



**HAL**  
open science

# Legendre Transform and Applications to Finite and Infinite Optimization

Cristopher Hermosilla

► **To cite this version:**

Cristopher Hermosilla. Legendre Transform and Applications to Finite and Infinite Optimization. Set-Valued and Variational Analysis, 2016, 10.1007/s11228-016-0368-5 . hal-01055917v1

**HAL Id: hal-01055917**

**<https://inria.hal.science/hal-01055917v1>**

Submitted on 14 Aug 2014 (v1), last revised 14 Apr 2016 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SQUARED HESSIAN RIEMANNIAN METRICS ON CONVEX SETS AND APPLICATIONS TO FINITE AND INFINITE OPTIMIZATION.

CRISTOPHER HERMOSILLA

ABSTRACT. We investigate convex constrained nonlinear optimization problems and optimal control with convex state constraints. For this purpose we endow the interior of constraints set with the structure of Riemannian manifold. In particular, we consider a class of Riemannian metric induced by the squared Hessian of a Legendre functions. We describe in details the geodesic curves on this manifolds and we propose a gradient-like algorithm for constrained optimization based on linear search along geodesics. We also use the Legendre change of coordinates to study the Value Function of a Mayer problem with state constraints. We provide a characterization of the Value Function for this problem as the unique viscosity solution of the Hamilton-Jacobi-Bellman equation.

## 1. INTRODUCTION.

It is natural in optimization models to include constraints to reflect the real-world restrictions. These may be due to the physical nature of the problem or to policies that some agent may set. In general, constrained models yield to more than technical difficulties when comparing with unconstrained ones, reason for which it has attracted the attention of many researchers in mathematical programming as well as in optimal control.

In this paper we study the case of convex constraints through the so-called *Legendre transform*. We are particularly interested in convex sets whose interior can be endowed with the structure of Riemannian manifold by mean of a class of penalization functions known as the *Legendre functions*. The principal advantage of using this transformation, as already remarked in [2], is that the under suitable assumptions, it provides an isometry between the interior of the constraints and  $\mathbb{R}^n$ . This fact is essential for our study because it allows us to transform the constrained problem of concern into an

---

2010 *Mathematics Subject Classification*. 53C25, 65K05, 90C51, 93C15, 49L25.

*Key words and phrases*. Riemannian manifold, Riemannian metrics, Convex functions of Legendre type, Legendre transform coordinates, geodesic search methods, Optimal control of differential inclusions, Mayer problem, Value Function, State constraints.

This research was partially financed by the Basal Project CMM, Universidad de Chile and by the European Union under the 7th Framework Programme **FP7-PEOPLE-2010-ITN** Grant agreement number 264735-SADCO .

unconstrained one. In particular, we present two applications to optimization, one to finite dimensional mathematical programs and another to an infinite dimensional problem (an optimal control problem of Mayer type).

The specific Riemannian metric we analyze is called the *squared Hessian Riemannian metric* and it allows us to characterize some of the principal elements of interest associated with a manifold in an explicit way. Namely, we can provide an expression for the geodesics and for the Riemannian distance, and a simple criterion for completeness as well.

The application we exhibit for the finite dimensional problem is based in the expression obtained for the geodesics, we use this to propose a geodesic search strategy for solving finite dimensional mathematical programs with convex constraints. This method is based on a dual lecture given to the original mathematical program in terms of the Legendre transform.

To the best of our knowledge, the algorithm we propose has not been directly studied before. However, there are in the literature some works of similar nature when seeing it as a penalization method; see for instance [4, 26]. Nevertheless, the approach using squared Hessian Riemannian metric, specially the choice of the descent directions, seems to be quite new.

On the other hand, we also use the Legendre transform to study the Value Function of a Mayer problem with convex state constraints. In the case with unrestricted state space, it is well-known that the Value Function of an optimal control problem is the unique solution in the viscosity sense of the so-called Hamilton-Jacobi-Bellman equation; we refer to [5, 32] for more details. However, in presence of state constraints the situation is different and without additional compatibility assumptions, this is no longer true. In the literature it is usually assumed that the control systems satisfy a qualification condition called the *inward pointing condition* (IPC); see for example [29, 15, 18] among many others. This condition basically says that at each point of the boundary of  $\Omega$  there exists a controlled vector field pointing into  $\Omega$ .

