe that $\langle \eta, V_{\alpha_q}(t)\rangle = 0$.

Let choose tangent vectors to the pr_1-fibers (vertical vectors), *i.e.*, $V_{(t,\alpha_q)} \in T_{(\alpha_q,t)}\mathrm{pr}_1^{-1}(t)$. Then the following decomposition holds

$$V_{(t,\alpha_q)} = 0_t + V_{\alpha_q}(t) = V_{\alpha_q}(t) \in T_{\alpha_q}T^*Q\,.$$

From the second equation in (1.17) we obtain

$$0 = \langle \eta_{(t,\alpha_q)}, V_{\alpha_q}(t)\rangle = \langle (\Psi_s^*\eta)_{(t,\alpha_q)}, V_{\alpha_q}(t)\rangle = \langle \eta_{\Psi_s(t,\alpha_q)}, T\Psi_s(V_{\alpha_q}(t))\rangle\,.$$

Hence, $T\Psi_s(V_{\alpha_q}(t)) \in T_{\Psi_{t,s}(\alpha_q)}T^*Q$ and

$$T\Psi_s(V_{\alpha_q}(t)) = 0_{t+s} + T\Psi_{t,s}(V_{\alpha_q}(t)) \equiv T\Psi_{t,s}(V_{\alpha_q}(t))\,.$$

Thus, from the first conservation property in (1.17) we derive

$$\begin{aligned}
(\omega_Q)_{\alpha_q}(V_{\alpha_q}(t), \tilde{V}_{\alpha_q}(t)) &= (\Omega_Q + \mathrm{d}H \wedge \mathrm{d}t)_{(\alpha_q,t)}(V_{\alpha_q}(t), \tilde{V}_{\alpha_q}(t)) \\
&= (\Omega_Q + \mathrm{d}H \wedge \mathrm{d}t)_{(\Psi_{t,s}(\alpha_q),t+s)}(T\Psi_s(V_{\alpha_q}(t)), T\Psi_s(\tilde{V}_{\alpha_q}(t))) \\
&= (\omega_Q)_{\Psi_{t,s}(\alpha_q)}(T\Psi_{t,s}(V_{\alpha_q}(t)), T\Psi_{t,s}(\tilde{V}_{\alpha_q}(t)))
\end{aligned}$$

where $V_{\alpha_q}(t), \tilde{V}_{\alpha_q}(t) \in T_{\alpha_q}T^*Q \equiv T_{(\alpha_q,t)}\mathrm{pr}_1^{-1}(t)$. Therefore, $\Psi_{t,s}^*\omega_Q = \omega_Q$. ∎

1.3 Discrete variational calculus

1.3.1 Discrete Euler-Lagrange equations

Our aim is to approximate some (continuous) Lagrangian system. The idea is to construct a discrete Lagrangian theory in such a way we can derive the equations of motion variationally, and automatically we will obtain conservation of the symplectic forms.

Consider our **state space** to be $Q \times Q$ as a discrete version of TQ and therefore $Q \times Q \times Q \times Q$ as a discrete analogue of $TQ \times TQ$ (Marsden and West, 2001; Marrero et al., 2015). Instead of curves on Q, the trajectories of the dynamical system are replaced by sequences of points in Q. Given some $N \in \mathbb{N}$, we use the notation

$$\mathcal{C}_d(Q) = \left\{ q_d \colon \; \{k\}_{k=0}^{N} \longrightarrow Q \right\} ,$$

for the set of possible solutions, which can be identified with the manifold $Q \times \overset{(N+1)}{\cdots} \times Q$. Now, a **discrete Lagrangian function** is a family of mappings $\left\{ L_d^k \colon Q \times Q \to \mathbb{R} \; : \; k = 0, \ldots, N-1 \right\}$, where parameters $k = 0, \ldots, N-1$ account for the possible time-dependence of the continuous Lagrangian that we want to approximate. The functions $L_d^k \colon Q \times Q \to \mathbb{R}$ can be obtained using some quadrature rule from a continuous action $\mathbb{S}$ in the interval $[kh, (k+1)h]$, $k = 0, \ldots, N-1$, where h is the step-size used for the discretization.

Define a functional, the **discrete action map**, on the space of sequences $\mathcal{C}_d(Q)$ by

$$\mathbb{S}_d(q_d) = \sum_{k=0}^{N-1} L_d^k(q_k, q_{k+1}), \quad q_d \in \mathcal{C}_d(Q).$$

Considering variations of q_d with fixed end points q_0 and q_N and extremizing $\mathbb{S}_d$ over $q_1, \ldots, q_{N-1}$, we obtain the analogous discrete result to Proposition 1.2.

Proposition 1.7 (Discrete Euler-Lagrange equations (DEL)). *The trajectory of a sequence $q_d \in \mathcal{C}_d(Q)$ extremizing the discrete action $\mathbb{S}_d$ is given from Hamilton's principle by*

$$D_1 L_d^k(q_k, q_{k+1}) + D_2 L_d^{k-1}(q_{k-1}, q_k) = 0 \quad \textit{for all} \quad k = 1, \ldots, N-1 \,. \tag{1.18}$$

In the proposition, $D_1 L_d^k(q_k, q_{k+1}) \in T_{q_k}^* Q$ and $D_2 L_d^k(q_k, q_{k+1}) \in T_{q_{k+1}}^* Q$ correspond to $\mathrm{d}L_d^k(q_k, q_{k+1})$ under the identification $T_{(q_k, q_{k+1})}^*(Q \times Q) \cong T_{q_k}^* Q \times T_{q_{k+1}}^* Q$, *i.e.*, $D_i L_d^k(q_k, q_{k+1}) = P_i \, dL_d^k(q_k, q_{k+1})$, where $P_i \colon T_{q_k}^* Q \times T_{q_{k+1}}^* Q \longrightarrow T_{q_{k+i-1}}^* Q$, i=1,2, are the respective projections.

Now the Hessian matrix of each L_d^k is given locally by

$$\mathrm{Hess}(L_d^k) = \left(\frac{\partial^2 L_d^k}{\partial q_k^i \partial q_{k+1}^j} \right)_{ij} .$$

If L_d^k is regular, that is, $\mathrm{Hess}(L_d^k)$ is non-singular, then we can define the so-called **discrete Lagrangian map**

$$\begin{array}{rccc} F_{L_d^k} \colon & Q \times Q & \longrightarrow & Q \times Q \\ & (q_{k-1}, q_k) & \longmapsto & (q_k, q_{k+1}(q_{k-1}, q_k)) \,, \end{array}$$

where q_{k+1} is the unique solution of the DEL (1.18) for the given pair (q_{k-1}, q_k). We can also assure that the discrete Lagrangian map is invertible, thus it is possible to write $q_{k-1} = q_{k-1}(q_k, q_{k+1})$ (Marsden and West, 2001, Theorem 1.5.1). In other words, the system is time-reversible.

In this discrete setting it is possible to define two **discrete Legendre transformations**

$$\mathcal{F}^{+}L_d^k, \mathcal{F}^{-}L_d^k \colon Q \times Q \longrightarrow T^*Q,$$

as each projection can be chosen for the base point. They are defined by

$$\begin{aligned}
\mathcal{F}^{+}L_d^k(q_k, q_{k+1}) &= (q_{k+1}, D_2 L_d^k(q_k, q_{k+1}))\,, \\
\mathcal{F}^{-}L_d^k(q_k, q_{k+1}) &= (q_k, -D_1 L_d^k(q_k, q_{k+1}))\,.
\end{aligned}$$

As usual, the momenta are defined by $p_k := D_2 L_d^{k-1}(q_{k-1}, q_k) = -D_1 L_d^k(q_k, q_{k+1})$. Moreover, take $\Omega_{L_d^k} := (\mathcal{F}^{+}L_d^k)^*\omega_Q = (\mathcal{F}^{-}L_d^k)^*\omega_Q$, with local expression

$$\Omega_{L_d^k}(q_k, q_{k+1}) = \frac{\partial^2 L_d^k}{\partial q_k^i \partial q_{k+1}^j} \mathrm{d}q_k^i \wedge \mathrm{d}q_{k+1}^j\,.$$

We have that $\Omega_{L_d^k}$ is a symplectic 2-form on $Q \times Q$ since L_d^k is regular.

Using these Legendre transformations, the **evolution of the discrete system** on the Hamiltonian side, $\widetilde{F}_{L_d^k} \colon T^*Q \longrightarrow T^*Q$, can also be defined by any of the formulas

$$\widetilde{F}_{L_d^k} = \mathcal{F}^{+}L_d^k \circ (\mathcal{F}^{-}L_d^k)^{-1} = \mathcal{F}^{+}L_d^k \circ F_{L_d^k} \circ (\mathcal{F}^{+}L_d^k)^{-1} = \mathcal{F}^{-}L_d^k \circ F_{L_d^k} \circ (\mathcal{F}^{-}L_d^k)^{-1}\,,$$

due to the commutativity of the following diagram:

$$\begin{tikzcd}
(q_{k-1}, q_k) \arrow[rr, "F_{L_d^k}"] \arrow[dr, "\mathcal{F}^{+}L_d^k"'] & & (q_k, q_{k+1}) \arrow[rr, "F_{L_d^k}"] \arrow[dl, "\mathcal{F}^{-}L_d^k"] \arrow[dr, "\mathcal{F}^{+}L_d^k"'] & & (q_{k+1}, q_{k+2}) \arrow[dl, "\mathcal{F}^{-}L_d^k"] \\
& (q_k, p_k) \arrow[rr, "\widetilde{F}_{L_d^k}"'] & & (q_{k+1}, p_{k+1}) &
\end{tikzcd}$$

From the properties above, the next result is obvious.

Theorem 1.8. *The discrete Hamiltonian map $\widetilde{F}_{L_d^k} \colon (T^*Q, \omega_Q) \longrightarrow (T^*Q, \omega_Q)$ is a symplectomorphism.*

1.3.2 Exact discrete Lagrangian

Up to this point, we have assumed as a starting point a discrete Lagrangian $\left\{L_d^k \colon Q \times Q \longrightarrow \mathbb{R} \,:\, k = 0, \dots, N-1\right\}$. Nevertheless, given a continuous Lagrangian and taking an appropriate discrete Lagrangian, then the DEL equations result in a geometric integrator for the continuous Euler-Lagrange system, called a **variational integrator**. Therefore, given a time-dependent regular Lagrangian function $L \colon \mathbb{R} \times TQ \longrightarrow \mathbb{R}$, we will obtain a discrete Lagrangian $\left\{L_d^k \colon Q \times Q \longrightarrow \mathbb{R} \,:\, k = 0, \dots, N-1\right\}$ as an approximation to the continuous action map. As a matter of fact, for a regular Lagrangian L, and appropriate h, q_0, q_1, we can define the **exact discrete Lagrangian** as

$$L^k_{ed}(q_0, q_1) = \int_{kh}^{(k+1)h} L(t, q_{0,1}(t), \dot{q}_{0,1}(t))\mathrm{d}t,$$

where $q_{0,1}(t)$ is the unique solution of the Euler-Lagrange equations for L (1.5) satisfying $q_{0,1}(kh) = q_0$ and $q_{0,1}\big((k+1)h\big) = q_1$ (Hartman, 2002; Marrero et al., 2021). This is well defined for sufficiently small $h > 0$. Remark the implicit dependence of L^k_{ed} on the time-step $h \in \mathbb{R}$.

The next result is just a particularization of the one in Marsden and West (2001, Theorem 4.6.3) taking constant time-step $t_{k+1} - t_k = h$.

Theorem 1.9. *Let $L\colon \mathbb{R} \times TQ \longrightarrow \mathbb{R}$ be a regular Lagrangian and L^k_{ed} the induced exact discrete Lagrangian. If the time-step h is sufficiently small, then for any solution $q\colon [t_0, t_N] \longrightarrow Q$ of the continuous Euler-Lagrange equations (1.5) we have that $q_d = \{q_k = q(kh)\}_{k=0}^N$ is a solution of the DEL equations (1.18).*

Conversely, for any solution $q_d := \{q_k\}_{k=0}^N$ of the DEL equations (1.18) for L^k_{ed}, define a curve $q\colon [t_0, t_N] \longrightarrow Q$ by

$$q(t) := q_{k,k+1}(t), \quad t \in [kh, (k+1)h]\,,$$

where $q_{k,k+1}$ is the unique solution of the Euler-Lagrange equations (1.5) for L such that $q_{k,k+1}(kh) = q_k$ and $q_{k,k+1}((k+1)h) = q_{k+1}$. Then $q(t)$ is a solution of the Euler-Lagrange equations (1.5) on $[t_0, t_N]$.

In practice, $L^k_{ed}(q_0, q_1)$ will not be provided explicitly, as we need to know the dynamics of the system first. Hence, we will take

$$L^k_d(q_0, q_1) \approx L^k_{ed}(q_0, q_1)\,,$$

using some quadrature rule. We obtain symplectic integrators in this way (Patrick and Cuell, 2009).

Now recall the result of Patrick and Cuell (2009) and Marsden and West (2001) for a *discrete Lagrangian function* $L^k_d\colon Q \times Q \to \mathbb{R}$. Note that now we work with a sample of a discrete Lagrangian $\{L^k_d\colon Q \times Q \to \mathbb{R} \,:\, k = 0, \dots, N\}$.

Definition 1.10. *Let $L^k_d\colon Q \times Q \to \mathbb{R}$, $k = 0, \dots, N$ be a discrete Lagrangian. We say that the discrete Lagrangian is a **discretization of order** r if there exist a relatively compact subset $U \subset TQ$ and constants $C > 0$, $h_1 > 0$ such that*

$$\|L^k_d(q(kh), q((k+1)h)) - L^k_{ed}(q(kh), q((k+1)h))\| \leq C\, h^{r+1}\,,$$

where $q\colon [kh, (k+1)h] \longrightarrow Q$ is the solution of the Euler-Lagrange equations (1.5) with initial conditions $(q_0, \dot{q}_0) \in U$ and for any $h \leq h_1$.

The following result on the order of a variational integrator is obtained (Marsden and West, 2001, Theorem 4.9.2), (Patrick and Cuell, 2009),

Theorem 1.11. *The evolution map* $\widetilde{F}_{L_d^k}\colon T^*Q \longrightarrow T^*Q$ *of a certain discrete Lagrangian function of order* r, $L_d^k\colon Q\times Q \to \mathbb{R}$, *satisfies*

$$\widetilde{F}_{L_d^k} = \widetilde{F}_{L_{ed}^k} + \mathcal{O}(h^{r+1})\,.$$

Equivalently, $\widetilde{F}_{L_d^k}$ *gives a symplectic integrator of order* r *for* $\widetilde{F}_{L_{ed}^k}$.

The order of a given discrete Lagrangian $L_d^k\colon Q\times Q\to\mathbb{R}$ can be computed by expanding $L_d^k(q(kh), q((k+1)h))$ in a Taylor series in h and comparing this to the same expansions for the exact Lagrangian. If the series agree up to r terms, then we have an order r discretization.

1.3.3 Forced discrete mechanics

A very important property of variational integrators involves the ability to adapt them to situations that are more complex, *e.g.*, systems with forces or constraints (Marsden and West, 2001, Section 3.2), (Sato Martín de Almagro, 2019, Section 4.2), (Campos et al., 2015).

The discretization of a continuous force $F\colon TQ\to T^*Q$ gives two different maps $F_d^+\colon Q\times Q\longrightarrow T^*Q$ and $F_d^-\colon Q\times Q\longrightarrow T^*Q$, in analogue with the two Legendre transformations. By definition, they are fibre-preserving mappings in such a way $\pi_Q\circ F_d^+ = \mathrm{pr}_2$ and $\pi_Q\circ F_d^- = \mathrm{pr}_1$.

Definition 1.12 (Discrete Lagrange-d'Alembert principle)**.** *Given a discrete Lagrangian function* $L_d^k\colon Q\times Q\longrightarrow\mathbb{R}$ *with discrete action map* $\mathbb{S}_d\colon \mathcal{C}_d(Q)\longrightarrow\mathbb{R}$ *and discrete forces* $(F_d^k)^\pm\colon Q\times Q\longrightarrow T^*Q$*, the discrete equations of motion are derived from the equation*

$$\langle\delta\mathbb{S}_d(q_d),\,\delta q_d\rangle + \sum_{k=1}^{N-1}\langle (F_d^{k-1})^+(q_{k-1},q_k) + (F_d^k)^-(q_k,q_{k+1}),\,\delta q_k\rangle = 0\,, \tag{1.19}$$

for all variations δq_k *with fixed end points, that is,* $\delta q_0=\delta q_N = 0$.

Simple computations result in the **forced Euler-Lagrange equations**,

$$D_2L_d^{k-1}(q_{k-1},q_k) + D_1L_d^k(q_k,q_{k+1}) + (F_d^{k-1})^+(q_{k-1},q_k) + (F_d^k)^-(q_k,q_{k+1}) = 0\,, \tag{1.20}$$

that implicitly define a discrete forced Lagrangian map $F_{f,L_d^k}\colon Q\times Q\to Q\times Q$.

Analogous to the non-forced case, we are able to define the associated discrete Legendre transformations $\mathcal{F}_f^\pm L_d\colon Q\times Q\to T^*Q$ given by

$$\begin{aligned}
\mathcal{F}_f^+L_d^k(q_k,q_{k+1}) &= \left(q_{k+1}, D_2L_d^k(q_k,q_{k+1}) + (F_d^k)^+(q_k,q_{k+1})\right),\\
\mathcal{F}_f^-L_d^k(q_k,q_{k+1}) &= \left(q_k, -D_1L_d^k(q_k,q_{k+1}) - (F_d^k)^-(q_k,q_{k+1})\right).
\end{aligned}$$

If the system is regular, *i.e.*, the discrete Legendre transformations $\mathcal{F}_f^\pm L_d^k$ are local diffeomorphisms, then we obtain an explicit discrete forced Lagrangian map F_{f,L_d^k}. Furthermore, we have a discrete forced Hamiltonian map

$$\widetilde{F}_{f,L_d^k}\colon T^*Q\longrightarrow T^*Q$$

defined as before by

$$\widetilde{F}_{f,L_d^k} = \mathcal{F}_f^{\pm} L_d^k \circ F_{f,L_d^k} \circ \left(\mathcal{F}_f^{\pm} L_d^k\right)^{-1} .$$

Finally, given a continuous time-dependent system, we can state an analogous result for the order of a discrete variational integrator (see Martín de Diego and Sato Martín de Almagro (2018), Theorem 7.1, for details).

Theorem 1.13. *The evolution map $\widetilde{F}_{f,L_d^k} \colon T^*Q \longrightarrow T^*Q$ of a certain discrete forced Lagrangian function of order r, $L_d^k \colon Q \times Q \to \mathbb{R}$, with discrete forces $(F_d^k)^{\pm} \colon Q \times Q \longrightarrow T^*Q$, satisfies*

$$\widetilde{F}_{f,L_d^k} = \widetilde{F}_{f,L_{ed}^k} + \mathcal{O}(h^{r+1}) .$$

Equivalently, $\widetilde{F}_{f,L_d^k}$ gives an integrator of order r for $\widetilde{F}_{f,L_{ed}^k}$.

1.4 Retraction maps

In this brief section, we set the basis of an object that allows to obtain discrete Lagrangians simply from a continuous Lagrangian. The idea is based on the exponential map (Absil et al., 2008). On a general manifold the tangent line of a curve does not have to be contained in the manifold. A retraction map is precisely the tool to define the notion of moving in a direction of a tangent vector while staying on the manifold. It maps one point and a velocity to another nearby point in such a way certain differential equation is preserved. Because of that, retraction maps play the role of generalizing the linear-search methods in Euclidean spaces to general manifolds and therefore they have been widely used to construct numerical integrators of ordinary differential equations.