For the case when the Value Function is merely lower semicontinuous another compatibility assumption of similar nature has been taken into account, this is known as the *outward pointing condition* (OPC). This condition states essentially the same as the IPC but for the backward dynamics; see for instance [19, 18]. In this paper we consider control systems for which the boundary of the state constraints is absorbing and so neither IPC nor the OPC are satisfied; we give more details in Section 6. We show that under reasonable hypothesis the Value Function can be characterized as the unique viscosity solution of the Hamilton-Jacobi-Bellman equation on  $\Omega$ .

In the literature, alternative ways to deal with the Value Function without imposing an IPC or an OPC has also been studied; see for example [8, 20].

We mentioned that the interest in optimization problems on manifolds has been increasing in the last years; we make reference to [28, 1] for nonlinear optimization, to [26, 30] for convex optimization and to [13, 11] for optimal control. This work contributes also in this direction by providing an explicit

study for the manifolds that are convex open set on the ambient space, which can be endowed with the squared Hessian Riemannian manifold.

**1.1. Notation.** Throughout this paper,  $\mathbb{R}$  denotes the sets of real numbers,  $|\cdot|$  is the Euclidean norm and  $\langle \cdot, \cdot \rangle$  is the Euclidean inner product on  $\mathbb{R}^n$ ,  $\mathbb{B}$  the unit open ball  $\{x \in \mathbb{R}^n : |x| < 1\}$  and  $\mathbb{B}(x, r) = x + r\mathbb{B}$ . For a given set  $S$ , we denote by  $\text{int}(S)$ ,  $\bar{S}$  and  $\text{bdry}(S)$  its interior, closure, and boundary, respectively. The effective domain of a function  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is defined by  $\text{dom } \varphi = \{x \in \mathbb{R}^n \mid \varphi(x) \in \mathbb{R}\}$ . If  $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$  is a set-valued map, then  $\text{dom } \Gamma$  is the set of points for which  $\Gamma(x) \neq \emptyset$ .

For any open set  $\mathcal{O} \subseteq \mathbb{R}^n$  and any  $p \in \mathbb{N}$ ,  $\mathcal{C}^p(\mathcal{O})$  stands for the set of all the  $p$ -times continuously differentiable real-valued functions on  $\mathcal{O}$ . If  $f \in \mathcal{C}^p(\mathcal{O})$ ,  $\nabla f(x)$  stands for the gradient of  $f$  at  $x$  and if  $p \geq 2$ , the Hessian matrix of  $g$  at  $x$  is denoted by  $\nabla^2 g(x)$ .

## 2. PRELIMINARIES.

**2.1. Some notions of Convex Analysis.** In this part we review some concepts of Convex Analysis; for a more extensive exposition on the topic we refer to the classical monograph [27] or to a more recent textbook [10]. We denote by  $\Gamma_0(\mathbb{R}^n)$  the class of all convex and lower semicontinuous functions defined on  $\mathbb{R}^n$  with values in  $\mathbb{R} \cup \{+\infty\}$  whose effective domain is nonempty. Let  $g \in \Gamma_0(\mathbb{R}^n)$ , we denote by  $g^*$  its *Fenchel-Legendre conjugate*

$$g^*(y) = \sup\{\langle x, y \rangle - g(x) \mid x \in \text{dom } g\},$$

and by  $\partial g$  its *subdifferential*

$$\partial g(x) = \{y \in \mathbb{R}^n \mid \forall z \in \mathbb{R}^n, g(x) + \langle y, z - x \rangle \leq g(z)\}.$$

Accordingly to [27, Section 26],  $g \in \Gamma_0(\mathbb{R}^n)$  is called *essentially smooth* if

- $\text{int}(\text{dom } g) \neq \emptyset$  and  $g$  is differentiable on  $\text{int}(\text{dom } g)$ .
- $|\nabla g(x_k)| \rightarrow +\infty$  for every sequence  $\{x_k\} \subseteq \text{int}(\text{dom } g)$  such that  $x_k \rightarrow \bar{x}$  for some  $\bar{x} \in \text{bdry}(\text{dom } g)$ .

This property has a dual interpretation in terms of the Legendre-Fenchel conjugate. Indeed,  $g$  is essentially smooth if and only if  $g^*$  is strictly convex on every convex subset of  $\text{dom } \partial g$ ; see [27, Theorem 26.3]. This last property is known as *essential strict convexity*. Hence, we have that

$g$  is essentially smooth if and only if  $g^*$  is essentially strictly convex.

This relation motivates the following definition.

**Definition 2.1** ([27, Section 26], [10, Definition 7.1.1]). *A function  $g \in \Gamma_0(\mathbb{R}^n)$  is called convex function of Legendre type or simply Legendre function provided it is essentially smooth and essentially strictly convex.*

**Remark 2.1.** *Some well-known examples of Legendre functions on  $\mathbb{R}$  for which the interior of their domain is  $(0, +\infty)$  are the log-barrier  $g_{\log}$  and the Boltzmann-Shannon entropy  $g_{ent}$  which are defined as follows:*

$$g_{\log}(x) := \begin{cases} -\log(x) & x > 0 \\ +\infty & x \leq 0 \end{cases} \quad \text{and} \quad g_{ent}(x) := \begin{cases} x \log(x) - x & x \geq 0 \\ +\infty & x < 0. \end{cases}$$

Consequently,  $g$  is a Legendre function if and only if  $g^*$  it is as well. In this case, by [27, Theorem 26.5],  $\nabla g : \text{int}(\text{dom } g) \rightarrow \text{int}(\text{dom } g^*)$  is a bijection known as the *Legendre transform*. Furthermore, we have

$$(1) \quad (\nabla g)^{-1}(y) = \nabla g^*(y), \quad \forall y \in \text{int}(\text{dom } g^*).$$

This last relation implies that whenever the Hessian matrix of  $g$  and  $g^*$  are well-defined and invertible, they satisfy the following

$$[\nabla^2 g(x)]^{-1} = \nabla^2 g^*(y), \quad \text{with } y = \nabla g(x).$$

**2.2. Riemannian Geometry framework.** Now, in order to make the exposition clearer for the readers who are not familiar with the Riemannian Geometry Theory, we make a short review on some of the definitions we are going to use in this paper; we refer to [16, 22] for further details.

Suppose that  $M$  is a differentiable manifold, we denote by  $TM$  the *tangent bundle* to  $M$  and by  $T_x M$  the *tangent space* to  $M$  at  $x$ . A chart on  $M$  is a pair  $(\mathcal{O}, \phi)$ , where  $\mathcal{O} \subseteq M$  is an open subset and  $\phi : \mathcal{O} \rightarrow \phi(\mathcal{O})$  is an homeomorphism from  $\mathcal{O}$  into an open  $\phi(\mathcal{O}) \subseteq \mathbb{R}^n$ .

A map  $\Psi : M \rightarrow N$  from  $M$  into another differentiable manifold  $N$  is said *differentiable* provided for every  $x \in M$  there exist a chart  $(\mathcal{O}, \phi)$  around  $x$  and a chart  $(\mathcal{U}, \varphi)$  around  $\Psi(x)$  such that  $\Psi(\mathcal{O}) \subseteq \mathcal{U}$  and the map  $\varphi \circ \Psi \circ \phi^{-1}$  is differentiable from  $\phi(\mathcal{O})$  into  $\varphi(\mathcal{U})$ . In this case, its differential is

$$d_x \Psi : T_x M \rightarrow T_{\Psi(x)} N.$$

A differentiable manifold  $M$  is a *Riemannian manifold* provided there exists a *Riemannian metric* on  $M$ , that is, a map that associates to each  $x \in M$  an inner dot  $(\cdot, \cdot)_x$  defined on the tangent space  $T_x M$ , which in addition is differentiable with respect to  $x$  in the following sense:

$$\forall \Psi_1, \Psi_2 : M \rightarrow TM \text{ differentiable, } x \mapsto (\Psi_1(x), \Psi_2(x))_x \text{ is differentiable.}$$

**Remark 2.2.** *Suppose that  $M$  is an open set of  $\mathbb{R}^n$ , then we can identify the tangent space to  $M$  with  $\mathbb{R}^n$ . Moreover, consider a differentiable mapping  $x \mapsto A(x)$  that associates to any  $x \in M$  a  $n \times n$  real symmetric and positive-definite matrix  $A(x)$ . The following defines a Riemannian metric on  $M$ :*

$$(u, v)_x^A := \langle A(x)u, v \rangle, \quad \forall u, v \in \mathbb{R}^n.$$

The Riemannian structure allows to identify  $T_x M$  with its dual, that is, the cotangent space  $T_x M^*$ , and thus to define a notion of gradient vector. Indeed, given  $f : M \rightarrow \mathbb{R}$  differentiable, the gradient of  $f$  is denoted by  $\nabla_M f$  and is uniquely determined by the following two conditions:

- Tangency: for all  $x \in M$ ,  $\nabla_M f(x) \in T_x M$ .
- Duality: for all  $x \in M$ ,  $v \in T_x M$ ,  $d_x f(v) = (\nabla_M f(x), v)_x$ .