Definition 1.14. *(Absil et al., 2008, Section 4.1) A **retraction map** on a manifold Q is a smooth mapping R from the tangent bundle TQ onto Q such that the restriction of R to T_qQ, denoted R_q, satisfies the following properties*

1. *$R_q(0_q) = q$ where $0_q \in T_qQ$.*
2. *With the canonical identification $T_{0_q}T_qQ \cong T_qQ$, R_q satisfies*

$$\mathrm{d}R_q(0_q) = T_{0_q}R_q = \mathrm{Id}_{T_qQ}, \tag{1.21}$$

where Id_{T_qQ} is the identity on T_qQ.

The condition (1.21) is known as **local rigidity condition** since as we announced before, given $v \in T_qQ$, the curve $\gamma_v(t) = R_q(tv)$ has initial velocity v,

$$\dot{\gamma}_v(t) = \langle dR_q(tv), v\rangle \text{ and, in consequence, } \dot{\gamma}_v(0) = \mathrm{Id}_{T_qQ}(v) = v .$$

We see that the exponential map is an obvious example of a retraction map.

A generalization of the notion of retraction map will be very useful later. For a point and a velocity, it will map them into two close points.

Definition 1.15. *(Barbero Liñán and Martín de Diego, 2022, Definition 2.2) A map $R_d\colon U \subset TQ \longrightarrow Q \times Q$ given by*

$$R_d(q,v) = (R^1(q,v),\, R^2(q,v))\,,$$

where U is an open neighborhood of the zero section of TQ, defines a ***discretization map*** *on Q if it satisfies*

1. *$R_d(0_q) = (q,q)$,*
2. *$T_{0_q}R^2_q - T_{0_q}R^1_q\colon T_{0_q}T_qQ \cong T_qQ \longrightarrow T_qQ$ is equal to the identity map on T_qQ for all $q \in Q$.*

Given a discretization map R_d and choosing $R^1(q,v) = q$ then we have that R^2 is a retraction map. Moreover, from the properties in Definition 1.15 it is easy to obtain that any discretization map R_d is a local diffeomorphism from some neighborhood of the zero section of TQ. Therefore, we will be able to use the inverse of some R_d to connect the kinetic energy of some continuous Lagrangian with the discrete one. The following example illustrate this idea.

Example 1.16. Consider $\mathbb{S}^2$ and a particle under the influence of a certain potential $f\colon \mathbb{S}^2 \longrightarrow \mathbb{R}$. The Lagrangian of the particle is

$$\begin{aligned} L\colon T\mathbb{S}^2 &\longrightarrow \mathbb{R} \\ (x,\dot{x}) &\longmapsto L(x,\dot{x}) := \tfrac{1}{2}\|\dot{x}\|^2 - f(x)\,. \end{aligned}$$

The known Euler-Lagrange equations (1.5) for this system are given by

$$\ddot{x} + \nabla f(x) = 0\,.$$

Choosing a discretization map $R_d(x,\dot{x}) = \left(x, \frac{x+\dot{x}}{\|x+\dot{x}\|}\right)$ we can take a discrete Lagrangian by

$$\begin{aligned} L_d^k(x_k, x_{k+1}) &:= L\big(x_k, R_d^{-1}(x_k, x_{k+1})\big) \\ &= \frac{1}{2h^2}\left(\frac{1}{(x_k \cdot x_{k+1})^2} - 1\right) - f(x_k)\,. \end{aligned}$$

Then, the DEL equations (1.18) result

$$\frac{1}{(x_k \cdot x_{k+1})^3}x_{k+1} + \frac{1}{(x_{k-1} \cdot x_k)^3}x_{k-1} + h^2\,\nabla f(x_k) = 0\,,$$

where $x \cdot y$ denotes the usual dot product on $\mathbb{S}^2$. Note that in this case the variational integrator is implicit.

We will see in the following chapters that this notion plays a central role in obtaining accelerated integrators in the case of Lie groups (and also general smooth manifolds). In fact, it will be necessary in order to obtain the family of Nesterov accelerated gradient methods that will depend on the choice of such a retraction.

Chapter 2

Accelerated optimization methods on Lie groups

2.1 Lie groups and retractions

Sophus Lie (1842-1899) developed his seminal work on *continuous transformation groups* during the winter of 1873-74, and a few years later Elie Cartan (1869 - 1951) began to describe Lie group theory in the abstract form we are now familiar with (see `https://mathshistory.st-andrews.ac.uk/Biographies/Lie/`). It would be surprising for these authors how here the general theory finds practical applications in obtaining methods for certain examples of interest a century later (Iserles et al., 2000), such as a neural network to recognize human actions through the rotation matrices that determine their joints (see Chapter 3).

A **Lie group** G (Serre, 1992) is a smooth manifold endowed with a structure of group that agrees with the differentiable structure in the sense multiplication and inversion are smooth mappings. Canonically we have that $\mathfrak{g} := T_eG$ is an algebra (identifying this space with the *left-invariant vector fields* and the usual Lie bracket on this vector space) and it is called the corresponding **Lie algebra**. Classical examples are $\mathrm{GL}(n, \mathbb{K})$, $\mathrm{SO}(n, \mathbb{K})$, $U(n)$, etc. An **action** of a Lie group G on a smooth manifold Q is a smooth mapping

$$\begin{aligned} \phi\colon G \times Q &\longrightarrow Q \\ (g, q) &\longmapsto \phi(g, q) \equiv \phi_g(q)\,, \end{aligned}$$

such that $\phi_e = Id$ and $\phi_{g_1 g_2} = \phi_{g_1}\phi_{g_2}$.

We will use extensively some examples of group action of a Lie group G on itself, such as

$$\begin{aligned} \textbf{Left translation: } & \mathcal{L}_g\colon G \longrightarrow G,\ h \mapsto \mathcal{L}_g(h) := gh\,, \\ \textbf{Right translation: } & \mathcal{R}_g\colon G \longrightarrow G,\ h \mapsto \mathcal{R}_g(h) := hg\,, \\ \textbf{Adjoint map: } & \mathrm{Ad}_g\colon G \longrightarrow G,\ h \mapsto \mathrm{Ad}_g(h) := ghg^{-1}\,. \end{aligned}$$

It is straightforward that both left and right actions are (simply) transitive and they commute. Indeed, they are diffeomorphisms. Particularly, $\mathrm{Ad}_g = \mathcal{L}_g \mathcal{R}_{g^{-1}}$.

Using the corresponding tangent maps, we can transport canonically every vector and 1-form (actually, every tensor) to the Lie algebra $\mathfrak{g}$. In fact, we have the following isomorphisms

$$(\mathcal{L}_{g^{-1}})_* = T_g\mathcal{L}_{g^{-1}},\ (\mathcal{R}_{g^{-1}})_* = T_g\mathcal{R}_{g^{-1}} \colon T_gG \longrightarrow \mathfrak{g}\,,$$
$$(\mathcal{L}_{g^{-1}})^* = T_g^*\mathcal{L}_{g^{-1}},\ (\mathcal{R}_{g^{-1}})^* = T_g^*\mathcal{R}_{g^{-1}} \colon T_g^*G \longrightarrow \mathfrak{g}^*\,.$$

Given a vector $v \in T_gG$ and a 1-form $\alpha \in T_g^*G$, it is usually denoted $(\mathcal{L}_{g^{-1}})_*(v) \equiv g^{-1}v \in \mathfrak{g}$, $(\mathcal{R}_{g^{-1}})_*(v) \equiv vg^{-1} \in \mathfrak{g}$, $(\mathcal{L}_{g^{-1}})^*(\alpha) \equiv g^{-1}\alpha \in \mathfrak{g}^*$ and $(\mathcal{R}_{g^{-1}})^*(\alpha) \equiv \alpha g^{-1} \in \mathfrak{g}^*$. The notation agrees with the usual multiplication of matrices in matrix groups. Clearly, the endomorphisms $(\mathrm{Ad}_g)_* \colon \mathfrak{g} \to \mathfrak{g}$ and $(\mathrm{Ad}_g)^* \colon \mathfrak{g}^* \to \mathfrak{g}^*$ are isomorphisms. It is usual to denote $\mathrm{Ad}_g \equiv (\mathrm{Ad}_g)_*$. Moreover, for $\xi \in \mathfrak{g}$, the **adjoint operator** $\mathrm{ad}_\xi \colon \mathfrak{g} \to \mathfrak{g}$ is defined as $\mathrm{ad}_\xi\,\eta := \langle T_e(\mathrm{Ad}_g\,\eta), \xi\rangle = [\xi, \eta]$, for all $\eta \in \mathfrak{g}$.

Now, these objects allow us to trivialize the tangent TG and cotangent bundle T^*G as $G \times \mathfrak{g}$ and $G \times \mathfrak{g}^*$, respectively. That is

$$\begin{aligned} TG &\longrightarrow G \times \mathfrak{g}, \qquad (g, \dot{g}) \longmapsto (g, T_g\mathcal{L}_{g^{-1}}\dot{g} \equiv g^{-1}\dot{g} = \xi) \\ T^*G &\longrightarrow G \times \mathfrak{g}^*, \qquad (g, \alpha) \longmapsto (g, T_e^*\mathcal{L}_g\alpha \equiv g\alpha = \mu) \end{aligned}$$

As a consequence, choosing a basis in $\mathfrak{g}$ gives a parametrization on every tangent and cotangent space of the Lie group.

2.1.1 Retractions on Lie groups

The classical concept of retraction on a Lie group is subtly different from the definition presented in section 1.4. Given a Lie group G, a **retraction** on G is a mapping $\tau : \mathfrak{g} \longrightarrow G$ satisfying the properties of Definition 1.14 and $\tau(\xi)\tau(-\xi) = e$, $\forall \xi \in \mathfrak{g}$, where e is the neutral element of G. An important example is the exponential map (Kirillov, 2008, Section 3.1), but others will appear in the next sections.

For obtaining variational integrators on Lie groups we will need the left- or right-trivialization of the inverse of a certain retraction. Given $\tau \colon \mathfrak{g} \longrightarrow G$ a retraction on G, define its **right-trivialized tangent** (Bou-Rabee and Marsden, 2009) as a mapping $\mathrm{d}\tau \colon \mathfrak{g} \times \mathfrak{g} \longrightarrow \mathfrak{g}$ such that

$$T_\xi\tau = (\mathcal{R}_{\tau(\xi)})_*\,\mathrm{d}\tau_\xi\,,$$

where $\mathrm{d}\tau_\xi := \mathrm{d}\tau(\xi, \cdot) \colon \mathfrak{g} \to \mathfrak{g}$, for every $\xi \in \mathfrak{g}$. The left-trivialized tangent is defined analogously.

$$\begin{tikzcd} \mathfrak{g} \arrow[rr, "T_\xi\tau"] \arrow[dr, "d\tau_\xi"'] & & T_{\tau(\xi)}G \\ & \mathfrak{g} \arrow[ur, "(\mathcal{R}_{\tau(\xi)})_*"'] & \end{tikzcd}$$

The **inverse right-trivialized tangent** of τ is the mapping $\mathrm{d}\tau^{-1} \colon \mathfrak{g} \times \mathfrak{g} \longrightarrow \mathfrak{g}$ such that

$$T_g\tau^{-1} = \mathrm{d}\tau_\xi^{-1}\,(\mathcal{R}_{g^{-1}})_*, \quad \mathrm{d}\tau_\xi^{-1} \circ \mathrm{d}\tau_\xi = \mathrm{Id}_{\mathfrak{g}}\,,$$

for all $\xi \in \mathfrak{g}$, where $g = \tau(\xi)$.

$$\begin{tikzcd}
T_gG \arrow[rr, "T_g\tau^{-1}"] \arrow[dr, "(\mathcal{R}_{g^{-1}})_*"'] & & \mathfrak{g} \\
& \mathfrak{g} \arrow[ur, "d\tau_\xi^{-1}"'] &
\end{tikzcd}$$

Note that in this case $g^{-1} = \tau(\xi)^{-1} = \tau(-\xi)$.

Proposition 2.1. *Given a retraction $\tau\colon \mathfrak{g} \longrightarrow G$ on a Lie group G, the following properties are satisfied:*

1. $\mathrm{d}\tau_\xi = \mathrm{Ad}_{\tau(\xi)}\, \mathrm{d}\tau_{-\xi}$,
2. $\mathrm{d}\tau_\xi^{-1} = \mathrm{d}\tau_{-\xi}^{-1}\, \mathrm{Ad}_{\tau(-\xi)}$.

Proof. Differentiate equation $\tau(\xi)\tau(-\xi) = e$ to obtain

$$T_{-\xi}\tau = \tau(-\xi)\, T_\xi\tau\, \tau(-\xi) \equiv (\mathcal{L}_{\tau(-\xi)})_*\, (\mathcal{R}_{\tau(-\xi)})_*\, T_\xi\tau\,.$$

Now, using the definition,

$$T_{-\xi}\tau = (\mathcal{R}_{\tau(-\xi)})_*\, \mathrm{d}\tau_{-\xi}\,.$$

Therefore, taking into account both expressions and simplifying, it turns out

$$(\mathcal{L}_{\tau(\xi)})_*\, \mathrm{d}\tau_{-\xi} = (\mathcal{R}_{\tau(\xi)})_*\, \mathrm{d}\tau_\xi\,,$$

which implies item (1).

On the other hand, compose in (1) by $\mathrm{d}\tau_\xi^{-1}$ to get

$$\mathrm{Id}_{\mathfrak{g}} = \mathrm{Ad}_{\tau(\xi)}\, \mathrm{d}\tau_{-\xi}\, \mathrm{d}\tau_\xi^{-1}\,.$$

It suffices to solve this expression for $\mathrm{d}\tau_\xi^{-1}$ to obtain item (2). ∎

2.1.2 The rotations group SO(3)

Some examples will be treated later to prove the performance of our accelerated methods and, in the case of Lie groups, they will be defined on the rotations group $SO(3)$. In fact, this space is ubiquitous when dealing with computational mechanics. As we have said, on every Lie group there is a well-defined **exponential map** (Kirillov, 2008), and in this particular space it has a simple expression. The so-called **Cayley map** will be defined as a more efficient example of retraction, in the sense that it has a lower computational cost than that of the exponential.

The **orthogonal group** O(3) is defined to be the set of all matrices $R \in \mathrm{GL}(3,\mathbb{R})$ such that $R^t = R^{-1}$, i.e., the linear maps on $\mathbb{R}^3$ that preserve the scalar product. Clearly, if $R \in \mathrm{O}(3)$ it turns out $\det R \in \{-1, 1\}$. The **special orthogonal group** SO(3) is the connected component of $\det^{-1}(\{1\})$ in O(3), that is, $R \in \mathrm{SO}(3)$ if and only if $R^t R = I_3$ and $\det R = 1$. It is usually called the *group of rotations* because it corresponds physically with the set of rotation maps on the Euclidean space $\mathbb{R}^3$. This definition provides SO(3) with a structure of connected and compact Lie group (Cushman and Bates, 2015). Its Lie algebra is

$$\mathfrak{so}(3) = T_e\mathrm{SO}(3) = \{A \in \mathrm{GL}(3,\mathbb{R}) \,:\, A + A^t = 0\}\,,$$

that is, $\mathfrak{so}(3)$ is the set of 3×3 skew-symmetric real matrices. The associated Lie bracket $[\,,]$ on $\mathfrak{so}(3)$ is given by the relations

$$[E_i, E_j] = \sum_{k=1}^{3} \epsilon_{ijk} E_k, \quad i,j,k \in \{1,2,3\}\,,$$

where ϵ_{ijk} is the usual Levi-Civita symbol ($\epsilon_{ijk} = 0$ if any of the subscripts match, $\epsilon_{ijk} = 1$ if (ijk) is an even permutation of (123) and $\epsilon_{ijk} = -1$ if (ijk) is an odd permutation) and $\{E_1, E_2, E_3\}$ is the standard basis of $\mathfrak{so}(3)$,

$$E_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Now, $\mathfrak{so}(3)$ has a natural scalar product

$$\begin{aligned} k\colon \mathfrak{so}(3) \times \mathfrak{so}(3) &\longrightarrow \mathbb{R} \\ (A,B) &\longmapsto k(A,B) = \frac{1}{2}\operatorname{tr}(AB^t) = -\frac{1}{2}\operatorname{tr}(AB), \end{aligned}$$

where tr is the trace map. k is positive definite and is called the **Killing metric**. In the following, denote $\|A\|$ the norm of $A \in \mathfrak{so}(3)$ corresponding to k, and $|x|$ the usual euclidean norm of $x \in \mathbb{R}^3$. Let see that this metric allows to identify $\mathfrak{so}(3)$ with $\mathbb{R}^3$. Consider the linear map $\vee\colon \mathfrak{so}(3) \longrightarrow \mathbb{R}^3$ defined by

$$\begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}^{\vee} = (a_1, a_2, a_3) =: a$$

Note that $(E_1)^{\vee} = (1,0,0), (E_2)^{\vee} = (0,1,0)$ e $(E_3)^{\vee} = (0,0,1)$. It is obviously an isomorphism and its inverse is denoted by $\wedge\colon \mathbb{R}^3 \longrightarrow \mathfrak{so}(3)$. Given $x \in \mathbb{R}^3$, denote $\hat{x} = \wedge(x)$. Remark the two following relations

$$\operatorname{tr}(\hat{x}^2) = -2|x|^2\,, \tag{2.1a}$$
$$\hat{x}^3 = -|x|^2\hat{x}\,, \tag{2.1b}$$

for all $x \in \mathbb{R}^3$ (see Appendix A).

Lemma 2.2. *1. $\vee\colon (\mathfrak{so}(3), k) \longrightarrow (\mathbb{R}^3, (\,,))$ is an isometry,*

2. $\vee\colon (\mathfrak{so}(3), [\,,]) \longrightarrow (\mathbb{R}^3, \times)$ is a Lie algebra isomorphism, where $\times$ denotes the vector product on $\mathbb{R}^3$.

Proof. Let $A = \sum_{i=1}^{3} a_i E_i$ and $B = \sum_{i=1}^{3} b_i E_i$. Then

$$k(A,B) = -\frac{1}{2}\mathrm{tr}(AB) = a_1 b_1 + a_2 b_2 + a_3 b_3 = a \cdot b,$$

where $a := (a_1, a_2, a_3) = A^\vee$, $b := (b_1, b_2, b_3) = B^\vee \in \mathbb{R}^3$. For the second point, it turns out

$$([A,B])^\vee = \left(\sum_{i,j,k=1}^{3} \epsilon_{ijk} a_i b_j E_k\right)^\vee = \left(\sum_{i,j=1}^{3} \epsilon_{ijk} a_i b_j\right)_{k=1,2,3} = a \times b = Ab.$$

■

Now, we can introduce the two main retractions defined on the rotations group. The exponential map $\exp\colon \mathfrak{so}(3) \longrightarrow \mathrm{SO}(3)$ is

$$\hat{x} \in \mathfrak{so}(3) \longmapsto \exp\hat{x} \equiv e^{\hat{x}} = \sum_{n=0}^{\infty} \frac{1}{n!}\hat{x}^n = I_3 + \hat{x} + \frac{\hat{x}^2}{2!} + \ldots + \frac{\hat{x}^n}{n!} + \ldots$$

Given $\hat{x} \in SO(3)$, we have easily that $(\exp\hat{x})^T = \exp(-\hat{x}) = (\exp\hat{x})^{-1}$. Thus, using that exp is continuous and $\det \circ \exp(0) = 1$, exp is well-defined. On the other hand, the Cayley map $\mathrm{cay}\colon \mathfrak{so}(3) \longrightarrow \mathrm{SO}(3)$ is given by (Iserles et al., 2000)

$$\hat{x} \in \mathfrak{so}(3) \longmapsto (I - \hat{x})^{-1}(I + \hat{x}) \in SO(3)\,.$$

Note that $(I - \hat{x})$ is always invertible, since $(I - \hat{x})y = 0$ implies

$$|y|^2 = y \cdot \hat{x}y = y \cdot (x \times y) = 0\,.$$

It is straightforward that $\mathrm{cay}(\hat{x}) \in \mathrm{SO}(3)$, for all $x \in \mathbb{R}^3$.

We will need the inverse of these two retraction maps wherever they are defined. In the case of the exponential, we will obtain its inverse if the elements have norm lower than π. Moreover, by definition, it is easy to obtain that $\mathrm{cay}^{-1}(R) = (R + I)^{-1}(R - I)$ if $(R + I)$ is an invertible matrix. The condition $(R + I)$ being invertible is equivalent to R not having -1 as an eigenvalue, or $\mathrm{tr}\, R \neq -1$. Equivalently, R is not a rotation of angle π.

Proposition 2.3. *Let consider $\hat{x} \in \mathfrak{so}(3)$ and $R \in \mathrm{SO}(3)$. If $r := |x| < \pi$, then the exponential map satisfies*

$$\exp(\hat{x}) = I + \tfrac{\sin r}{r}\hat{x} + \tfrac{1}{2}\tfrac{\sin^2(r/2)}{(r/2)^2}\hat{x}^2\,, \tag{2.2a}$$

$$\log(R) = \tfrac{\phi(R)}{2\sin\phi(R)}(R - R^t)\,, \tag{2.2b}$$

where $\phi(R) := \cos^{-1}(\frac{1}{2}\,\mathrm{tr}\,R - \frac{1}{2})$. If $R = \exp(\hat{x})$, then $\phi(R) = r$.

On the other hand, if $\mathrm{tr}\,R \neq -1$, the Cayley map verifies

$$\mathrm{cay}(\hat{x}) = I + \tfrac{2}{1+r^2}(\hat{x} + \hat{x}^2)\,, \tag{2.3a}$$

$$\mathrm{cay}^{-1}(R) = \tfrac{1}{1+\mathrm{tr}(R)}(R - R^t)\,. \tag{2.3b}$$

Proof. From expression (2.1b), and using Cayley-Hamilton theorem,

$$\begin{aligned}\hat{x}^{2n+1} &= (-1)^n r^{2n}\hat{x},\\ \hat{x}^{2n+2} &= (-1)^n r^{2n}\hat{x}^2, \quad \text{for } n \geq 0\,.\end{aligned}$$

Directly, we get

$$\begin{aligned}\exp(\hat{x}) &= \sum_{n=0}^{\infty}\frac{\hat{x}^n}{n!}\\ &= I + \frac{1}{r}\left(\sum_{n=0}^{\infty}(-1)^n\frac{r^{2n+1}}{(2n+1)!}\right)\hat{x} - \frac{1}{r^2}\left(\sum_{n=0}^{\infty}(-1)^{n+1}\frac{r^{2n+2}}{(2n+2)!}\right)\hat{x}^2\\ &= I + \frac{\sin r}{r}\hat{x} + \frac{1-\cos r}{r^2}\hat{x}^2\\ &= I + \frac{\sin r}{r}\hat{x} + \frac{1}{2}\frac{\sin^2(r/2)}{(r/2)^2}\hat{x}^2\,.\end{aligned}$$

Now, to obtain log consider $R = \exp(\hat{x})$. Split R in even and odd parts

$$R = I + \tfrac{1}{2}(R - R^t) + \tfrac{1}{2}(R + R^t - 2I)\,.$$

Then, using (2.2a) we get

$$\begin{aligned}\log(R) = \hat{x} &= \tfrac{r}{2\sin r}(R - R^t)\\ \hat{x}^2 &= \tfrac{(r/2)^2}{\sin^2(r/2)}(R + R^t - 2I)\,.\end{aligned}$$

We see now why we have chosen $r \neq \pi$. Since $\lim_{r\to 0}\frac{\sin r}{r} = 0$, the expressions are well-defined for $r = 0$. To obtain r in terms of R take traces in this second expression,

$$-2r^2 = \operatorname{tr}(\hat{x}^2) = \tfrac{r^2}{4\sin^2(r/2)}\operatorname{tr}(R + R^t - 2I) = \tfrac{r^2}{4\sin^2(r/2)}2(\operatorname{tr}(R) - 3)\,.$$

Thus, solving for r and using $\cos(r) = 1 - 2\sin^2(r/2)$, we get

$$\phi(R) := r = \cos^{-1}(\tfrac{1}{2}\operatorname{tr} R - \tfrac{1}{2})\,.$$

On other hand, to obtain the expressions for cay, split

$$(I - \hat{x})^{-1} = I + c_1\hat{x} + c_2\hat{x}^2\,,$$

where we have taken into account equation (2.1b). Then, using

$$(I - \hat{x})^{-1}(I - \hat{x}) = I\,,$$

we get

$$(I - \hat{x})^{-1} = I + \lambda(\hat{x} + \hat{x}^2)\,,$$

where $\lambda = \frac{1}{1+r^2}$. By definition of cay, equation (2.3a) is obtained. Now, if $R = \mathrm{cay}(\hat{x})$, in an analogous fashion to the previous derivation we get

$$\hat{x} = \tfrac{1}{4\lambda}(R - R^t),$$
$$\hat{x}^2 = \tfrac{1}{4\lambda}(R + R^t - 2I)\,.$$

Therefore, taking traces in the second expressions and solving for r^2, it implies $r^2 = \frac{3-\mathrm{tr}\,R}{1+\mathrm{tr}\,R}$. Hence, equation (2.3b) is obtained. ■

From expression (2.3b) we can derive a useful relation between $\mathrm{tr}\,R$ and the norm of the skew-symmetric part of some $R \in \mathrm{SO}(3)$, with $\mathrm{tr}\,R \neq -1$. Consider $\hat{x} = \mathrm{cay}^{-1}(R)$. Then, $r = \frac{1}{1+\mathrm{tr}\,R}\|R - R^t\|$ and, from $r^2 = \frac{3-\mathrm{tr}\,R}{1+\mathrm{tr}\,R}$, we get

$$1 + r^2 = \tfrac{4}{1+\mathrm{tr}\,R}\,.$$

Thus,

$$1 + \left(\tfrac{1}{1+\mathrm{tr}\,R}\|R - R^t\|\right)^2 = \tfrac{4}{1+\mathrm{tr}\,R}.$$

Reorganizing this expression, it is easy to obtain

$$\left(\frac{\mathrm{tr}\,R - 1}{2}\right)^2 + \left\|\frac{R - R^t}{2}\right\|^2 = 1\,. \tag{2.4}$$

Observe that the limit $\mathrm{tr}\,R = -1$ implies $R = R^t$, and R must be a rotation with two eigenvalues $\lambda = -1$. Therefore, exp is a diffeomorphism from $\{\hat{x} \in \mathfrak{so}(3) : 0 \leq |x| < \pi\}$ onto $\{R \in \mathrm{SO}(3) : \mathrm{tr}\,R > -1\}$ and, in fact, this formula shows that the function ϕ is well-defined. Moreover, if $\hat{x} = \mathrm{cay}^{-1}(R)$, for $R \in \mathrm{SO}(3)$ with $\mathrm{tr}\,R \neq -1$, then from (2.4) we get

$$|x|^2 = \frac{4 - (\mathrm{tr}\,R - 1)^2}{(\mathrm{tr}\,R + 1)^2} \geq 0\,.$$

Therefore, cay is a diffeomorphism from $\mathfrak{so}(3)$ onto $\{R \in \mathrm{SO}(3) : \mathrm{tr}\,R > -1\}$.

To end this section about the rotations group, let compute the right-trivialized tangents and inverse maps of these two retractions. Through the identification of the Lie algebra $\mathfrak{so}(3)$ with $\mathbb{R}^3$, suppose the right-trivialized tangents as defined on $\mathbb{R}^3$.

$$\begin{tikzcd}
\mathfrak{so}(3) \equiv T_{\hat{x}}\mathfrak{so}(3) \arrow[r, "T_{\hat{x}}\,\mathrm{cay}", "T_{\hat{x}}\exp"'] & T_R SO(3) \arrow[r, "\mathcal{R}_{\mathrm{cay}(-\hat{x})}", "\mathcal{R}_{\exp(-\hat{x})}"'] & \mathfrak{so}(3) \arrow[d, "\vee"] \\
\mathbb{R}^3 \arrow[u, "\wedge"] \arrow[rr, shift left, "\mathrm{dcay}(\hat{x}),\,\mathrm{dexp}(\hat{x})"] & & \mathbb{R}^3 \arrow[ll, shift left, "\mathrm{dcay}^{-1}(\hat{x}),\,\mathrm{dexp}^{-1}(\hat{x})"]
\end{tikzcd}$$

Proposition 2.4. *Let $\hat{x} \in \mathfrak{so}(3)$. If $r := |x| \in]0, \pi[$, then it is satisfied*

$$\mathrm{dexp}(\hat{x}) = I + \tfrac{1}{2}\tfrac{\sin(r/2)}{(r/2)^2}\hat{x} + \tfrac{r - \sin(r)}{r^3}\hat{x}^2\,, \tag{2.5a}$$
$$\mathrm{dlog}(\hat{x}) = I - \tfrac{1}{2}\hat{x} + \tfrac{1}{2}\tfrac{2 - r\cot(r/2)}{r^2}\hat{x}^2\,. \tag{2.5b}$$

If $x = 0$, then $\operatorname{dexp}(0) = \operatorname{dlog}(0) = I$.

On the other hand, it turns out

$$\operatorname{dcay}(\hat{x}) = \tfrac{2}{1+r^2}(I + \hat{x})\,, \tag{2.6a}$$

$$\operatorname{dcay}^{-1}(\hat{x}) = \tfrac{1}{2}((1+r^2)I - \hat{x} + \hat{x}^2)\,. \tag{2.6b}$$

Proof. Firstly, consider $y \in \mathbb{R}^3$ and a non-vanishing curve $x \colon]-\epsilon, \epsilon[\to \mathbb{R}^3$ such that $|x(t)| < \pi$, $\hat{x}(0) = \hat{x}$ and $\frac{\mathrm{d}}{\mathrm{d}t}\big|_0 \hat{x}(t) = \hat{y}$. Then

$$\begin{aligned} T_{\hat{x}} \exp(\hat{y}) &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_0 \exp(\hat{x}(t)) \\ &= a' \tfrac{x \cdot y}{r} \hat{x} + a\hat{y} + b' \tfrac{x\cdot y}{r} \hat{x}^2 + b(\hat{x}\hat{y} + \hat{y}\hat{x})\,, \end{aligned}$$

where $a(r) := \frac{\sin r}{r}$ and $b(r) := \frac{1-\cos r}{r^2}$. Using equation (2.2a), we get

$$\begin{aligned} \widehat{\operatorname{dexp}(\hat{x})(y)} &= T_{\hat{x}} \exp(\hat{y}) \exp(-\hat{x}) \\ &= a\hat{y} + (\tfrac{a'}{r} + ab'r + ab - a'br)(x \cdot y)\hat{x} + \tfrac{1}{2}(a^2 + b^2 r^2)\widehat{x \times y}\,. \end{aligned}$$

Then, from the expressions in Appendix A and computing the terms in r, $\operatorname{dexp}(\hat{x}) = \left[aI + \frac{1-a}{r^2} x \otimes x + \frac{1}{2} b\hat{x}\right](y)$. Taking into account formula (A.1b), we finally get (2.5a).

On the other hand, it turns out

$$\left(I - \frac{1}{2}\hat{x} + \frac{1 - \frac{1}{2}\frac{a}{b}}{r^2}\right) \operatorname{dexp}(\hat{x}) = I\,.$$

Hence, equation (2.5b) is obtained.

To get the tangents of cay, consider $\alpha(r) := \frac{2}{1+r^2}$. For a curve $x \colon]-\epsilon, \epsilon[\to \mathbb{R}^3$ such that $\hat{x}(0) = \hat{x}$ and $\frac{\mathrm{d}}{\mathrm{d}t}|_0 \hat{x}(t) = \hat{y} \in \mathfrak{so}(3)$, the tangent map of cay is given by

$$\begin{aligned} T_{\hat{x}} \operatorname{cay}(\hat{y}) &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_0 \operatorname{cay}(\hat{x}(t)) \\ &= \alpha\hat{y} + \alpha(\hat{x}\hat{y} + \hat{y}\hat{x}) + \alpha' \tfrac{x\cdot y}{r}(\hat{x} + \hat{x}^2)\,. \end{aligned}$$

Then, using equation (2.3a) and the formulae in Appendix A,

$$\begin{aligned} \widehat{\operatorname{dcay}(\hat{x})(y)} &= T_{\hat{x}} \operatorname{cay}(\hat{y}) \operatorname{cay}(-\hat{x}) \\ &= \alpha\hat{y} + \left(\frac{\alpha'}{r} + \alpha^2\right)(x\cdot y)\hat{x} + \left(\frac{\alpha'}{r} - \frac{\alpha\alpha'}{r} - \alpha\alpha' r + \alpha^2\right)(x \cdot y)\hat{x}^2 + \alpha\widehat{x \times y} \\ &= \alpha(\hat{y} + \widehat{x \times y})\,, \end{aligned}$$

where we have $\frac{\alpha'}{r} + \alpha^2 = 0$ and $\frac{\alpha'}{r} - \frac{\alpha\alpha'}{r} - \alpha\alpha' r + \alpha^2 = 0$. Therefore,

$$\operatorname{dcay}(\hat{x}) = \alpha(I + \hat{x})\,.$$

Finally, from

$$\operatorname{dcay}(\hat{x})\tfrac{1}{2}((1+r^2)I - \hat{x} + \hat{x}^2) = I\,,$$

we obtain equation (2.6b). ∎

The corresponding expressions for left-trivialized tangents are given by

$$\operatorname{dexp}(\hat{x}) = I - \tfrac{1}{2}\tfrac{\sin(r/2)}{(r/2)^2}\hat{x} + \tfrac{r-\sin(r)}{r^3}\hat{x}^2\,, \tag{2.7a}$$

$$\operatorname{dlog}(\hat{x}) = I + \tfrac{1}{2}\hat{x} + \tfrac{1}{2}\tfrac{2-r\cot(r/2)}{r^2}\hat{x}^2\,, \tag{2.7b}$$

$$\operatorname{dcay}(\hat{x}) = \frac{2}{1+r^2}(I-\hat{x})\,, \tag{2.7c}$$

$$\operatorname{dcay}^{-1}(\hat{x}) = \tfrac{1}{2}\big((1+r^2)I + \hat{x} + \hat{x}^2\big)\,. \tag{2.7d}$$

2.2 Euler-Lagrange and Hamilton equations on a Lie group

We have introduced the basic properties of Lie groups and an important example space where we can test our results. At this point, it suffices to develop the discrete mechanics theory to obtain the proposed accelerated methods. First of all, let translate the continuous equations of motion of a Lagrangian or Hamiltonian system to the case of Lie groups using left-trivialization. The corresponding right-trivialization is analogous.

2.2.1 Lagrangian picture

Given a time-dependent Lagrangian $\check{L}\colon \mathbb{R}\times TG \to \mathbb{R}$, we define its left-trivialized version, $L\colon \mathbb{R}\times G\times\mathfrak{g}\to\mathbb{R}$, by the relation

$$L(t,g,\xi) = \check{L}(t,g,(\mathcal{L}_g)_*\xi) \equiv \check{L}(t,g,g\xi)\,.$$

With this trivialized Lagrangian, the classical Euler-Lagrange equations can be rewritten as

Proposition 2.5 (Left-trivialized Euler-Lagrange equations). *The Euler-Lagrange equations for L with fixed-endpoints variations are given by*

$$\tfrac{d}{dt}D_\xi L = \operatorname{ad}^*_\xi(D_\xi L) + \mathcal{L}^*_g\left(D_g L\right)\,, \tag{2.8}$$

together with the so-called ***reconstruction equation***

$$\dot{g} = g\xi\,. \tag{2.9}$$

Proof. As usual, the action functional is $\mathbb{S} = \int_{t_a}^{t_b} L(t,g(t),\xi(t))\mathrm{d}t$. Then, Hamilton principle imply

$$0 = \delta\mathbb{S} = \int_{t_a}^{t_b}\Big(\langle D_g L,\delta g\rangle + \langle D_\xi L,\delta\xi\rangle\Big)\mathrm{d}t\,.$$

Define $\Sigma = g^{-1}\delta g$. Then, using the reconstruction equation, $\delta\dot{g} = \delta g\,\xi + g\,\delta\xi$, and it turns out

$$\begin{aligned}\dot{\Sigma} &= -g^{-1}\dot{g}g^{-1}\delta g + g^{-1}\delta\dot{g}\\ &= -\xi\Sigma + \Sigma\xi + \delta\xi\,.\end{aligned}$$

Hence, integration by parts and using $\delta\xi = \dot{\Sigma} + \mathrm{ad}_\xi \Sigma$, it turns out

$$0 = \int_{t_a}^{t_b} \langle \mathcal{L}_g^* \left(D_g L \right) - \frac{\mathrm{d}}{\mathrm{d}t} \left(D_\xi L \right) + \mathrm{ad}_\xi^* D_\xi L, \Sigma \rangle \, \mathrm{d}t \,,$$

for all variations Σ with vanishing end-points. ■

Note that this equations are the so-called Euler-Poincaré equations (Holm et al., 1998; Martín de Diego, 2018) with an extra term that accounts for the Lagrangian not being left-invariant. In subsection 2.3.1, we will see that the corresponding equations of motion for the rigid body are derived from the Euler-Poincaré equations.

Example 2.6 (Nesterov Lagrangian)**.** Consider the so-called Nesterov Lagrangian $L \in C^\infty(\mathbb{R} \times SO(3) \times \mathfrak{so}(3))$ (Su et al., 2016; Wibisono et al., 2016) given by

$$L = t^3 \left(\tfrac{1}{2} \|\xi\|^2 - \phi(R) \right) ,$$

for any $t \in \mathbb{R}$, $R \in \mathrm{SO}(3)$ and $\xi \in \mathfrak{so}(3)$. In Su et al. (2016), they proved the second-order convergence (in t) of the solutions to the minimum of ϕ. The extension given by Wibisono et al. (2016), Theorem 1.1, and the generalization to Riemannian manifolds by Duruisseaux and Leok (2022), Theorem 3.4, gives the same convergence result to the dynamical trajectories of this Lagrangian system. In Chapter 3, we will see how this is translated to the discrete case.

We have $D_g L = -t^3 \mathrm{d}\phi(R)$ and $D_\xi L = t^3 \xi$. Now, let $\xi = \hat{x}\,, \eta = \hat{y} \in \mathfrak{so}(3)$ and $\mathfrak{so}(3)^* \ni \mu \equiv \hat{z} \in \mathbb{R}^3$. Taking into account equation (A.1e), $\mathrm{ad}_{\hat{x}} \hat{y} = \widehat{x \times y}$, and then

$$\begin{aligned} \langle \mathrm{ad}_\xi^* \mu, \eta \rangle &= \langle \mu, [\xi, \eta] \rangle \equiv k(\hat{z}, \widehat{x \times y}) \\ &= z \cdot (x \times y) = y \cdot (z \times x) \,, \end{aligned}$$

i.e., $\mathrm{ad}_{\hat{x}}^* \mu \equiv -\widehat{x \times z} = -\,\mathrm{ad}_{\hat{x}} \hat{z}$. Therefore, using $\mathrm{ad}_{\hat{x}} \hat{x} = \widehat{x \times x} = 0$, the Euler-Lagrange equations for this system are given by

$$\left. \begin{aligned} \dot{x} + 3\frac{x}{t} + \left(\mathcal{L}_R^* \, \mathrm{d}\phi(R) \right)^\vee &= 0 \\ \dot{R} &= R \, \hat{x} \,. \end{aligned} \right\} \tag{2.10}$$

The variable $x(t)$ is nothing more than the angular velocity of the rotations $R(t)$, so we are dealing with a damped harmonic oscillator with potential the left-trivialization of $\mathrm{d}\phi(R(t))$. In fact, discretizing this system with the trivialization techniques that we are going to introduce in the following sections, they will produce integrators that oscillate around the extremal of the function ϕ and will be faster than the gradient descent due to the dissipation that we see in the linear term in the velocities. In this case, such dissipation is time-dependent, which at large times will produce the desired acceleration (Su et al., 2016). ◇

It should be remarked that there are some equivalent derivations of this left-trivialized equations. Later we will use a Pontryagin version to obtain the discrete Euler-Lagrange equations, that is, including the reconstruction equation explicitly in the extremizing property of the action. The following theorem is the time-dependent version of Bou-Rabee and Marsden (2009), Theorem 3.4.