If we are in the situation described in Remark 2.2, the Riemannian gradient is denoted by  $\nabla_A f$  to emphasis the dependence on matrix-valued map  $A(\cdot)$ .

On Riemannian manifold it is also possible to compute the length of a piecewise differentiable curve  $\gamma : [0, 1] \rightarrow M$ . Indeed, it is given by

$$\ell(\gamma) = \sum_{i=0}^{n-1} \int_{a_i}^{a_{i+1}} \sqrt{(\dot{\gamma}(t), \dot{\gamma}(t))_{\gamma(t)}} dt,$$

where  $0 = a_0 < \dots < a_k = 1$  is a finite partition of  $[0, 1]$  such that  $\gamma$  is differentiable on each open interval  $(a_i, a_{i+1})$ .

Furthermore, we can define a distance on  $M$  which endows it with the structure of metric space. This distance is defined as follows:  $\forall x, y \in M$

$$\text{dist}_M(x, y) = \inf\{\ell(\gamma) \mid \gamma \text{ piecewise differentiable, } \gamma(0) = x \text{ and } \gamma(1) = y\}.$$

If  $M$  and  $N$  are two Riemannian manifolds, a diffeomorphism  $\Psi : M \rightarrow N$  is called an *isometry between  $M$  and  $N$*  provided

$$(2) \quad (u, v)_x = (d_x \Psi(u), d_x \Psi(v))_{\Psi(x)}, \quad \forall x \in M, \forall u, v \in T_x M.$$

On the other hand, roughly speaking, a *geodesic* on a differentiable manifold is a curve whose "acceleration" is zero. To define what "acceleration" means it is required to introduced the so-called *affine connections*, which are basically operators that allow to define directional derivatives of vector fields on manifolds. In the sequel we only consider geodesics on *flat Riemannian manifolds*, i.e., those one that are locally isometric to  $\mathbb{R}^n$  endowed with the Euclidean metric (each open subset of  $M$  is isometric to an open subset of  $\mathbb{R}^n$ ). For a definition in a more general context we refer to [22, Chapter 5].

**Definition 2.2.** *Let  $M$  be a flat Riemannian manifold. A curve  $\gamma : (a, b) \rightarrow M$  with  $a < 0 < b$ , is a geodesic with initial point  $x$  and initial velocity  $v$  provided*

$$\gamma(t) = \Psi^{-1}(\Psi(x) + td_x \Psi(v)), \quad \forall t \in (a, b),$$

where  $\Psi$  is the local isometry between  $M$  and  $\mathbb{R}^n$  around  $x$ .

Note that the definition of geodesic is only local in time and may not have sense for some particular times. When this situation does not occur, that is, any geodesic can be defined for all times, it yields to a notion of completeness on the manifold. More precisely, a Riemannian manifold  $M$  is called *geodesically complete* if the maximal interval of time on which any arbitrary geodesic can be defined is the whole space  $\mathbb{R}$ .

We finish this section by introducing some generalized notions of convexity on Riemannian manifolds. First we introduce an extension of convex set. A subset  $C \subseteq M$  is called *totally convex* if for every  $x, y \in C$ , the geodesic  $\gamma : [0, 1] \rightarrow M$  for which  $\gamma(0) = x$  and  $\gamma(1) = y$ , must also satisfy

$$\gamma(t) \in C, \quad \forall t \in [0, 1].$$

**Definition 2.3.** Let  $M$  be a Riemannian manifold and  $C \subseteq M$  a totally convex set. A function  $f : C \rightarrow \mathbb{R}$  is called convex on  $M$  if for any  $x, y \in C$

$$f(\gamma(t)) \leq (1-t)f(x) + tf(y), \quad \forall t \in [0, 1],$$

where  $\gamma : [0, 1] \rightarrow M$  is the geodesic that satisfies  $\gamma(0) = x$  and  $\gamma(1) = y$ .

### 3. SQUARED HESSIAN METRICS INDUCED BY LEGENDRE FUNCTIONS.

Now we are in position to introduce the main object of study of this paper and its principal properties. Hereinafter, the symbol  $\Omega$  is reserved to denote an open nonempty convex set of  $\mathbb{R}^n$  and  $g$  is a function in the class  $\Gamma_0(\mathbb{R}^n)$ .

**3.1. Basic definitions.** We begin this subsection with a definition that will play an important role throughout this manuscript.

**Definition 3.1.** Let  $g \in \Gamma_0(\mathbb{R}^n)$  and let  $\Omega \subseteq \mathbb{R}^n$  be a nonempty open convex set. We say that the pair  $(\Omega, g)$  is Legendre zone consistent if  $g$  is a Legendre function with  $\text{int}(\text{dom } g) = \Omega$ .

**Remark 3.1.** In general,  $\Omega$  is given and we look for a Legendre function such that the pair  $(\Omega, g)$  is Legendre zone consistent. The existence and construction of a Legendre function for an arbitrary open convex set is not a trivial task, however, as shown in [9], for many important cases it is possible to give a positive answer; see also [10, Chapter 7.4].

The main purpose of introducing the concept of Legendre zone consistent pairs lies in the possibility of defining a Riemannian metric on a given open convex set. Let  $(\Omega, g)$  be a Legendre zone consistent pair with  $g \in \mathcal{C}^2(\Omega)$ . For any  $x \in \Omega$  consider the bilinear mapping defined on  $\mathbb{R}^n \times \mathbb{R}^n$

$$(3) \quad \rho_x(u, v) := \langle \nabla^2 g(x)u, \nabla^2 g(x)v \rangle, \quad \forall u, v \in \mathbb{R}^n.$$

In order to endow  $\Omega$  with the structure of Riemannian manifold it is necessary to impose some conditions over the Legendre function  $g$ .

**Lemma 3.1.** Let  $(\Omega, g)$  be a Legendre zone consistent pair with  $g \in \mathcal{C}^2(\Omega)$ . Then, the family  $\{\rho_x\}_{x \in \Omega}$  is a Riemannian metric on  $\Omega$  provided

$$(4) \quad g \in \mathcal{C}^3(\Omega) \text{ and } \nabla^2 g(x) \text{ is positive-definite } \forall x \in \Omega.$$

*Proof.* By (4),  $[\nabla^2 g(x)]^2$  is symmetric and positive-definite for any  $x \in \Omega$ . So,  $\rho_x$  is an inner product on  $\mathbb{R}^n$ . Besides, since  $\Omega$  is an open set, the tangent space to  $\Omega$  at  $x$  can be identified with  $\mathbb{R}^n$  for any  $x \in \Omega$ . Furthermore, given that  $g$  is thrice continuously differentiable on  $\Omega$ , for any pair of differentiable vector fields  $\Psi_1, \Psi_2 : \Omega \rightarrow \mathbb{R}^n$ , the map  $x \mapsto \rho_x(\Psi_1(x), \Psi_2(x))$  is differentiable. Hence,  $\{\rho_x\}_{x \in \Omega}$  is a Riemannian metric on  $\Omega$ .  $\square$

If no confusion can arise, we use the notation  $G(x) := \nabla^2 g(x)$ , so

$$G_{ij}(x) = \frac{\partial^2 g}{\partial x_i \partial x_j}(x), \quad i, j = 1, \dots, n.$$

**Definition 3.2.** Let  $(\Omega, g)$  be a Legendre zone consistent pair that satisfies (4). The family of inner dots given by (3) is called the squared Hessian Riemannian metric on  $\Omega$  induced by the Legendre function  $g$ , and it is denoted by  $(\cdot, \cdot)_x^{G^2}$ , that is,  $\forall x \in \Omega$ ,

$$(u, v)_x^{G^2} := \langle G(x)^2 u, v \rangle = \langle G(x)u, G(x)v \rangle, \quad \forall u, v \in \mathbb{R}^n.$$

In this case, we say that  $(\Omega, G^2)$  is a Squared Hessian Riemannian manifold.

**Remark 3.2.** Consider  $\Omega = \mathbb{B}$  in  $\mathbb{R}^n$ , then some suitable choices of Legendre functions that make  $(\Omega, G^2)$  a squared Hessian Riemannian manifolds are

$$g_1(x) = \begin{cases} -\log(1 - |x|^2) & |x| < 1 \\ +\infty & |x| \leq 1 \end{cases} \quad \text{and} \quad g_2(x) = \begin{cases} -\log(\cos(\frac{\pi}{2}|x|)) & |x| < 1 \\ +\infty & |x| \leq 1. \end{cases}$$

**3.2. Properties.** We turn our attention into the properties of squared Hessian Riemannian manifolds. The following lemma is the basic result we use to derive properties for this class of manifolds.