Theorem 2.7. *The following statement are equivalent:*

1. *The Lagrangian $\check{L}$ satisfy Hamilton principle*

$$\delta \int_{t_a}^{t_b} \check{L}(t, g, \dot{g}) \mathrm{d}t = 0\,,$$

for all variations δg with fixed endpoints, $\delta g(t_a) = \delta g(t_b) = 0$.

2. *The left-trivialized Lagrangian L satisfy*

$$\delta \int_{t_a}^{t_b} L(t, g, \xi) \mathrm{d}t = 0\,,$$

for all variations $\delta\xi = \dot{\Sigma} + \mathrm{ad}_{\xi}\Sigma$, with $\Sigma(t_a) = \Sigma(t_b) = 0$, and together with the reconstruction equation (2.9).

3. *The Lagrangian $\check{L}$ satisfy the Hamilton-Pontryagin principle*

$$\delta \int_{t_a}^{t_b} \big(\check{L}(t, g, v) + \langle p, \dot{g} - v\rangle\big) \mathrm{d}t = 0\,,$$

for all variations δg, δv, δp with fixed endpoints $\delta g(a) = \delta g(b) = 0$.

4. *L satisfy the left-trivialized Hamilton-Pontryagin principle*

$$\delta \int_{t_a}^{t_b} \big(L(t, g, \xi) + \langle \mu, g^{-1}\dot{g} - \xi\rangle\big) \mathrm{d}t = 0\,,$$

for all variations δg, $\delta\xi$, $\delta\mu$ with fixed endpoints $\delta g(a) = \delta g(b) = 0$.

Proof. Above we proved the equivalence between (1) and (2). The corresponding to (3) and (4) is equivalent. For example, let prove the one related to (1) and (3). For the independent variations δp it results

$$\dot{g} = v\,.$$

The remaining terms turn out

$$\begin{aligned} 0 &= \int_{t_a}^{t_b} \big(\langle D_g\check{L}(t, g, v), \delta g\rangle + \langle D_v\check{L}(t, g, v), \delta v\rangle + \langle p, \delta\dot{g}\rangle - \langle p, \delta v\rangle\big) \mathrm{d}t \\ &= \int_{t_a}^{t_b} \big(\langle D_g\check{L}(t, g, v), \delta g\rangle - \langle \tfrac{\mathrm{d}}{\mathrm{d}t} D_v\check{L}(t, g, v), \delta g\rangle\big) \mathrm{d}t\,, \end{aligned}$$

where we have used integration by parts and the first equation $\dot{g} = v$. Hence, we obtain the same equations of motion than the corresponding to (1). ■

Now, the Lagrangian energy

$$E_L(t, g, \xi) = \big\langle D_\xi L(t, g, \xi), \xi\big\rangle - L(t, g, \xi),$$

is not preserved since the system is explicitly time-dependent. In fact, from the first chapter it was already known that

$$\frac{\mathrm{d}E_L}{\mathrm{d}t} = \frac{\partial L}{\partial t}\,.$$

something that will be important in the construction of accelerated methods on Lie groups.

As before, the Euler-Lagrange equations are modified under the presence of external forces, $\check{F}\colon \mathbb{R}\times TG \to T^*G$. The left-trivialized version $F\colon \mathbb{R}\times G\times\mathfrak{g}\to\mathfrak{g}^*$ is defined by

$$F(t,g,\xi) = \mathcal{L}_g^*\left(\check{F}(t,g,(\mathcal{L}_g)_*\,\xi)\right).$$

The resulting equations can be obtained by applying the Lagrange-D'Alembert principle and using left-trivialization as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \xi}\right) = \mathrm{ad}_\xi^*\frac{\partial L}{\partial \xi} + \mathcal{L}_g^*\left(D_g L\right) + F. \tag{2.11}$$

The derivation is analogous to that of Chapter 1.

2.2.2 Hamiltonian picture

Given a time-dependent left-trivialized Lagrangian $L\colon \mathbb{R}\times G\times\mathfrak{g}\to\mathbb{R}$, the left-trivialized Legendre transformation

$$\begin{array}{rrcl} \mathcal{F}L\colon & \mathbb{R}\times G\times\mathfrak{g} & \longrightarrow & \mathbb{R}\times G\times\mathfrak{g}^* \\ & (t,g,\xi) & \longmapsto & (t,g,\mu = D_\xi L) \end{array}$$

is a local diffeomorphism and Equations (2.8) are rewritten as

$$\dot{\mu} = \mathrm{ad}_{D_\mu H}\mu - \mathcal{L}_g^*(D_g H), \tag{2.12}$$

$$\dot{g} = g\, D_\mu H, \tag{2.13}$$

where $\mu\in\mathfrak{g}^*$ and $H\circ\mathcal{F}L = E_L$.

The canonical structures of the cotangent bundle, the Liouville 1-form θ_G and the canonical symplectic 2-form ω_G on T^*G, are now translated to $G\times\mathfrak{g}^*$. That is, using left-trivialization they result (Cushman and Bates, 2015, Appendix A.2, example 2')

$$(\theta_G)_{(g,\alpha)}(\xi_1,\nu_1) \;=\; \langle\alpha,\xi_1\rangle\,, \tag{2.14}$$

$$(\omega_G)_{(g,\alpha)}\left((\xi_1,\nu_1),(\xi_2,\nu_2)\right) \;=\; -\langle\nu_1,\xi_2\rangle + \langle\nu_2,\xi_1\rangle + \langle\alpha,[\xi_1,\xi_2]\rangle\,, \tag{2.15}$$

with $(g,\alpha)\in G\times\mathfrak{g}^*$, where $\xi_i\in\mathfrak{g}$ and $\nu_i\in\mathfrak{g}^*$, $i=1,2$, and we have used the previous identifications. Observe that we are identifying the elements of $T_{\alpha_g}T^*G$ with the pairs $(\xi,\nu)\in\mathfrak{g}\times\mathfrak{g}^*$.

Taking the projections $\mathrm{pr}_1\colon \mathbb{R}\times G\times\mathfrak{g}^*\to\mathbb{R}$ and $\mathrm{pr}_2\colon \mathbb{R}\times G\times\mathfrak{g}^*\to G\times\mathfrak{g}^*$ and the respective pullbacks $\eta = {\mathrm{pr}_1}^*\mathrm{d}t$, $\Omega_G = {\mathrm{pr}_2}^*\omega_G$ and $\Omega_H = \Omega_G + \mathrm{d}H\wedge\eta$, then Equations 2.12 are equivalent to the cosymplectic equations:

$$\imath_{X_H}\Omega_H = 0\,,\qquad \imath_{X_H}\eta = 1\,. \tag{2.16}$$

If we write $X_H(t,g,\mu) = (k_H, \xi_H, \nu_H)$ where $k_H \in \mathbb{R}$, $\xi_H \in \mathfrak{g}$ and $\nu_H \in \mathfrak{g}^*$ then

$$\begin{aligned} k_H &= 1 \\ \xi_H &= \frac{\partial H}{\partial \xi}(t,g,\mu)\,, \\ \nu_H &= \mathrm{ad}^*_{\xi_H}\mu - \mathcal{L}^*_g\left(\frac{\partial H}{\partial g}(t,g,\mu)\right)\,. \end{aligned}$$

In other words, taking $\dot{g} = g\xi_H$, the integral curves of X_H are specifically the solutions of equations (2.12) and (2.13) together with $\dot{t} = 1$.

Finally, mimicking the simplecticity results in Section 1.2.6, we get the conservation of the left-trivialized symplectic structure on the fibers. From equations (2.16), the flow of X_H satisfies

$$\mathcal{L}_{X_H}\Omega_H = 0 \qquad \mathcal{L}_{X_H}\eta = 0 \tag{2.17}$$

Denote $\Psi_s\colon \mathcal{U} \subset \mathbb{R}\times G\times\mathfrak{g}^* \longrightarrow \mathbb{R}\times G\times\mathfrak{g}^*$ the flow of the evolution vector field X_H, with $\mathcal{U}$ an open subset of $\mathbb{R}\times G\times\mathfrak{g}^*$. Again, it splits

$$\Psi_s(t,\alpha_q) = (t+s, \Psi_{t,s}(\alpha_q)), \qquad \alpha_q \in T^*_qQ,$$

where

$$\Psi_{t,s}\colon \mathcal{U}_t \subseteq G\times\mathfrak{g}^* \to G\times\mathfrak{g}^*$$

and $\mathcal{U}_t = \{\alpha_q \in G\times\mathfrak{g}^* \mid (t,\alpha_q)\in\mathcal{U}\}$. From equations (2.17) it turns out

$$\Psi^*_s(\Omega_H) = \Omega_H, \qquad \Psi^*_s(\eta) = \eta\,.$$

The following theorem is just the left-trivialization of Theorem 1.6 and so it is the proof.

Theorem 2.8. $\Psi_{t,s}\colon \mathcal{U}_t \subseteq G\times\mathfrak{g}^* \to G\times\mathfrak{g}^*$ *is a symplectomorphism, i.e.,* $\Psi^*_{t,s}\omega_G = \omega_G$.

Certainly, the same derivation and results hold for right-trivialization.

2.3 Discrete Euler-Lagrange equations on Lie groups

In analogue with the derivation of the Euler-Lagrange equations in the continuous case, the time dependency of the left-trivialized Lagrangian L now will be modelled in the discrete picture using a family of Lagrangians $L^k_d\colon G\times G\to\mathbb{R}$ with $k = 0,\dots,N-1$ (Campos et al., 2021). The structure of Lie group allows to simplify the discrete equations via left/right-trivialization. Using the reconstruction equation, the discrete Lagrangian family is defined by $L^k_d\colon G\times G\to\mathbb{R}$ with arguments (g_k, f_k), where

$$g_{k+1} = g_k f_k, \quad k = 0,\dots N-1\,, \tag{2.18}$$

is the so-called **discrete reconstruction equation**. The trajectories are now $g_d := \{g_k\}_{k=0}^N$, and $f_d := \{f_k\}_{k=0}^{N-1}$ sequences of points in the Lie group G. The discrete action sum is then

$$\mathbb{S}_d(g_d) = \sum_{k=0}^{N-1} L_d^k(g_k, f_k)\,.$$

Since $G \times G$ is a product space, write $dL_d^k = (D_1 L_d^k, D_2 L_d^k)$, with

$$D_1 L_d^k(g_k, f_k) \in T^*_{g_k} G \quad \text{and} \quad D_2 L_d^k(g_k, f_k) \in T^*_{f_k} G\,.$$

Proposition 2.9 (Discrete Euler-Lagrange Equations)**.** *Fixing endpoints $g_0 \in G$ and $g_N \in G$, the discrete left-trivialized Euler-Lagrange equations associated with the discrete action $\mathbb{S}_d$ and the discrete Lagrangian family $\{L_d^k : k = 0, \ldots, N-1\}$ are*

$$\mathcal{L}^*_{g_k}(D_1 L_d^k) - \mathcal{R}^*_{f_k}(D_2 L_d^k) + \mathcal{L}^*_{f_{k-1}}(D_2 L_d^{k-1}) = 0, \quad 1 \le k \le N-1 \tag{2.19}$$

where $D_i L_d^k$ is a shorthand notation for $D_i L_d^k(g_k, f_k)$ for $i = 1, 2$.

Proof. For any $\delta g_k \in T_{g_k} G$ we can write $\delta g_k = g_k \Sigma_k$ where $\Sigma_k \in \mathfrak{g}$ are arbitrary with $\Sigma_0 = 0$ and $\Sigma_N = 0$. Since $f_k = g_k^{-1} g_{k+1}$ then its variations satisfy

$$\begin{aligned} \delta f_k &= -g_k^{-1}\,\delta g_k\, g_k^{-1} g_{k+1} + g_k^{-1} \delta g_{k+1} \\ &= -\Sigma_k g_k^{-1} g_{k+1} + g_k^{-1} g_{k+1} \Sigma_{k+1} \\ &= -\Sigma_k f_k + f_k \Sigma_{k+1} = \operatorname{ad}_{f_k} \Sigma_k. \end{aligned}$$

Then

$$\begin{aligned} \delta \mathbb{S}_d &= \sum_{k=0}^{N-1} \Big[\langle D_1 L_d^k, g_k \Sigma_k \rangle + \langle D_2 L_d^k, -\Sigma_k f_k + f_k \Sigma_{k+1} \rangle \Big] \\ &= \sum_{k=0}^{N-1} \Big[\langle (T_e \mathcal{L}_{g_k})^* (D_1 L_d^k) - (T_e \mathcal{R}_{f_k})^* (D_2 L_d^k) + (T_e \mathcal{L}_{h_{k-1}})^* (D_2 L_d^{k-1}), \Sigma_k \rangle \Big]\,, \end{aligned}$$

for all $\Sigma_k \in \mathfrak{g}$, $k = 1, \ldots, N-1$, and thus we obtain equation (2.19). ∎

As before, if each one of the (left-trivialized) Lagrangians L_d^k are regular, they define a discrete Lagrangian map

$$\begin{aligned} F_{L_d^k} \colon G \times G &\longrightarrow G \times G \\ (g_k, f_k) &\longmapsto \big(g_{k+1}, f_{k+1}(g_k, f_k)\big)\,, \end{aligned}$$

where $g_{k+1} = g_k f_k$ and f_{k+1} is the unique solution of (2.19) for a given pair (g_k, f_k). Then, the two discrete Legendre transformations are given by

$$\begin{aligned} \mathcal{F}^+ L_d^k, \mathcal{F}^- L_d^k \colon G \times G &\longrightarrow G \times \mathfrak{g}^* \\ \mathcal{F}^+ L_d^k(g_k, f_k) &= (g_{k+1}, p_{k+1})\,, \\ \mathcal{F}^- L_d^k(g_k, f_k) &= (g_k, p_k)\,, \end{aligned}$$

where

$$p_{k+1} = \mathcal{L}^*_{f_k} D_2 L_d^k, \tag{2.20}$$

$$p_k = -\mathcal{L}^*_{g_k} D_1 L_d^k + \mathcal{R}^*_{f_k} D_2 L_d^k. \tag{2.21}$$

These two equations are called the **left-trivialized discrete Hamilton equations**. Therefore, the (left-trivialized) evolution of the discrete system on the Hamiltonian side, $\widetilde{F}_{L_d^k}\colon G \times \mathfrak{g}^* \longrightarrow G \times \mathfrak{g}^*$, is defined by any of the formulae

$$\widetilde{F}_{L_d^k} = \mathcal{F}^+ L_d^k \circ (\mathcal{F}^- L_d^k)^{-1} = \mathcal{F}^+ L_d^k \circ F_{L_d^k} \circ (\mathcal{F}^+ L_d^k)^{-1} = \mathcal{F}^- L_d^k \circ F_{L_d^k} \circ (\mathcal{F}^- L_d^k)^{-1}.$$

The crucial point is that the evolution map inherits the conservation of the simplecticity on the fibers $G \times \mathfrak{g}^*$, i.e., $(\widetilde{F}_{L_d^k})^* \omega_G = \omega_G$.

Summarizing, every variational integrator (now trivialized using the Lie group structure) gives a discrete evolution that preserves the structure of the system. Since the methods simulate some mechanical system, that is equivalent to conserve the mechanical structure of the continuous system. In fact, we saw in Section 1.3.2 that if the discrete Lagrangian is of order r, then the evolution map is a symplectic integrator of order r with respect to the exact discrete evolution map. This is the reason behind the good behaviour of the variational (symplectic) methods.

2.3.1 Rigid body

Rigid solid is defined as a collection of particles where we consider that the parameters governing their dynamics are reduced to those corresponding to translations and rotations. In this way, this type of mechanical system appears naturally to which we can apply our theory of variational integrators and obtain its discrete equations of motion. It should be noted that although this is not our objective, it provides an interesting application beyond the accelerated methods.

Consider a Lagrangian

$$L(t, \Omega) = \tfrac{1}{2} e^{ct} (I_1 \Omega_1^2 + I_2 \Omega_2^2 + I_3 \Omega_3^2) = \tfrac{1}{2} e^{ct} \Omega \cdot \mathbb{I} \Omega,$$

where $\Omega \in \mathbb{R}^3$ is the angular momentum, $\mathbb{I}$ is the inertia tensor and $c \in \mathbb{R}$, and the reconstruction equation

$$\dot{R} = R\,\widehat{\Omega}.$$

This system has as an invariant group $SO(3)$ and the corresponding dynamical equations are called **Euler-Poincaré equations** (Holm et al., 1998; Martín de Diego, 2018),

$$\frac{\mathrm{d}}{\mathrm{d}t} D_\Omega L = \mathrm{ad}^*_\Omega D_\Omega L.$$

Then, $D_\Omega L = e^{ct} \mathbb{I} \Omega$ and $\mathrm{ad}^*_\Omega D_\Omega L \equiv -\mathrm{ad}_\Omega D_\Omega L = -e^{ct} \Omega \times \mathbb{I} \Omega$. Simplifying the expressions, the Euler-Poincaré equations (2.8) result

$$\dot{\Omega} + c\Omega + \mathbb{I}^{-1} (\Omega \times \mathbb{I} \Omega) = 0.$$

These are just the classical rigid body equations with a constant dissipation $c \in \mathbb{R}$.

Now, consider a discrete Lagrangian

$$L_d^k := L(kh, \Omega_k) = \frac{1}{2} e^{ckh} \Omega_k \cdot \mathbb{I} \Omega_k \,,$$

where $\widehat{\Omega}_k \simeq \frac{1}{h} R_k^t (R_{k+1} - R_k) = \frac{1}{h}(f_k - I)$, $I \in \mathbb{R}^{3\times 3}$ the identity matrix, is a discretization of the angular velocity. Define $L_d^k(f_k) = L(kh, \frac{1}{h}(f_k - I))$. Therefore,

$$L_d^k(f_k) = -\frac{1}{2h^2} e^{ckh} \operatorname{tr}(f_k \mathbb{I}_d) \,,$$

where $\mathbb{I}_d = \frac{1}{2}\operatorname{tr}(\mathbb{I}) I - \mathbb{I}$. The left-/right-trivialized tangents are given by

$$\begin{aligned}
\langle \mathcal{L}_{f_k}^* DL_d^k, \hat{\chi} \rangle &= -\frac{1}{2h^2} e^{ckh} \operatorname{tr}(f_k \hat{\chi} \mathbb{I}_d) \\
&= -\frac{1}{2h^2} e^{ckh} \chi \cdot (\mathbb{I}_d f_k - f_k^t \mathbb{I}_d)^\vee \,, \\
\langle \mathcal{R}_{f_k}^* DL_d^k, \hat{\chi} \rangle &= -\frac{1}{2h^2} e^{ckh} \operatorname{tr}(\hat{\chi} f_k \mathbb{I}_d) \\
&= -\frac{1}{2h^2} e^{ckh} \chi \cdot (f_k \mathbb{I}_d - \mathbb{I}_d f_k^t)^\vee \,.
\end{aligned}$$

Hence, the discrete Euler-Poincaré equations, i.e., the variational integrator (2.19) of the Rigid Body for this discretization is given by

$$f_k \mathbb{I}_d - \mathbb{I}_d f_k^t = e^{-ch} (\mathbb{I}_d f_{k-1} - f_{k-1}^t \mathbb{I}_d) \,. \tag{2.22}$$

2.4 Discrete Variational Integrators on Lie groups derived from a continuous Lagrangian

It is possible to define a discretization of a mechanical system using simply the continuous Lagrangian, via some retraction. This gives a simple and efficient tool to obtain integrators. Let $L\colon \mathbb{R} \times G \times \mathfrak{g} \longrightarrow \mathbb{R}$ be a left-trivialized Lagrangian. If we consider a method in the Pontryagin version, including an updated version of the discrete reconstruction equation (2.18) $g_k^{-1} g_{k+1} = \tau(h\xi_k)$, the discrete action can be selected as

$$\mathbb{S}_d = \sum_{k=0}^{N-1} \left[h L_d^k(g_k, \xi_k) + \langle p_k, \tau^{-1}(g_k^{-1} g_{k+1}) - h\xi_k \rangle \right] , \tag{2.23}$$

where $L_d^k(g_k, \xi_k)$ is a discretization of a continuous Lagrangian $L(t, g, \xi)$ taking into account the change of variable $g_{k+1} = g_k \tau(h\xi_k)$, p_k is a momenta sequence in $\mathfrak{g}^*$ and $h > 0$ is the time-step. The particular discretization can be obtained from a quadrature rule such as the mid-point or trapezoidal rule. For example, we can take naively a left-hand rectangle approximation $L_d^k(g_k, \xi_k) = L(kh, g_k, \xi_k)$. See Bou-Rabee and Marsden (2009) for extensions to variational partitioned Runge–Kutta (VPRK) methods on Lie groups.

Proposition 2.10. *The discrete Lie variational integrator (LVI) associated with the discrete action (2.23) is given by*

$$g_k = g_{k-1}\tau(h\xi_{k-1})\,, \tag{2.24a}$$
$$(\mathrm{d}\tau^{-1}_{h\xi_k})^*(p_k) = (\mathrm{d}\tau^{-1}_{-h\xi_{k-1}})^*(p_{k-1}) + h\mathcal{L}^*_{g_k} D_1 L^k_d\,, \tag{2.24b}$$
$$p_k = D_2 L^k_d\,, \tag{2.24c}$$

where $L^k_d \equiv L^k_d(g_k,\xi_k)$ is a discrete Lagrangian and $\mathrm{d}\tau^{-1}$ is the inverse right-trivialized tangent of τ.

Proof. Firstly, extremizing the independent variations of ξ_k implies

$$\delta\xi_k\colon\; hD_2L^k_d - hp_k = 0 \implies p_k = D_2L^k_d\,. \tag{2.25}$$

Analogously for the variations of p_k,

$$\delta p_k\colon\; \tau^{-1}(f_k) - h\xi_k = 0 \implies \tau(h\xi_k) = f_k\,. \tag{2.26}$$

Finally, for the variations of g_k we have

$$\begin{aligned} 0 &= \langle hD_1L^k, \delta g_k\rangle + \langle p_k, T_{f_k}\tau^{-1}(-g_k^{-1}\delta g_k f_k)\rangle + \langle p_{k-1}, T_{f_{k-1}}\tau^{-1}(g_{k-1}^{-1}\delta g_k)\rangle \\ &= \langle hD_1L^k - \mathcal{L}^*_{g_k^{-1}}\mathcal{R}^*_{f_k} T^*_{f_k}\tau^{-1}(p_k) + \mathcal{L}^*_{g_{k-1}^{-1}} T^*_{f_{k-1}}\tau^{-1}(p_{k-1}), \delta g_k\rangle\,, \end{aligned}$$

for each δg_k. Now, using the definition of inverse right-trivialized tangent, $T^*_g\tau^{-1} = \mathcal{R}^*_{g^{-1}}(d\tau^{-1}_\xi)^*$, for $g = \tau(\xi) \in G$. From (2.26), $\tau(h\xi_k) = f_k$, and then it turns out

$$\begin{aligned} 0 &= hD_1L^k - \mathcal{L}^*_{g_k^{-1}}\mathcal{R}^*_{f_k}\mathcal{R}^*_{f_k^{-1}}(\mathrm{d}\tau^{-1}_{h\xi_k})^*(p_k) + \mathcal{L}_{g_{k-1}^{-1}}\mathcal{R}^*_{f_{k-1}^{-1}}(\mathrm{d}\tau^{-1}h\xi_{k-1})^*(p_{k-1}) \\ &= hD_1L^k - \mathcal{L}^*_{g_k^{-1}}(\mathrm{d}\tau^{-1}_{h\xi_k})^*(p_k) + \mathcal{L}_{g_{k-1}^{-1}}\mathcal{R}^*_{f_{k-1}^{-1}}(\mathrm{d}\tau^{-1}_{h\xi_{k-1}})^*(p_{k-1})\,. \end{aligned}$$

Therefore, applying $\mathcal{L}^*_{g_k}$ and using Proposition 2.1, we get

$$\begin{aligned} (\mathrm{d}\tau^{-1}_{h\xi_k})^*(p_k) &= \mathcal{L}^*_{g_k}\mathcal{L}^*_{g_{k-1}^{-1}}\mathcal{R}^*_{f_{k-1}^{-1}}(\mathrm{d}\tau^{-1}_{h\xi_{k-1}})^*(p_{k-1}) + h\mathcal{L}^*_{g_k}D_1L^k_d \\ &= (\mathrm{d}\tau^{-1}_{-h\xi_{k-1}})^*(p_{k-1}) + h\mathcal{L}^*_{g_k}D_1L^k_d\,. \end{aligned}$$

■

Given $\{g_{k-1},\xi_{k-1},p_{k-1}\}$ we obtain $\{g_k,\xi_k,p_k\}$ from (2.24). Note that the first equation is explicit, while the others are possibly implicit.

Finally, consider a system of discrete forces

$$F^+_k, F^-_k : G\times G \longrightarrow T^*G\,, \tag{2.27}$$

such that $F^+_k(g_k,g_{k+1}) \in T^*_{g_{k+1}}G$ and $F^-_k(g_k,g_{k+1}) \in T^*_{g_k}G$. Then the forced LVI (FLVI) associated with the discrete action (2.23) and the discrete forces (2.27) is given from (1.20) and (2.24) by

$$\begin{aligned} (\mathrm{d}\tau^{-1}_{h\xi_k})^*(p_k) &= (\mathrm{d}\tau^{-1}_{-h\xi_{k-1}})^*(p_{k-1}) + h\mathcal{L}^*_{g_k}D_1L^k_d \\ &\quad + \mathcal{L}^*_{g_k}\left[F^-_k(g_k, g_k\tau(h\xi_k)) + F^+_{k-1}(g_{k-1},g_k)\right]\,, \end{aligned} \tag{2.28}$$

and together with equations (2.24a) and (2.24c).

2.4.1 Variational integrators on the rotations group SO(3)

Consider the special orthogonal group $G = SO(3)$, whose corresponding Lie algebra is $\mathfrak{g} = \mathfrak{so}(3)$, the skew-symmetric matrices. From Section 2.1.2, the expressions for right-trivialized tangents of log and cay are

$$\begin{aligned} \mathrm{dlog}(\hat{x}) &= I - \tfrac{1}{2}\hat{x} + \tfrac{1}{2}\tfrac{2-r\cot(r/2)}{r^2}\hat{x}^2\,, \\ \mathrm{dcay}^{-1}(\hat{x}) &= \tfrac{1}{2}\big((1+r^2)I - \hat{x} + \hat{x}^2\big)\,, \end{aligned}$$

where $r = |x|$, $x \in \mathbb{R}^3$. If we identify $\mathfrak{so}(3)^* \equiv \mathbb{R}^3 \equiv \mathfrak{so}(3)$, take into account that

$$\begin{aligned} \langle (\mathrm{d}\tau_{\hat{x}}^{-1})^* \mu, \hat{y}\rangle &= \langle \mu, \mathrm{d}\tau_{\hat{x}}^{-1}(y)\rangle \\ &\equiv k(\hat{z}, \mathrm{d}\tau_{\hat{x}}^{-1}(y)) = z \cdot \mathrm{d}\tau_{\hat{x}}^{-1}(y)\,. \end{aligned}$$

where k is the Killing metric on $\mathfrak{so}(3)$ and $\mu \equiv \hat{z} \in \mathfrak{so}(3)$ and $\hat{x}, \hat{y} \in \mathfrak{so}(3)$. Then the pullback of the inverse retraction is its transposed matrix, $(\mathrm{d}\tau_{\hat{x}}^{-1})^* = (\mathrm{d}\tau_{\hat{x}}^{-1})^t$.

Hence, considering $\hat{x}_k := h\xi_k \in \mathfrak{so}(3)$, the LVI (2.24) associated with $\tau = \exp$ is

$$\left.\begin{aligned} R_k &= R_{k-1}\exp(\hat{x}_{k-1}) \\ (I + \tfrac{1}{2}\hat{x}_k + \tfrac{1}{2}\tfrac{2-r_k\cot(r_k/2)}{r_k^2}\hat{x}_k^2)(p_k) &= A_k + h(\mathcal{L}^*_{R_k} D_1 L_d^k)^\vee \\ p_k &= (D_2 L_d^k)^\vee\,, \end{aligned}\right\} \tag{2.29}$$

where $A_k = (I - \frac{1}{2}\hat{x}_{k-1} + \frac{1}{2}\frac{2-r_{k-1}\cot(r_{k-1}/2)}{r_{k-1}^2}\hat{x}_{k-1}^2)p_{k-1}$ and $p_k \in \mathbb{R}^3 \equiv \mathfrak{so}(3)^*$. The corresponding LVI (2.24) to $\tau = \mathrm{cay}$ is

$$\left.\begin{aligned} R_k &= R_{k-1}\,\mathrm{cay}(\hat{x}_{k-1}) \\ \tfrac{1}{2}((1+r_k^2)I + \hat{x}_k + \hat{x}_k^2)(p_k) &= B_k + h(\mathcal{L}^*_{R_k} D_1 L_d^k)^\vee \\ p_k &= (D_2 L_d^k)^\vee\,, \end{aligned}\right\} \tag{2.30}$$

where $B_k = \frac{1}{2}\big((1+r_{k-1}^2)I - \hat{x}_{k-1} + \hat{x}_{k-1}^2\big)p_{k-1}$ and $p_k \in \mathbb{R}^3 \equiv \mathfrak{so}(3)^*$. Given a certain discrete Lagrangian L_d^k and $(R_{k-1}, x_{k-1}, p_{k-1})$, to obtain (R_k, x_k, p_k) the possible difficulty could be to compute the second and third terms for x_k, p_k. Observe that the terms $A_k, B_k \in \mathbb{R}^3$ are known at every step, since they depend only on $\{x_{k-1}, p_{k-1}\}$.

Example 2.11 (Discretization of Nesterov Lagrangian)**.** Recall the definition of the Nesterov Lagrangian from example 2.6 as $L \in C^\infty(\mathbb{R} \times SO(3) \times \mathfrak{so}(3))$,

$$L(t, R, \xi) := t^3\big(\tfrac{1}{2}\|\xi\|^2 - \phi(R)\big),$$

where ϕ eventually will be a function to be optimized (see Section 2.5). Consider the discrete trajectories as $\{R_k\}_{k=0}^N \subset \mathrm{SO}(3)$, $\{\hat{x}_k = h\xi_k\}_{k=0}^N \subset \mathfrak{so}(3)$ and $\{p_k\}_{k=0}^N \subset \mathbb{R}^3 \equiv \mathfrak{so}(3)^*$. Then, the left-hand rectangle discretization gives

$$L_d^k = L_d^k(R_k, \hat{x}_k) = L(kh, R_k, \tfrac{1}{h}\hat{x}_k) \quad \text{and} \quad p_k \equiv (D_2 L_d^k)^\vee = hk^3 x_k\,,$$

for $k = 0, \ldots, N$. Therefore, if $\alpha_k = \alpha(r_k) = \frac{2 - r_k \cot(r_k/2)}{r_k^2}$, the LVI (2.29) is

$$\left.\begin{aligned} R_k &= R_{k-1} \exp(\hat{x}_{k-1}) \\ (I + \tfrac{1}{2}\hat{x}_k + \tfrac{1}{2}\alpha_k \hat{x}_k^2) h k^3 x_k &= A_k - h^4 k^3 (\mathcal{L}^*_{R_k} \mathrm{d}\phi(R_k))^\vee \\ p_k &= h^3 k^3 x_k \,, \end{aligned}\right\}$$

where now $A_k = (I - \frac{1}{2}\hat{x}_{k-1} + \frac{1}{2}\alpha_{k-1}\hat{x}_{k-1}^2) h k^3 x_{k-1}$. Note that $\hat{x}(x) = x \times x = 0$. Therefore, $A_k = h(k-1)^3 x_{k-1}$ and the LVI for the exponential map is given by

$$\left.\begin{aligned} R_k &= R_{k-1} \exp(\hat{x}_{k-1}) \\ x_k &= (\tfrac{k-1}{k})^3 x_{k-1} - h^3 (\mathcal{L}^*_{R_k} \mathrm{d}\phi(R_k))^\vee \,. \end{aligned}\right\} \tag{2.31}$$

Note that the integrator is independent of the momentum term.

On the other hand, the LVI (2.30) is now

$$\left.\begin{aligned} R_k &= R_{k-1} \operatorname{cay}(\hat{x}_{k-1}) \\ \tfrac{1}{2}\big((1 + r_k^2) I + \hat{x}_k + \hat{x}_k^2\big) h k^3 x_k &= \\ &= B_k - h^4 k^3 (\mathcal{L}^*_{R_k} \mathrm{d}\phi(R_k))^\vee \\ p_k &= h k^3 x_k \,, \end{aligned}\right\}$$

where $B_k = \frac{1}{2}\big((1 + r_{k-1}^2) I - \hat{x}_{k-1} + \hat{x}_{k-1}^2\big) p_{k-1} = \frac{1}{2} h (k-1)^3 (1 + r_{k-1}^2) x_{k-1}$. Therefore, the LVI for the Cayley map turns out

$$\left.\begin{aligned} R_k &= R_{k-1} \operatorname{cay}(\hat{x}_{k-1}) \\ \tfrac{1}{2}(1 + r_k^2) x_k &= \tfrac{1}{2}(\tfrac{k-1}{k})^3 (1 + r_{k-1}^2) x_{k-1} - h^3 (\mathcal{L}^*_{R_k} \mathrm{d}\phi(R_k))^\vee \\ p_k &= h k^3 x_k \,. \end{aligned}\right\} \tag{2.32}$$

This method is eventually explicit. Taking norms, we have to solve the cubic equation

$$(1 + r_k^2) r_k = 2 s_k \,,$$

for $s_k = \left| \frac{1}{2}(\frac{k-1}{k})^3 (1 + r_{k-1}^2) x_{k-1} - h^3 (\mathcal{L}^*_{R_k} \mathrm{d}\phi(R_k))^\vee \right|$. The unique real solution is given by

$$r_k = \left(u_k - \tfrac{1}{3u_k}\right), \quad \text{for} \quad u_k = \sqrt[3]{s_k + \sqrt{s_k^2 + \tfrac{1}{27}}} \,.$$

Once the value of r_k is known, the LVI (2.32) is totally explicit.

Although (2.31) is a more efficient option from this viewpoint, in the following section we will see that the increased efficiency of the Cayley map with respect to the exponential will produce a more accelerated integrator. $\diamond$

2.5 Optimal accelerated methods on Lie groups

Let denote Classical Momentum method, also called Polyak's Heavy Ball, by CM, and Nesterov's Accelerated Gradient by NAG. For a convex $\mathcal{C}^1$ function

$\phi\colon D \subset \mathbb{R}^n \to \mathbb{R}$, where $D \subset \mathbb{R}^n$ is a convex open region, the discrete equations that define CM and NAG are (Campos et al., 2021)

$$\Delta x_k = \mu_k \Delta \left[x_{k-1} - \varepsilon \eta_{k-1} \nabla \phi(x_{k-1})\right] - \eta_k \nabla \phi(x_k)\,, \tag{2.33}$$

where Δ is the forward difference operator, μ_k and η_k are suitable coefficients (the method's strategy), and ε is a boolean coefficient: $\varepsilon = 0$ for CM and $\varepsilon = 1$ for NAG. The terms accompanying ε are associated to a force, hence NAG are in fact CM with forces. Observe that taking $\mu_k = 0$ we obtain gradient descent.

Both families, and more particularly NAG, are usually written in the form

$$y_{k+1} = x_k - \eta_k \nabla \phi(x_k)\,, \tag{2.34a}$$
$$x_{k+1} = y_{k+1} + \mu_k \Delta z_k\,, \tag{2.34b}$$

where the letter z has a different meaning depending on the family of choice, $z \equiv x_{\cdot-1}$ for CM and $z \equiv y$ for NAG. Equation (2.34a) should be viewed as an auxiliary definition that transforms (2.33) into (2.34b) and vice versa. Although x's and y's follow a trajectory towards the argument minimum of ϕ, strictly speaking x_k is the natural one, referred here as "on track", while we refer to y_k as the "off road" trajectory.

Consider now a real-valued $\mathcal{C}^1$-function, $\phi\colon G \to \mathbb{R}$, defined on a given Lie group G. Assume furthermore that ϕ possess a single local minimum in G,

$$g^* = \operatorname{argmin} \phi(g)\,, \quad g \in G\,. \tag{2.35}$$

We define on $G \times \mathfrak{g}$ the discrete time-dependent Lagrangian system with forces

$$L_d^k(g_k, \xi_k) := a_k \tfrac{1}{2} \|\xi_k\|^2 - b_k^- \phi(g_k) - b_{k+1}^+ \phi(g_k \tau(h\xi_k))\,, \tag{2.36a}$$
$$F_k^-(g_k, \xi_k) := -\tfrac{a_{k-1}}{a_k} h(b_k^- + b_k^+) \mathrm{d}\phi(g_k) \in T^*_{g_k} G\,, \tag{2.36b}$$
$$F_k^+(g_k, \xi_k) := \qquad h(b_k^- + b_k^+) \mathcal{R}^*_{\tau(-h\xi_k)}\, \mathrm{d}\phi(g_k) \in T^*_{g_{k+1}} G\,, \tag{2.36c}$$

where $a_k, b_k^{\pm}$ are arbitrary coefficients, τ is certain retraction map on G and $h > 0$ is a fixed constant time-step. Note that we are using a trapezoidal-type rule to approximate the action obtained from the Nesterov Lagrangian. In the case of Example 2.11, we chose $a_k = b_k^- = (hk)^3$ and $b_k^+ = 0$.