**Lemma 3.2.** Let  $(\Omega, g)$  be a Legendre zone consistent pair that satisfies (4) and set

$$(5) \quad \Omega^* := \text{int}(\text{dom } g^*) = \nabla g(\Omega).$$

Then  $\nabla g : \Omega \rightarrow \Omega^*$  is an isometry between the Riemannian manifold  $(\Omega, G^2)$  induced by  $g$  and  $\Omega^*$  endowed with the Euclidean metric. In particular, if  $\gamma : [a, b] \rightarrow \Omega$  is a piecewise differentiable curve, then  $\ell(\gamma) = \ell(\tilde{\gamma})$ , where  $\tilde{\gamma} : [a, b] \rightarrow \Omega^*$  is given by  $\tilde{\gamma}(t) = \nabla g(\gamma(t))$  for any  $t \in [a, b]$ .

*Proof.* Since  $g$  is a Legendre function with nonsingular Hessian, it follows from the inverse function theorem that  $\nabla g$  is a diffeomorphism. Furthermore, let  $u, v \in \mathbb{R}^n$  and  $x \in \Omega$ , hence

$$(u, v)_x^{G^2} = \langle G(x)^2 u, v \rangle = \langle \nabla^2 g(x)u, \nabla^2 g(x)v \rangle = \langle d_X(\nabla g)(u), d_X(\nabla g)(v) \rangle.$$

Therefore, (2) is verified and so  $\nabla g$  is an isometry between  $(\Omega, G^2)$  and  $\Omega^*$  endowed with the Euclidean metric of  $\mathbb{R}^n$ . The equality for the length of the curves is easily verified from its definition.  $\square$

In the light of the previous lemma, we can completely characterize the geodesics on a squared Hessian Riemannian manifold.

**Proposition 3.1 (Geodesics).** Let  $(\Omega, g)$  be a Legendre zone consistent pair that satisfies (4). The geodesic on  $(\Omega, G^2)$  that satisfies  $X(0) = x_0 \in \Omega$  and  $\dot{X}(0) = v_0 \in \mathbb{R}^n$  is given by

$$(6) \quad X(t) = \nabla g^*(\nabla g(x_0) + t\nabla^2 g(x_0)v_0), \quad \forall t \in (T_{min}, T_{max}).$$

*Proof.* Recall that by the Fundamental Lemma of Riemannian Geometry [22, Theorem 5.4] the Levi-Civita is the unique linear connection on  $(\Omega, G^2)$ , so by the Existence and Uniqueness of Geodesics Theorem [22, Theorem 4.10] such geodesic is uniquely determined on an interval of time that contains zero. Let  $t \mapsto X(t)$  be the geodesic which is defined on maximal



interval of time  $(T_{min}, T_{max})$ . By virtue of Lemma 3.2 and [22, Proposition 5.6], we have that  $Y(t) = \nabla g(X(t))$  is a geodesic in  $\Omega^*$  endowed with the Euclidean metric, that is,  $\exists y_0, w_0 \in \mathbb{R}^n$ , such that  $Y(t) = y_0 + tw_0$ . Therefore,  $y_0 = \nabla g(x_0)$  and since  $\dot{Y}(t) = \nabla^2 g(X(t))\dot{X}(t)$ , we deduce that  $w_0 = \dot{Y}(0) = \nabla^2 g(x_0)v_0$ . Thus by (1) we get (6).  $\square$

**Remark 3.3.** Consider  $\Omega = \mathbb{B}$  in  $\mathbb{R}^n$  and take  $g = g_2$  where  $g_2$  is given by Remark 3.2. Hence using the basic calculus rules for the Legendre-Fenchel transform we get that

$$g^*(y) = \frac{2}{\pi}|y| \arctan\left(\frac{2}{\pi}|y|\right) - \frac{1}{2} \log\left(1 + \frac{4}{\pi^2}|y|^2\right), \quad \forall y \in \mathbb{R}^N.$$

Furthermore,  $\nabla g^*(y) = \frac{2 \arctan(\frac{2}{\pi}|y|)}{\pi|y|}y$  for every  $y \in \mathbb{R}^n \setminus \{0\}$ . Thus, by Proposition 3.1 the geodesic of  $(\Omega, G^2)$  starting from  $x_0 \in \Omega$  with initial velocity  $\nabla^2 g(x_0)^{-1}d_0 \in \mathbb{R}^n$  is given by

$$X(t) = \frac{2 \arctan\left(\frac{2}{\pi} \left| \frac{\pi \tan\left(\frac{\pi}{2}|x_0|\right)}{2|x_0|}x_0 + td_0 \right| \right)}{\pi \left| \frac{\pi \tan\left(\frac{\pi}{2}|x_0|\right)}{2|x_0|}x_0 + td_0 \right|} \left( \frac{\pi \tan\left(\frac{\pi}{2}|x_0|\right)}{2|x_0|}x_0 + td_0 \right), \quad \forall t \in \mathbb{R}.$$

The geodesics of a squared Hessian Riemannian manifold are intrinsically related to constrained optimization problems with linear objective functions. Indeed, these arcs can be seen as central path trajectories of a penalized problem, where the penalization term is the *Bregman pseudo-distance* associated with a Legendre function  $g$

$$D_g(x, y) := g(x) - g(y) - \langle \nabla g(y), x - y \rangle, \quad \forall x, y \in \Omega,$$

and where the time variable  $t$  acts like a penalization parameter. For more details about central path methods we refer the reader to [31, Chapter 25].