The LVI for this system can be computed from (2.24) as follows,

$$\begin{aligned}
D_1 L_d^k &= -b_k^- d\phi(g_k) - b_{k+1}^+ \mathcal{R}^*_{\tau(h\xi_k)} d\phi(g_{k+1})\,,\\
p_k = D_2 L_d^k &= a_k \xi_k^\flat - h b_{k+1}^+ T^*_{h\xi_k} \tau\, \mathcal{L}^*_{g_k} d\phi(g_{k+1})\\
&= a_k \xi_k^\flat - h b_{k+1}^+ (d\tau_{h\xi_k})^*\, \mathcal{R}^*_{f_k} \mathcal{L}^*_{g_k} d\phi(g_{k+1})\,.
\end{aligned}$$

Again $f_k = g_k^{-1} g_{k+1} = \tau^{-1}(h\xi_k) = \tau(-h\xi_k)$, and we obtain

$$\begin{aligned}
(d\tau^{-1}_{h\xi_k})^*(p_k) &= a_k (d\tau^{-1}_{h\xi_k})^*(\xi_k^\flat) - h b_{k+1}^+ \operatorname{Ad}^*_{g_k} \mathcal{R}^*_{g_{k+1}} d\phi(g_{k+1})\,,\\
(d\tau^{-1}_{-h\xi_k})^*(p_k) &= \operatorname{Ad}^*_{f_k} (d\tau^{-1}_{\xi_k})^*(p_k)\\
&= a_k \operatorname{Ad}^*_{f_k} (d\tau^{-1}_{h\xi_k})^*(\xi_k^\flat) - h b_{k+1}^+ \mathcal{L}^*_{g_{k+1}} d\phi(g_{k+1})\,,\\
h \mathcal{L}^*_{g_k} D_1 L_d^k &= -h b_k^- \mathcal{L}^*_{g_k} d\phi(g_k) - h b_{k+1}^+ \mathcal{L}^*_{g_k} \mathcal{R}^*_{f_k} d\phi(g_{k+1})\,.
\end{aligned}$$

The discrete variational integrator of a free/forced system is given from (2.24) and (2.28) by

$$a_{k+1}(d\tau^{-1}_{h\xi_{k+1}})^*(\xi^\flat_{k+1}) - \cancel{hb^+_{k+2}\operatorname{Ad}^*_{g_{k+1}}\mathcal{R}^*_{g_{k+2}}d\phi(g_{k+2})} = a_k \operatorname{Ad}^*_{f_k}(d\tau^{-1}_{\xi_k})^*(\xi^\flat_k)$$
$$- hb^+_{k+1}\mathcal{L}^*_{g_{k+1}}d\phi(g_{k+1}) - hb^-_{k+1}\mathcal{L}^*_{g_{k+1}}d\phi(g_{k+1}) - \cancel{hb^+_{k+2}\mathcal{L}^*_{g_{k+1}}\mathcal{R}^*_{f_{k+1}}d\phi(g_{k+2})} + \epsilon\mathcal{L}^*_{g_{k+1}}(F^-_{k+1}+F^+_k)\,,$$

that is,

$$\begin{aligned} a_{k+1}(d\tau^{-1}_{h\xi_{k+1}})^*(\xi^\flat_{k+1}) =& a_k \operatorname{Ad}^*_{f_k}(d\tau^{-1}_{h\xi_k})^*(\xi^\flat_k) - hb^+_{k+1}\mathcal{L}^*_{g_{k+1}}d\phi(g_{k+1}) \\ & - hb^-_{k+1}\mathcal{L}^*_{g_{k+1}}d\phi(g_{k+1}) + \epsilon\mathcal{L}^*_{g_{k+1}}(F^-_{k+1}+F^+_k)\,, \end{aligned}$$

together with $g_{k+1} = g_k\tau(h\xi_k)$ and $p_k = D_2L^k_d$. As before, ϵ is a boolean coefficient: $\epsilon = 0$ for a free system, $\epsilon = 1$ for a forced system. Define

$$\Delta X_k := \tfrac{1}{h}\operatorname{Ad}^*_{g_k^{-1}}(d\tau^{-1}_{h\xi_k})^*(\xi^\flat_k)\,,$$

therefore applying $\operatorname{Ad}^*_{g^{-1}_{k+1}}$ and dividing by ha_{k+1} we get

$$\Delta X_{k+1} = \mu_{k+1}\left(\Delta X_k - \varepsilon\Delta\left[\eta_k\mathcal{R}^*_{g_k}\mathrm{d}\phi(g_k)\right]\right) - \eta_{k+1}\mathcal{R}^*_{g_{k+1}}\mathrm{d}\phi(g_{k+1}) \in \mathfrak{g}^*\,, \quad (2.37)$$

where $\mu_k := \frac{a_{k-1}}{a_k}$ and $\eta_k := \frac{b^-_k+b^+_k}{a_k}$. Note that solely the bracketed Δ is the usual difference operator, whereas ΔX is merely a suggestive notation.

Although (2.37) is formally identical to (2.33), due to the non-linearity of G, here the Δx terms cannot be split in two as an explicit difference which prevents to rewrite it as in (2.34). Nonetheless (2.37) still defines a family of twin methods, to which we refer to as momentum methods for Lie groups, CM when $\varepsilon = 0$ and NAG when $\varepsilon = 1$.

We can however take advantage of the linear structure of $\mathfrak{g}^*$ to recover something similar to (2.34). Consider the difference of two consecutive equations (2.37), that is,

$$\Delta X_{k+1} = \Delta X_k + \Delta\left[\mu_k\left(\Delta X_{k-1} - \varepsilon\Delta\left[\eta_{k-1}\mathcal{R}^*_{g_{k-1}}\mathrm{d}\phi(g_{k-1})\right]\right) - \eta_k\mathcal{R}^*_{g_k}\mathrm{d}\phi(g_k)\right] \in \mathfrak{g}^*\,, \quad (2.38)$$

which is equivalent to

$$\Delta X_k = \tfrac{1}{h}\operatorname{Ad}^*_{g_k^{-1}}(d\tau^{-1}_{h\xi_k})^*(\xi^\flat_k)\,, \quad (2.39a)$$
$$\Delta Y_{k+1} = \Delta X_k - \Delta\left[\eta_k\mathcal{R}^*_{g_k}\mathrm{d}\phi(g_k)\right]\,, \quad (2.39b)$$
$$\Delta X_{k+1} = \Delta Y_{k+1} + \Delta\left[\mu_k\Delta Z_k\right]\,, \quad (2.39c)$$

where $\Delta Z \equiv \Delta X_{\cdot-1}$ for CM and $\Delta Z \equiv \Delta Y$ for NAG. Equations (2.39b)-(2.39c) are valid for $k \geq 1$, but Equation (2.37) must be considered for $k = 0$, giving

$$\Delta Y_1 = \Delta X_0 - \varepsilon\Delta\left[\eta_0\mathcal{R}^*_{g_0}\mathrm{d}\phi(g_0)\right]\,, \quad (2.40a)$$
$$\Delta X_1 = \mu_1\Delta Y_1 - \eta_1\mathcal{R}^*_{g_1}\mathrm{d}\phi(g_0)\,, \quad (2.40b)$$

where $\varepsilon = 0$ for CM and $\varepsilon = 1$ for NAG.

2.5.1 CM and NAG on the rotations group SO(3)

Equations (2.39) and (2.40) are fully explicit, except for the reconstruction equation (2.39a) that is implicit for ξ_k, and equivalently for g_{k+1}. However, they can be made explicit for the particular case of the group of rotations in $\mathbb{R}^3$, in a similar fashion as the two integrators in the previous section.

Consider the sequences $\{R_k\} \subset \mathrm{SO}(3)$ and $\{x_k\} \subset \mathbb{R}^3$ such that $R_{k+1} = R_k \tau(x_k)$, for $\tau\colon \mathbb{R}^3 \to \mathrm{SO}(3)$ the corresponding retraction. Using the same computations from above, for the exponential map we have

$$\mathrm{dlog}(\hat{x}_k)^*(x_k^\flat) \equiv \mathrm{dlog}(\hat{x}_k)^t(x_k) = x_k\,,$$

where again $\mathfrak{so}(3)^* \equiv \mathbb{R}^3$. Therefore, expression (2.39a) turns out now

$$\Delta X_k = \mathrm{Ad}^*_{R_k^t}\, x_k^\flat \equiv \mathrm{Ad}_{R_k}\, x_k\,. \tag{2.41}$$

Hence, the integrator is fully explicit. In the case of the Cayley map, we get

$$\mathrm{dcay}^{-1}(\hat{x}_k)^*(x_k^\flat) \equiv \mathrm{dcay}^{-1}(\hat{x}_k)^t(x_k) = \tfrac{1}{2}(1 + |x_k|^2)x_k\,.$$

Again, expression (2.39a) can be made explicit using that the resulting third degree polynomial equation for $|x_k|$ is $(1 + |x_k|^2)|x_k| = 2s_k$, where $s_k = |\Delta X_k|$. Therefore, we obtain

$$|x_k| = u_k - \tfrac{1}{3u_k}\,, \tag{2.42}$$

with $u_k = \sqrt[3]{s_k + \sqrt{s_k^2 + \frac{1}{27}}}$, and expression (2.39a) is

$$\Delta X_k = \tfrac{1}{2}\big(1 + \big(u_k - \tfrac{1}{3u_k}\big)^2\big)\, \mathrm{Ad}_{R_k}\, x_k\,. \tag{2.43}$$

Note finally that although the CM and NAG integrators introduced here on the rotations group SO(3) have a lower computational cost for the exponential map from the point of view of equation (2.39a), the methods will be often faster when the Cayley map is used. As we said, this is due to the increased computational efficiency of cay over exp. Of course, there could be some other suitable retractions that improved these integrators. What we have just obtained is a family of CM and NAG methods.

Chapter 3

Simulations

In this chapter, we compute several simulations where we prove the computational efficiency of the CM and NAG methods given in previous sections. These have been implemented in Julia Language v1.7.3 `https://julialang.org/`. The library Plots.jl (Breloff, 2021) is the unique non-native one used here.

Throughout this section, the coefficients μ_k and η_k are selected from a discretization via the trapezoidal rule from the Nesterov Lagrangian (see previous sections and example 2.6). Particularly, the coefficients in (2.36) will be chosen as

$$a_k = \frac{(kh)^3 + ((k+1)h)^3}{2h^2} = \frac{h}{2}(k^3 + (k+1)^3) \quad \text{and} \quad b_k = \frac{(kh)^3}{2},$$

where we have approximate the velocity term to be constant on every interval $[kh, (k+1)h[$. Therefore,

$$\mu_k = \frac{k^3 + (k-1)^3}{k^3 + (k-1)^3} \simeq \frac{2k-3}{2k+3} \quad \text{and} \quad \eta_k = \frac{2k^3}{k^3 + (k+1)^2} h^2 \simeq \frac{2k}{2k+3} h^2, \tag{3.1}$$

where we consider the approximate version of them due to k will be sufficient large in the simulations given here.

As we saw in Example 2.6, we have convergence theorems for the continuous case (Su et al., 2016; Wibisono et al., 2016; Duruisseaux and Leok, 2022) and these techniques allow to translate the mechanical information to the discrete equations. Nevertheless, at this moment there are not results on the convergence of the discrete methods. That is why we compare them using experiments on computers.

As a remark, some other coefficients are possible to be chosen, but here we focus on the case of the Nesterov Lagrangian and dicretization from the trapezoidal rule. See Campos et al. (2021), Subsection 8.1, for other options.

3.1 Rosenbrock function on SO(3)

Rosenbrock function (Rosenbrock, 1960) (also called Rosenbrock's valley or Rosenbrock's banana function) is known as a test function to prove the performance of optimization algorithms. It is defined as

$$\begin{array}{rcl} f\colon \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & f(x,y) := (a-x)^2 + b(y-x^2)^2 , \end{array}$$

where $a, b \in \mathbb{R}$ are usually selected as $a = 1$ and $b = 100$. It represents a valley surrounded by steep walls that makes very difficult the search for minima. Moreover, it is not a convex function but is indeed geodesically convex.

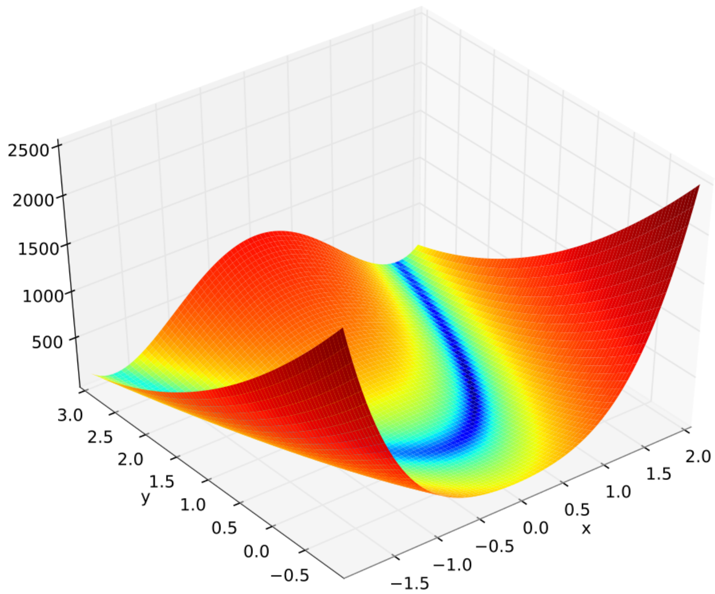


Figure 3.1: Rosenbrock function on $\mathbb{R}^2$ for $a = 1$ and $b = 100$.

In Campos et al. (2021) they give a generalization of this function on more dimensional euclidean spaces, and it takes the following form in the case of $\mathbb{R}^3$,

$$\begin{array}{rcl} f\colon \mathbb{R}^3 & \longrightarrow & \mathbb{R} \\ (x,y,z) & \longmapsto & (1-x)^2 + 100(y-x^2)^2 + (1-y)^2 + 100(z-y^2)^2 , \end{array}$$

with a unique global minimum at $(1,1,1)$ such that $f(1,1,1) = 0$.

We can give a generalization of Rosenbrock function on SO(3) via the natural isomorphism between $\mathbb{R}^3$ and $\mathfrak{so}(3)$, and using a retraction on $\mathfrak{so}(3)$. Abusing the notation, consider $\tau\colon \mathbb{R}^3 \to \mathrm{SO}(3)$ a retraction, and define

$$\begin{array}{rcl} \phi\colon U \subset SO(3) & \longrightarrow & \mathbb{R} \\ R & \longmapsto & f(\tau^{-1}(R)) , \end{array}$$

where U is an open subset containing the global minimum and such that τ^{-1} is well-defined. The geodesic convexity of the Rosenbrock function carries over to ϕ. Since the global minimum of f is placed at $(1,1,1)$, then ϕ counts with a global minimum at $R^* = \tau(1,1,1)$. If $\tau = \mathrm{cay} \circ \wedge$, then the global minimum is

$$R_1^* = \mathrm{cay}(\widehat{(1,1,1)}) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} ,$$

and else if $\tau = \exp \circ \wedge$ then

$$R_2^* = \exp(\widehat{(1,1,1)}) = \begin{pmatrix} 0.226296 & -0.183008 & 0.956712 \\ 0.956712 & 0.226296 & -0.183008 \\ -0.183008 & 0.956712 & 0.226296 \end{pmatrix}.$$

To obtain the optimization methods associated to these functions, it suffices to compute the term $\mathcal{L}_R^*(\mathrm{d}\phi_R)$ or $\mathcal{R}_R^*(\mathrm{d}\phi_R)$ for every $R \in SO(3)$. Given $R \in SO(3)$ and $\hat{y} \in \mathfrak{so}(3)$, it turns out

$$\begin{aligned} \langle \mathcal{L}_R^* \,\mathrm{d}\phi(R), \hat{y} \rangle &= \nabla\phi(\tau^{-1}(R)) \cdot T_R \tau^{-1}(R\hat{y}) \\ &= \nabla\phi(\tau^{-1}(R)) \cdot \mathrm{d}^L \tau^{-1}_{\tau^{-1}(R)}(y)\,, \end{aligned}$$

where d^L refers now to the left-trivialized tangent (see subsection 2.1.1) and "$\cdot$" is the usual scalar product on $\mathbb{R}^3$. Therefore,

$$\left(\mathcal{L}_R^* \,\mathrm{d}\phi(R)\right)^\vee = \left(d^L \tau^{-1}_{\tau^{-1}(R)}\right)^t \nabla f(\tau^{-1}(R))\,. \tag{3.2}$$

Equivalently, the right-trivialized tangent is given by

$$\left(\mathcal{R}_R^* \,\mathrm{d}\phi(R)\right)^\vee = \left(d^R \tau^{-1}_{\tau^{-1}(R)}\right)^t \nabla f(\tau^{-1}(R))\,. \tag{3.3}$$

3.1.1 Numerical experiments

We have computed several numerical experiments, testing CM and NAG with respect to the gradient descent. Here we present an illustrating sample of them. All figures have y-axis in logarithmic scale.

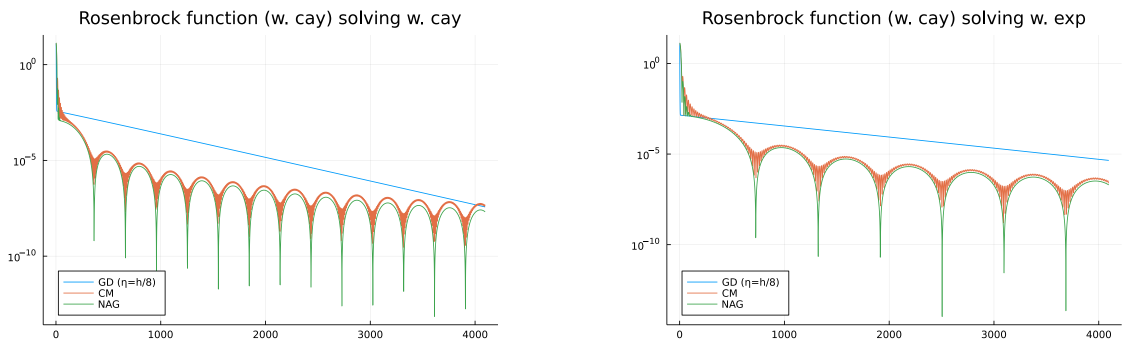


Figure 3.2: Rosenbrock function defined on SO(3) using cay, and the methods obtained from exp and cay, with $R_0 = \tau(1.1, 0.9, 1.0)$ and $h = 0.005$.

In figures 3.2 and 3.3, we choose $h = 0.005$ and $R_0 = \tau(1.1, 0.9, 1.0)$, for the step-size and the initial rotation, respectively, where τ is the corresponding retraction defining ϕ. The step-size used for gradient descent is $\eta = h/8$, but we could choose another smaller value up to h^2 for comparison because it is the one in expressions (3.1). We plot the value of ϕ with respect to the step k. If we

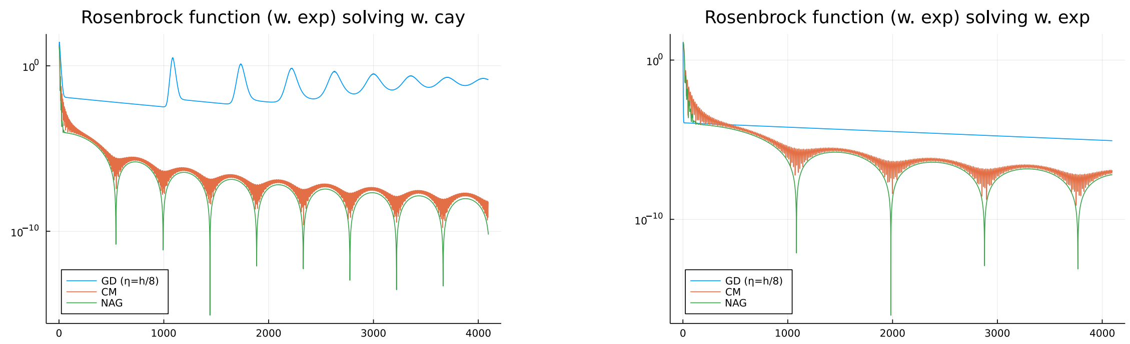


Figure 3.3: Rosenbrock function defined on SO(3) using exp, and the methods obtained from exp and cay, with $R_0 = \tau(1.1, 0.9, 1.0)$ and $h = 0.005$.

define Rosenbrock function on SO(3) using the exponential map and compute gradient descent with the Cayley map, we see that it gets stuck. It has to do with the shape of Rosenbrock function and is a fact common to many of the simulations. It is solved using a smaller step-size.