**Remark 3.4.** The fact that for a linear objective function the Hessian Riemannian steepest descent trajectory (8) can be viewed as a sort of central path (cf. (7)) has been already remarked in the literature; see for instance [6, 17, 23] for the log-barrier and [21] for more general barrier functions. For a complete study of such trajectories including the case of a nonlinear objective function  $f$ , we make reference to [2].

**Proposition 3.2** (Central Path). *Let  $(\Omega, g)$  be a Legendre zone consistent pair that satisfies (4). Let  $t \mapsto X(t)$  be the geodesic on  $(\Omega, G^2)$  that satisfies  $X(0) = x_0 \in \Omega$  and  $\dot{X}(0) = -\nabla^2 g(x_0)^{-1}d_0$  for some  $d_0 \in \mathbb{R}^n$ . Then*

$$(7) \quad X(t) = \arg \min \left\{ \langle d_0, x \rangle + \frac{1}{t} D_g(x, x_0) \right\}, \quad \forall t \in (0, T_{max}).$$

Moreover,  $t \mapsto X(t)$  is the steepest descent trajectory for the linear function  $f(x) = \langle d_0, x \rangle$  on the Hessian Riemannian manifold  $(\Omega, G)$ , that is,

$$(8) \quad \begin{cases} \dot{X}(t) = -\nabla^2 g(X(t))^{-1} d_0, & \forall t \in (0, T_{max}), \\ X(0) = x_0. \end{cases}$$

*Proof.* Let  $t \mapsto X(t)$  be the maximal geodesic starting from  $x_0$  with initial velocity  $v_0$ . By (6) and (1),  $X(t)$  solves

$$(9) \quad t d_0 + \nabla g(x) - \nabla g(x_0) = 0.$$

Since,  $X(t) \in \Omega$ , this equation corresponds to the first-order optimality condition for the following minimization problem:

$$\inf \{ t \langle d_0, x \rangle + g(x) - \langle \nabla g(x_0), x \rangle \mid x \in \Omega \},$$

which is equivalent to (7) for any  $(0, T_{max})$ , in terms of optimal solutions. Note that the previous problem is convex, so  $X(t)$  solves (7), and by strict convexity of  $g$  on  $\Omega$ , it is unique.

Evaluating (9) at  $x = X(t)$  and then differentiating with respect to  $t$

$$d_0 + \nabla^2 g(X(t)) \dot{X}(t) = 0.$$

Hence  $X(\cdot)$  solves (8), which corresponds to the steepest descent trajectory for the objective function  $f(x) = \langle d_0, x \rangle$  on the set  $\Omega$  endowed with the Hessian Riemannian metric given by  $G$ , as we claimed.  $\square$

Since Riemannian isometries are arc-length preserving, it is possible to characterize the Riemannian distance on the manifold as well.

**Corollary 3.1** (Geodesic distance). *The geodesic distance  $\text{dist}_\Omega(\cdot, \cdot)$  on the squared Hessian Riemannian manifold associated with a Legendre zone consistent pair  $(\Omega, g)$  satisfies*

$$(10) \quad \text{dist}_\Omega(x_0, x_1) = |\nabla g(x_0) - \nabla g(x_1)|, \quad \forall x_0, x_1 \in \Omega.$$