On the other hand, choosing $h = 0.01$ and $R_0 = \tau(0.5, 0.5, 0.5)$, we get figure 3.4. Due to the higher efficiency in computing the Cayley map, we see that improvement of the methods when are obtained using it instead of the exponential map. The same fact holds for the gradient descent (GD).

The oscillations obtained for CM and NAG in every image are also consequence of the challenge to minimize this function. They pass close to the minimum but the inertia causes them to advance a bit and then go back to look for it. We see also that the forces terms introduced in the case of NAG produce a small improvement of the CM method, leading to a reduction in the number of oscillations, as can be seen in the figure. All the situations show that we have obtained indeed accelerated methods, and they are optimizing this complicated function in a very few steps.

Finally, a known fact when comparing CM and NAG is that the former is more stable. For example, when we choose $h = 0.025$ and $R_0 = \tau(-1, 0, 0)$, the latter does not converge while CM does. Because of that, this is often used instead of NAG, since the difference between their performances is not too important.

3.2 Skeleton-based Action Recognition

Recent advances in Technology make it more important every day to achieve better knowledge on effective techniques to pick up the interaction between humans and computers. Computer vision, psychology, artificial intelligence and more can be gathered in what is called Multimodal Human-Computer interaction (MMHCI) (Sebe et al., 2005). Most research lines usually focus on each modality and later integrate the results in the application phase. In our case, we focus on human action recognition.

In recent years, computer vision researchers have been interested in human

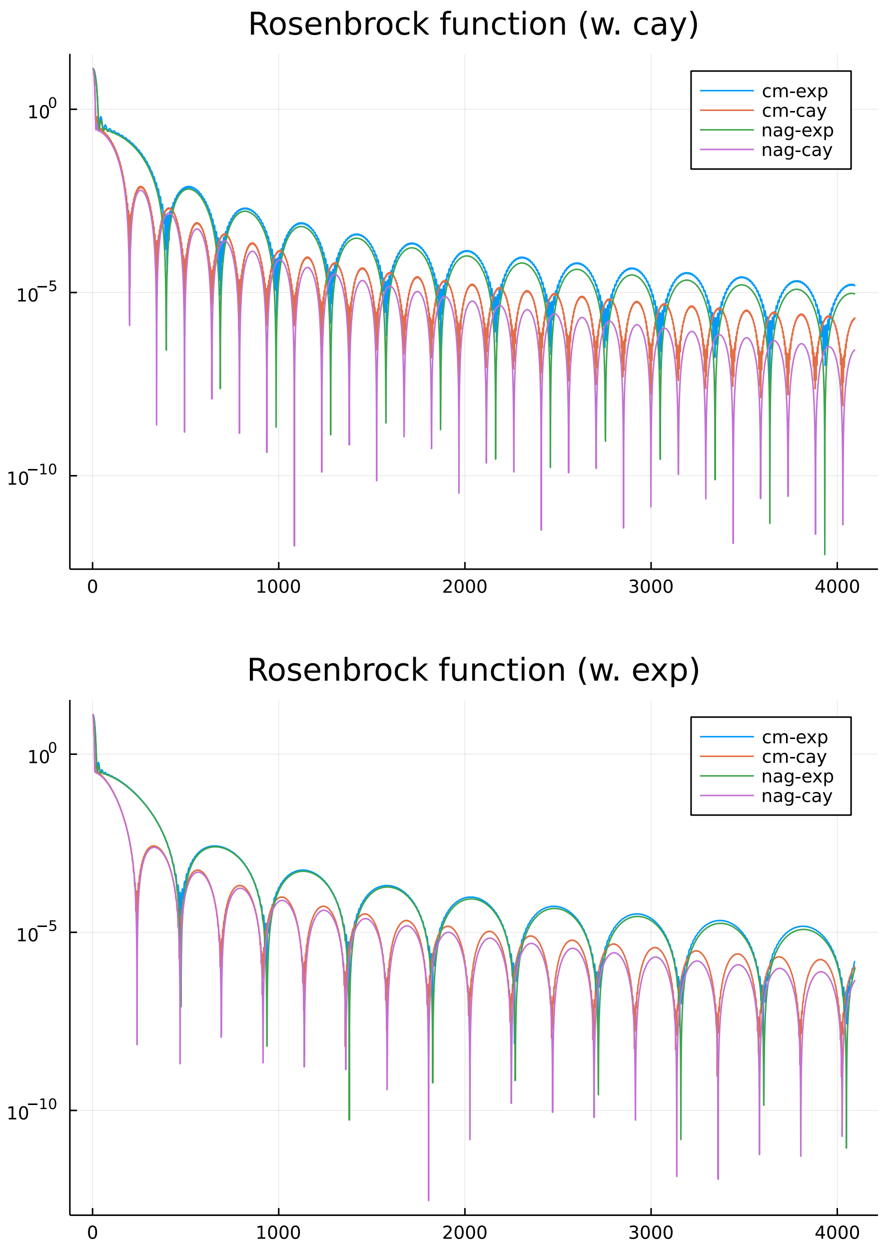


Figure 3.4: Rosenbrock function defined on $\mathrm{SO}(3)$ from two different retractions, for $R_0 = \tau(0.5, 0.5, 0.5)$ and $h = 0.01$.

action recognition as a 3D classification problem (Hu et al., 2004; Yang et al., 2019). The objective is to identify actions performed by a human subject, in our case through the information of a set of rotations that represent the relative position of the skeleton joints in a sequence of images (Huang et al., 2017). Hence, our approach is based on the representation of the human skeleton as an element of the Lie group formed by the Cartesian product of a certain number of SO(3) groups. This involves a very high dimensionality. We will present a simple deep network that can distinguish between a finite number of actions and then will test our accelerated methods in the training task.

Let us define the announced deep network. After performing a linearizing mapping layer, the output can be concatenated to a regular neural network using its Euclidean nature. The linearizing layer will consist of the use of a retraction to transport the group elements to the Lie algebra. In this case, to the Euclidean space $\mathbb{R}^3$. We won't use any additional fully-connected layer for simplicity. Let $\tau\colon U \subset \mathbb{R}^3 \to \mathrm{SO}(3)$ be a retraction on SO(3), diffeomorphism onto the image $\tau(U)$, and $\phi\colon \mathrm{SO}(3) \to \mathbb{R}$ the function $\phi(R) = \operatorname{acos}\left(\frac{1}{2}\operatorname{tr} R - \frac{1}{2}\right)$ if $\tau = \log$, and $\phi(R) = \frac{3-\operatorname{tr} R}{1+\operatorname{tr} R}$ if $\tau = \mathrm{cay}$. Note that $\phi(\tau(x)) = |x|$, for $x \in U$. For every $R, W \in \mathrm{SO}(3)^{MP}$, consider the following deep network consisting of four layers:

$$R \xrightarrow{\text{rot}} WR \xrightarrow{\text{pooling}} V \xrightarrow{\tau^{-1}} f^{\tau^{-1}}(V) \xrightarrow{\sigma} f^{\sigma} \circ f^{\tau^{-1}}(V)\,,$$

where

$$\begin{aligned} R &= (R_1^1, \ldots, R_1^P, \ldots, R_M^1, \ldots, R_M^P) \in \mathrm{SO}(3)^{MP}\,, \\ WR &= (W_1^1 R_1^1, \ldots, W_1^P R_1^P, \ldots, W_M^1 R_M^1, \ldots, W_M^P R_M^P) \in SO(3)^{MP}\,, \\ V &= (V_1, \ldots, V_M) \in \mathrm{SO}(3)^M\,, \end{aligned}$$

for $V_i := \operatorname*{argmax}_{j=1,\ldots,P} \phi(W_i^j R_i^j)$, and define

$$\begin{aligned} f^{\tau^{-1}}(V) &= (\tau^{-1}(V_1), \ldots, \tau^{-1}(V_M)) \in (\mathbb{R}^3)^M\,, \\ f^{\sigma}(x_1, \ldots, x_M) &= (\sigma(x_1), \ldots, \sigma(x_M)) \in \mathbb{R}^M\,. \end{aligned}$$

The term $\sigma\colon \mathbb{R}^3 \to \mathbb{R}$ is a sigmoid function defined by

$$x = (x_1, x_2, x_3) \mapsto \sigma(x) := \frac{2}{1 + \exp^{-(x_1^2 + x_2^2 + x_3^2)}} - 1\,.$$

Note that $\sigma(x) \in [0,1], \forall x \in \mathbb{R}^3$, and therefore, the final output may be thought of as a vector with the probabilities of observing a certain action. The subscripts denote the number of joints considered, $i = 1, \ldots, M$ (spatial part), and the superscripts the number of frames of each of the actions $j = 1, \ldots, P$ (temporal part). The components $W \in \mathrm{SO}(3)^{MP}$ denote the parameters to be optimize in the training task. Moreover, the second layer is pooling with respect to the temporal part, which is used to reduce the training computational effort. Observe that the neural network $R \in \mathrm{SO}(3)^{MP} \mapsto f^{\sigma} \circ f^{\tau^{-1}}(V)$ could be non-smooth due to precisely that layer, but because of the local nature of the methods, the experiments does not present any problem.

As we have said, for every set of frames for a certain human action, the output of this neural network is a vector whose coordinates represent the probability that the input set corresponds to a particular action. Then, if $\{(R_n, k_n)\}_{n=1,\ldots,N} \subset SO(3) \times \mathbb{Z}$ is the **training data** with k_n an integer representing that R_n corresponds to the k_n-action, our deep learning **total loss function** is

$$\begin{array}{rcl} F\colon (SO(3))^{MP} & \longrightarrow & \mathbb{R} \\ W := (W_1^1, \ldots, W_1^P, \ldots, W_M^1, \ldots, W_M^P) & \longmapsto & F(W) := \dfrac{1}{N} \displaystyle\sum_{n=1}^{N} l_{k_n} \tilde{F}(R_n, W) \end{array}$$

where

$$\tilde{F}(R_n, W) := \left(\sigma \circ \tau^{-1} \Big(\operatorname*{argmax}_{j=1,\ldots,P} \phi(W_1^j R_{n1}^j) \Big), \ldots, \sigma \circ \tau^{-1} \Big(\operatorname*{argmax}_{j=1,\ldots,P} \phi(W_M^j R_{nM}^j) \Big) \right),$$

and l_{k_n} is the so-called **loss function**. It can be selected as

$$\begin{array}{rcl} l_k\colon \mathbb{R}^M & \longrightarrow & \mathbb{R} \\ v = (v_1, \ldots, v_M) & \longmapsto & l_k(v) := \displaystyle\sum_{i \neq k} v_i^2 + (v_k - 1)^2. \end{array}$$

We suppose $N = M$, for simplicity.

The objective is to minimize the total loss function F to obtain the optimal parameters W accordingly to the loss function. Obviously, there will be a minimum due to the shape of the cost function. We will use the presented accelerated methods, thus we need to compute the left-/right-trivialized tangent of F. The following result gives the needed right-trivialized tangent of F.

Proposition 3.1. *Let $\{(R_n, k_n)\}_{n=1,\ldots,M} \subset SO(3) \times \mathbb{Z}$ be a training data. Given $W \in \mathrm{SO}(3)^{MP}$, if $x_{ni} = \tau^{-1}(V_{ni})$ and $X_{ni} = \nabla\sigma(x_{ni})$, for $i = 1, \ldots, M$, then*

$$\begin{aligned} \left(\mathcal{R}_W^* \, \mathrm{d}F(W)\right)^\vee = \frac{1}{M} \sum_{n=1}^{M} \Big(& \underbrace{0, \ldots, 0, \overbrace{a_{n1} (\mathrm{d}\tau_{\hat{x}_{n1}}^{-1})^t X_{n1}}^{j_{n1}}, 0, \ldots, 0}_{j=1,\ldots,P}, \ldots, \\ & \ldots, \underbrace{0, \ldots, 0, \overbrace{a_{nM} (\mathrm{d}\tau_{\hat{x}_{nM}}^{-1})^t X_{nM}}^{j_{nM}}, 0, \ldots, 0}_{j=1,\ldots,P} \Big), \end{aligned} \tag{3.4}$$

where $j_{ni} \in \mathbb{Z}$ are the indices corresponding to the one in $V_{ni} = W_i^{j_{ni}} R_{ni}^{j_{ni}} \in \mathrm{SO}(3)$, $i = 1, \ldots, M$, for all $n = 1, \ldots, M$.

Proof. Let $\xi = (\hat{y}_1^1, \ldots, \hat{y}_1^P, \ldots, \hat{y}_M^1, \ldots, \hat{y}_M^P) \in \mathfrak{so}(3)^{MP}$. Denote $a_n = \nabla l_{k_n}(\tilde{F}(R_n, W)) = \nabla l_{k_n}(\sigma(x_{n1}), \ldots, \sigma(x_{nM})) \in \mathbb{R}^M$ and consider $y_{ni} = y_i^{j_{ni}} \in \mathbb{R}^3$ to be the corresponding vector with the same index $j_{ni} = 1, \ldots, P$ than the one in $V_{ni} =$

$W_i^{j_{ni}} R_{ni}^{j_{ni}} \in \mathrm{SO}(3)$. Then,

$$\begin{aligned}
\langle \mathcal{R}_W^* \, \mathrm{d}F(W), \xi \rangle &= \langle \mathrm{d}F(W), \xi W \rangle \\
&= \frac{1}{M} \sum_{n=1}^{M} a_n \cdot \left(X_{ni} \cdot \left(T_{V_{ni}} \tau^{-1} (\hat{y}_{ni} V_{ni}) \right) \right)_{1 \leq i \leq M} \\
&= \frac{1}{M} \sum_{n=1}^{M} a_n \cdot \left((\mathrm{d}\tau_{\hat{x}_{ni}}^{-1})^t X_{ni} \cdot y_{ni} \right)_{1 \leq i \leq M} ,
\end{aligned}$$

where $\mathrm{d}\tau^{-1}$ refers to the right-trivialized tangent of τ^{-1}, the "$\cdot$" outside the parenthesis refers to the usual dot product in $\mathbb{R}^M$ and the inner one to the dot product in $\mathbb{R}^3$. Expression (3.4) is straightforward from this. ■

The left-trivialized tangent is obtained analogously, taking into account that it would appear an adjoint map Ad accounting for the misalignment of the term $W_i^{j_{ni}} \hat{y}_i^{j_{ni}} R_{ni}^{j_{ni}}$. For a given training data $\{(R_n, k_n)\}_{n=1,\ldots,N} \subset SO(3) \times \mathbb{Z}$, and parameters $W \in \mathrm{SO}(3)^{MP}$, note that

$$X_{ni} = \frac{4e^{-|x_{ni}|^2}}{(1+e^{-|x_{ni}|^2})^2} x_{ni} = \frac{2}{1+\cosh(|x_{ni}|^2)} x_{ni} \in \mathbb{R}^3, \quad n, i = 1, \ldots, M .$$

Therefore, in a similar fashion than the development in Subsection 2.5.1, we have

$$\begin{aligned}
(\mathrm{d}\log_{\hat{x}_{ni}})^t X_{ni} &= \frac{2}{1+\cosh(|x_{ni}|^2)} x_{ni} , \\
(\mathrm{d}\,\mathrm{cay}_{\hat{x}_{ni}}^{-1})^t X_{ni} &= \frac{1+|x_{ni}|^2}{1+\cosh(|x_{ni}|^2)} x_{ni} ,
\end{aligned}$$

for $n, i = 1, \ldots, M$.

3.2.1 Numerical experiments

To prove the improved computational efficiency of our methods with respect to the gradient descent, the one used by Huang et al. (2017), we have chosen the retraction defining F to be $\tau = \exp$. As before, an illustrating sample of numerical experiments testing CM and NAG with respect to the gradient descent are presented here. All figures have y-axis in logarithmic scale. Moreover, we consider the initial value to be $W_0 = (I, \ldots, I) \in \mathrm{SO}(3)^{MP}$, with I the identity matrix, and the training data $\{(R_n, k_n)\}$ with R_n basis rotations over a random axis and with a random angle in the interval $]0, 1[$. Therefore, the deep network total loss function F is not the same for each simulation, and neither is its minimum.

In figure 3.5 we plot the accelerated integrators for $M = 15$ training data and joints, and $P = 6$ frames. We can see that effectively CM and NAG are really efficient methods to optimize the total loss function F. Although it is not appreciated in the pictures, CM and NAG produce different sequences with a variation between them of order 10^{-4}. Moreover, after a few iterations, NAG

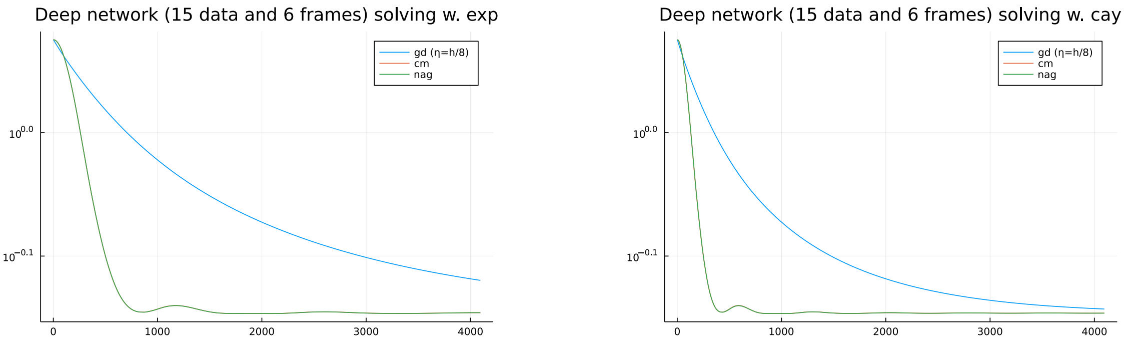


Figure 3.5: Deep network defined by the exponential map, for $h = 0.01$.

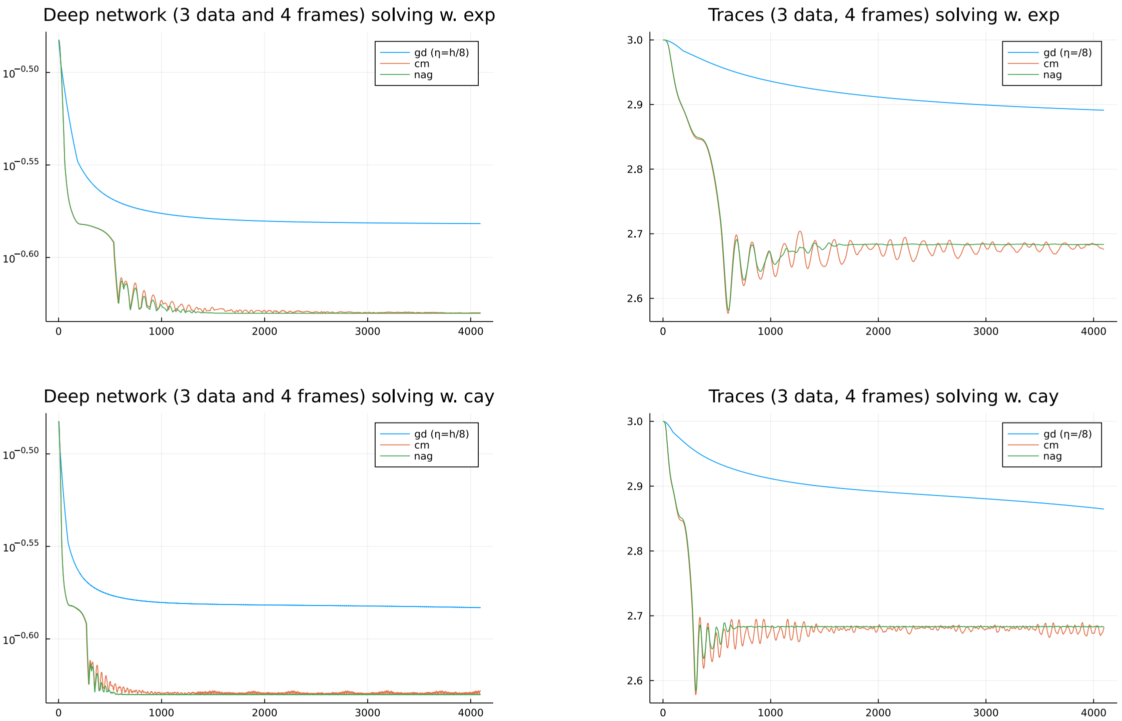
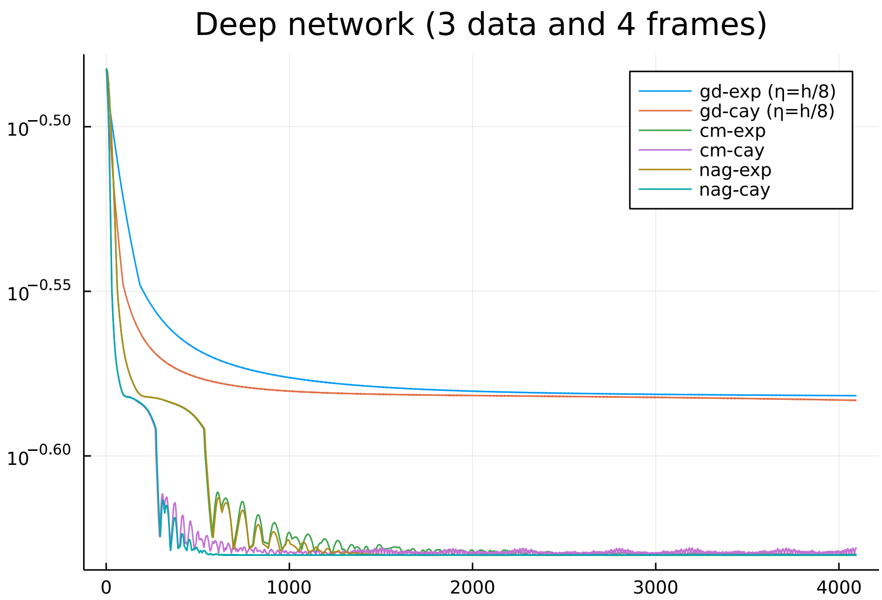


Figure 3.6: Deep network defined by the exponential map, for $h = 0.05$, and the traces of the sequence of rotations on each step.

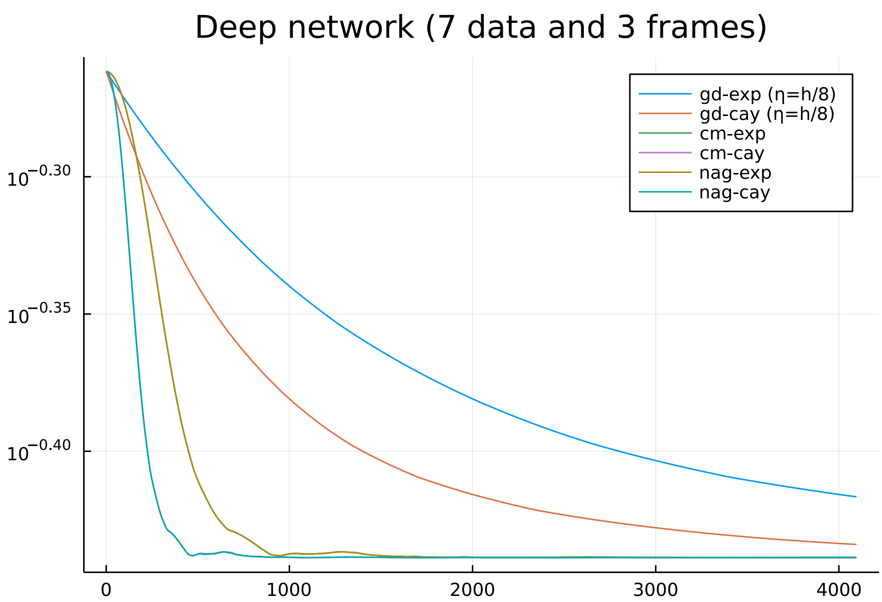
again finds lower values compared to CM. Moreover, as in the previous case, the latter produces more oscillations in the dynamical sequences. In Figure 3.6 we have plotted also the traces of the discrete trajectories to visualize this clearly.

Again, the higher efficiency of the Cayley map with respect to the exponential produces faster methods, even taking into account the extra computations for solving Equation (2.43). This can be clearly seen in Figure 3.7.

Eventually, these algorithms could be upgraded using the standard techniques in Machine Learning, such as a version of the Stochastic Gradient Descent (Hu et al., 2009; Jain et al., 2018). This would improve the acceleration in the deep network training procedure and therefore increases the relevance of these.



(a) $h = 0.05$



(b) $h = 0.025$

Figure 3.7: Deep network defined by the exponential map with two different configurations of training data and step-sizes.

Conclusion and future work

Throughout this work we have introduced the basis of discrete geometric mechanics that allow the development of a certain family of accelerated methods on Lie groups. In Chapter 1, we focused on the general case, where Lagrangian and Hamiltonian mechanics on smooth manifolds were introduced, using the so-called Legendre transformation (Section 1.2). The theory of discrete variations allowed us to obtain numerical methods from the so-called discrete Euler-Lagrange equations (Section 1.3), automatically obtaining simplecticity results (on the fibers T^*Q of the phase space $T^*Q \times \mathbb{R}$), Theorem 1.8. Special emphasis was given to the modification of the equations by the action of forces on a system, and also due to the possible time-dependence of the Lagrangian or Hamiltonian. The latter was crucial to obtain accelerated methods from the Nesterov Lagrangian. In Chapter 2, we focused on the modification of the discrete equations induced by trivializations of the tangent and cotangent bundles of a Lie group, leading to the left-/right-trivialized discrete Euler-Lagrange and Hamilton equations (Section 2.3). In fact, the general properties on smooth manifolds induced these to be variational (symplectic) integrators for the discrete mechanical systems considered. The objective of this work was satisfied in Section 2.5, where we presented a retraction-dependent family of accelerated integrators analogous to the Classical Momentum (CM) and Nesterov Accelerated Gradient (NAG) methods in the case of linear spaces. We deal with particular examples in the group of rotations SO(3), since two simple retractions are well known (the exponential and Cayley maps, see subsection 2.1.2) and also allow to solve the possible implicit reconstruction equation (2.39a) of the accelerated methods, as evidenced in subsection 2.5.1.

The performance of the proposed accelerated methods is shown in Chapter 3, including the case of Rosenbrock function (Section 3.1) and of minimizing the total lost function for the presented deep neural network for human motion analysis (Section 3.2). We saw that the proposed methods are indeed accelerating in contrast to the gradient descent, with a high efficiency. Moreover, the Cayley map on the rotations group is proved as a more powerful retraction to develop accelerated methods than the exponential map, due to its reduced complexity and even taking into account the extra computations for solving Equation (2.43).

Finally, the geometrical point of view arises as the correct one to analyze this type of methods and, in particular, the phenomenon of acceleration. This is evident in the derivation of NAG, where by adding the accurate forces to the mechanical system, we obtain an enhancement of the CM methods (see Section 2.5). Even so, it must be kept in mind that the latter are more stable and, therefore, this small improvement can be neglected in order to work with more solid integrators.

Several possible research lines are now open from this development. First of all, we have only obtained a numerical (qualitative) statement about the convergence of these accelerated methods. It would be gratifying at the scientific level to obtain that certainly these methods preserve the convergence ratio of the dynamical solutions of the continuous equations of motion. Moreover, other possible discrete systems, apart from the one corresponding to the Nesterov Lagrangian, can be considered and analyzed (Campos et al., 2021).

On the other hand, many other applications of accelerated numerical optimization appear in the general case of smooth manifolds (see for example Ma et al. (2000), Adler et al. (2002)). A natural question is how these integrators could be extended to this case. A promising approach is based on the notions of discretization map (Definition 1.15) and vector transport (Absil et al., 2008, Section 8.1). Considering a Riemannian manifold (Q, g), we can define a discrete forced Lagrangian system as

$$\begin{aligned}
L_d^k(q_0, q_1) &= a_k K_g \circ R_d^{-1}(q_0, q_1) - b_k^- \phi(q_0) - b_{k+1}^+ \phi(q_1)\,, \\
(F_d^k)^- &= -\frac{a_{k-1}}{a_k}(b_k^- + b_k^+)\, d\phi(q_0) \in T_{q_0}Q\,, \\
(F_d^k)^+ &= (b_k^- + b_k^+)\, \mathcal{T}_{R_d^{-1}(q_0,q_1)}(d\phi(q_0)) \in T_{q_1}Q\,,
\end{aligned}$$

where R_d is a discretization map of the form $R_d(q, v_q) = (q, R_q(v_q))$, that is, $R_q : T_qQ \to Q$ is a retraction map, $\mathcal{T}$ is the associated vector transport to this R (Absil et al., 2008, subsection 8.1.2), and $K_g(v_q) = \frac{1}{2} g_q(v_q, v_q)$ is the quadratic form associated to the Kinetic energy. Therefore, applying the discrete Lagrange-D'Alembert principle and reorganizing the associated forced Euler-Lagrange equations (1.20), we get

$$\begin{aligned}
D_1 L_d^k(q_k, q_{k+1}) + D_2 L_d^{k-1}(q_{k-1}, q_k) - \eta_k\, d\phi(q_k) &= \\
&= \mu_k \Big(\eta_k\, d\phi(q_k) - \eta_{k-1}\, \mathcal{T}_{R_d^{-1}(q_{k-1}, q_k)} d\phi(q_{k-1}) \Big)\,.
\end{aligned}$$

The terms $D_1 L_d^k$ and $D_2 L_d^k$ can be computed using the notion of *complete lift* (Yano and Ishihara, 1973). These constitute an analogous CM and NAG integrators (see equation (2.37)) that we will analyze and compare with the ones on Lie groups and linear spaces in the future.

Bibliography

Abraham, R. and Marsden, J. (1978). *Foundations of Mechanics*. Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass.

Absil, P., Mahony, R., and Sepulchre, R. (2008). *Optimization Algorithms on Matrix Manifolds*, volume 310. Princeton University Press.

Adler, R. L., Dedieu, J.-P., Margulies, J. Y., Martens, M., and Shub, M. (2002). Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, 22(3):359–390.

Barbero Liñán, M. and Martín de Diego, D. (2022). Retraction maps: A seed of geometric integrators. *Found Comput Math.*

Betancourt, M., Jordan, M. I., and Wilson, A. C. (2018). On symplectic optimization. *arXiv: Computation*.

Blanes, S. and Casas, F. (2016). *A concise introduction to geometric numerical integration*. Monographs and Research Notes in Mathematics. CRC Press, Boca Raton, FL.

Boothby, W. M. (2003). *An introduction to differentiable manifolds and Riemannian geometry*. Academic Press.

Bou-Rabee, N. and Marsden, J. E. (2009). Hamilton-Pontryagin integrators on Lie groups part I: Introduction and structure-preserving properties. *Found. Comput. Math.*, 9(2):197–219.

Breloff, T. (2021). Plots.jl.

Campos, C. M., Mahillo, A., and Martín De Diego, D. (2021). A Discrete Variational Derivation of Accelerated Methods in Optimization. DOI: 10.48550/arxiv.2106.02700, https://arxiv.org/abs/2106.02700v2.

Campos, C. M., Ober-Blöbaum, S., and Trélat, E. (2015). High order variational integrators in the optimal control of mechanical systems. *Discrete Contin. Dyn. Syst.*, 35(9):4193–4223.

Cappelletti-Montano, B., Nicola, A. D., and Yudin, I. (2013). A survey on cosymplectic geometry. *Rev. Math. Phys.*, 25(10):1343002, 55.

Cartan, E. (1945). *Les systèmes differentiels extérieurs et leur applications géométriques*. Actualités scientifiques et industrielles. Hermann & cie.

Cushman, R. H. and Bates, L. M. (2015). *Global Aspects of Classical Integrable Systems*. Birkhäuser Verlag.

de León, M. and Rodrigues, P. R. (1989). *Methods of differential geometry in analytical mechanics*, volume 158 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam.

Duruisseaux, V. and Leok, M. (2022). A variational formulation of accelerated optimization on riemannian manifolds. *SIAM Journal on Mathematics of Data Science*, 4(2):649–674.

Gelfand, I. M. and Fomin, S. V. (1963). *Calculus of variations*. Prentice-Hall.

Giaquinta, M. and Hildebrandt, S. (2004). Calculus of variations. *Springer-Verlag*, 310. Vol. I.

Hairer, E., Lubich, C., and Wanner, G. (2006). *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer Berlin, Heidelberg. Structure-preserving algorithms for ordinary differential equations.

Hartman, P. (2002). *Ordinary differential equations*, volume 38 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Corrected reprint of the second (1982) edition [Birkhäuser, Boston, MA; MR0658490 (83e:34002)].

Holm, D. D., Marsden, J. E., and Ratiu, T. S. (1998). The Euler–Poincaré Equations and Semidirect Products with Applications to Continuum Theories. *Adv. Math.*, 137(1):1–81.

Hu, C., Pan, W., and Kwok, J. (2009). Accelerated gradient methods for stochastic optimization and online learning. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., and Culotta, A., editors, *Adv Neural Inf Process Syst*, volume 22. Curran Associates, Inc.

Hu, W., Tan, T., Wang, L., and Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34(3):334–352.

Huang, Z., Wan, C., Probst, T., and Van Gool, L. (2017). Deep Learning on Lie Groups for Skeleton-Based Action Recognition. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1243–1252.

Iserles, A., Munthe-Kaas, H. Z., Nørsett, S. P., and Zanna, A. (2000). Lie-group methods. *Acta Numer.*, 9:215–365.

Jain, P., Kakade, S. M., Kidambi, R., Netrapalli, P., and Sidford, A. (2018). Accelerating stochastic gradient descent for least squares regression. In Bubeck, S., Perchet, V., and Rigollet, P., editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 545–604. PMLR.

Kirillov, Jr, A. (2008). *An Introduction to Lie Groups and Lie Algebras*. Cambridge Studies in Advanced Mathematics(CSAM). Cambridge University Press (CUP).

Lang, S. (1995). *Differential and Riemannian Manifolds*. Graduate Texts in Mathematics. Springer New York.

Lee, T., Tao, M., and Leok, M. (2021). Variational symplectic accelerated optimization on lie groups. *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 233–240.

Libermann, P. and Marle, C. M. (1987). *Symplectic geometry and analytical mechanics*, volume 35 of *Mathematics and its Applications*. D. Reidel Publishing Co., Dordrecht. Translated from the French by Bertram Eugene Schwarzbach.

Lu, M. and Li, F. (2020). Survey on Lie group machine learning. *Big Data Mining and Analytics*, 3(4):235–258.

Ma, Y., Kosecka, J., and Sastry, S. S. (2000). Linear differential algorithm for motion recovery: A geometric approach. *Int. J. Comput. Vis.*, 36:71–89.

Marrero, J. C., de Diego, D. M., and Martínez, E. (2021). Local convexity for second order differential equations on a Lie algebroid. *J. Geom. Mech.*, 13(3):477–499.

Marrero, J. C., Martín de Diego, D., and Martínez, E. (2015). The local description of discrete mechanics. In *Geometry, mechanics, and dynamics*, volume 73 of *Fields Inst. Commun.*, pages 285–317. Springer, New York.

Marsden, J. and Ratiu, T. (1999). *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. Texts in Applied Mathematics. Springer New York.

Marsden, J. E. and West, M. (2001). Discrete mechanics and variational integrators. *Acta Numer.*, 10:357–514.

Martín de Diego, D. and Sato Martín de Almagro, R. (2018). Variational order for forced Lagrangian systems. *Nonlinearity*, 31(8):3814–3846.

Martín de Diego, D. (2018). Lie-Poisson integrators. *Rev. de la Academia Canaria de Ciencias*, XXX:9–30.

Nemirovskii, A. S. and Yudin, D. (1983). *Problem complexity and method efficiency in optimization.* Wiley-Interscience series in discrete mathematics. Wiley.

Nesterov, Y. E. (1983). A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR*, 269(3):543–547.

Patrick, G. W. and Cuell, C. (2009). Error analysis of variational integrators of unconstrained Lagrangian systems. *Numer. Math.*, 113(2):243–264.

Polyak, B. (1964). Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.*, 4(5):1–17.

Rosenbrock, H. H. (1960). An Automatic Method for Finding the Greatest or Least Value of a Function. *The Computer Journal*, 3(3):175–184.

Sanz-Serna, J. M. and Calvo, M. P. (1994). *Numerical Hamiltonian problems*, volume 7 of *Appl. Math. Comput.* Chapman & Hall, London.

Sato Martín de Almagro, R. T. (2019). *Discrete mechanics for forced and constrained systems*. PhD thesis, Universidad Complutense de Madrid.

Sebe, N., Lew, M., and Huang, T. S., editors (2005). *Computer Vision in Human-Computer Interaction*, volume 3766. Springer Berlin Heidelberg.

Serre, J. P. (1992). Lie Algebras and Lie Groups. *Springer-Verlag*. 1964 Lectures given at Harvard University.

Su, W., Boyd, S., and Candès, E. J. (2016). A differential equation for modeling Nesterov's Accelerated Gradient method: theory and insights. *Journal of Machine Learning Research(JMLR)*, 17(153):1–43.

Wibisono, A., Wilson, A. C., and Jordan, M. I. (2016). A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 113(47):E7351–E7358.

Yang, H., Yuan, C., Li, B., Du, Y., Xing, J., Hu, W., and Maybank, S. J. (2019). Asymmetric 3d convolutional neural networks for action recognition. *Pattern Recognition*, 85:1–12.

Yano, K. and Ishihara, S. (1973). *Tangent and cotangent bundles: differential geometry*. New York (N.Y.) : Dekker.

Appendix A

Useful formulae in SO(3)

Given $x, y \in \mathbb{R}^3$, the following properties are satisfied (Iserles et al., 2000)

$$(x \otimes x)(y) = (x \cdot y)x \tag{A.1a}$$
$$x \otimes x = |x|^2 I + \hat{x}^2 \tag{A.1b}$$
$$\hat{x}y = x \times y \tag{A.1c}$$
$$\hat{x}\hat{y} = y \otimes x - (x \cdot y)I \tag{A.1d}$$
$$[\hat{x}, \hat{y}] = \hat{x}\hat{y} - \hat{y}\hat{x} = \widehat{x \times y} \tag{A.1e}$$
$$\hat{x}\hat{y}\hat{x} = -(x \cdot y)\hat{x} \tag{A.1f}$$
$$\hat{x}^2\hat{y} + \hat{y}\hat{x}^2 = -|x|^2\hat{y} - (x \cdot y)\hat{x} \tag{A.1g}$$

where $(x \otimes y)(z) := (y \cdot z)x$, for all $x, y, z \in \mathbb{R}^3$. These equations can be obtained from the expression of every skew-symmetric matrix,

$$\hat{x} = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix} \in \mathfrak{so}(3)\,,$$

for any $x \in \mathbb{R}^3$. That is, computing the coordinates of both sides of the expressions, they arise to be the same.