*In particular, if  $X(t)$  is the geodesic starting from  $x_0 \in \Omega$  with initial velocity  $v_0 = -\nabla^2 g(x_0)^{-1} d_0$  then*

$$\text{dist}_\Omega(X(t), x_0) = t |d_0|, \quad t \geq 0.$$

*Proof.* Set  $y_0 = \nabla g(x_0)$  and  $y_1 = \nabla g(x_1)$ . By convexity of  $\Omega^*$  we have that

$$Y(t) = t y_1 + (1-t) y_0 \in \Omega^*, \quad \forall t \in [0, 1].$$

This curve is a geodesic in  $\Omega^*$  and so, By virtue of [22, Proposition 5.6], the curve  $X(t) = \nabla g^*(Y(t))$  is a geodesic on  $\Omega$ . Hence, Lemma 3.2 implies

$$\text{dist}_\Omega(x_0, x_1) \leq \ell(X) = \ell(Y) = |y_0 - y_1|.$$

On the other hand, for any  $\varepsilon > 0$ , there exists a piecewise differentiable curve  $\gamma_\varepsilon : [0, 1] \rightarrow \Omega$  joining  $x_0$  and  $x_1$  such that

$$\ell(\gamma) \leq \text{dist}_\Omega(x_0, x_1) + \varepsilon.$$

Since  $\tilde{\gamma} = \nabla g \circ \gamma$  is a piecewise differentiable curve joining  $y_0$  and  $y_1$  we have that  $\ell(\tilde{\gamma}) \geq |y_0 - y_1|$ . Therefore, by Lemma 3.2 the length of  $\gamma$  and  $\tilde{\gamma}$  have

the same value, so letting  $\varepsilon \rightarrow 0$  we get (10). The last equation is a direct consequence of Proposition 3.1.  $\square$

On the other hand, the completeness of a Riemannian manifold can also be studied. There are some conditions that ensure this property in the so called *self-concordant case*; see, for instance [24]. We present a sort of dual characterization for squared Hessian Riemannian metrics on convex sets.

**Corollary 3.2** (Completeness). *The Riemannian manifold associated with a Legendre zone consistent pair  $(\Omega, g)$  is geodesically complete if and only if  $\Omega^* = \mathbb{R}^n$ , where  $\Omega^*$  is given by (5). Furthermore,  $(\Omega, \text{dist}_\Omega)$  is a complete metric space if and only if  $\Omega^* = \mathbb{R}^n$ .*

**Remark 3.5.** *Notice that in Remark 3.3, we can verify that  $\Omega^* = \mathbb{R}^n$  and the geodesics are defined for all times.*

*Proof.* Notice that (6) provides an explicit formula for the geodesic starting from  $x_0 \in \Omega$  with initial velocity  $v_0 = -\nabla^2 g(x_0)^{-1} d_0$ . Therefore, we can check that the geodesic is defined for every  $t \in \mathbb{R}$  if and only if  $\nabla g(x_0) - t d_0 \in \Omega^*$  for every  $t \in \mathbb{R}$ . Since  $\Omega$  is an open set, we have no restriction over the vector  $d_0$ , so  $\Omega^*$  must be the entire space  $\mathbb{R}^n$ . The last equivalence is a direct consequence of the Hopf-Rinow's Theorem ([22, Theorem 6.13]).  $\square$

#### 4. APPLICATION TO OPTIMIZATION.

In this section we investigate constrained mathematical programs over sets that can be seen as a squared Hessian Riemannian manifold. We show that it is possible to study this kind of problems by means of other optimization problems which are likely unconstrained.

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a given function. We consider the mathematical program

$$(P) \quad \inf \{f(x) \mid x \in \Omega\}.$$

Let  $\varphi_g : \Omega^* \rightarrow \mathbb{R}$  be the function defined by

$$(11) \quad \varphi_g(y) = f(\nabla g^*(y)), \quad \forall y \in \Omega^*.$$

By virtue of the Legendre coordinate transformation  $y = \nabla g(x)$

$$(12) \quad \forall x \in \Omega, \exists y \in \Omega^* \text{ such that } f(x) = \varphi_g(y).$$

$$(13) \quad \forall y \in \Omega^*, \exists x \in \Omega \text{ such that } f(x) = \varphi_g(y).$$

Therefore, we can associate to (P) an auxiliary optimization problem defined on the dual space  $\Omega^*$  which plays the role of *dual problem*

$$(D) \quad \inf \{\varphi_g(y) \mid y \in \Omega^*\}.$$

**Remark 4.1.** *If  $f \in \Gamma_0(\mathbb{R}^n)$  then the duality scheme proposed here is different from the classical notions of duality in Convex Analysis. The reason is simple: in the classical cases, the convexity of the problem is preserved*

under duality. However, in our case, the dual (resp. primal) function is not necessarily convex even if the primal (resp. dual) function is convex.

Furthermore, we can easily deduce the following result.

**Theorem 4.1.** *Let  $(\Omega, g)$  be a Legendre zone consistent pair satisfying (4) and let  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a given. Suppose  $\{x_k\} \subseteq \Omega$ , then it is a minimizing sequence for (P) if and only if  $\{\nabla g(x_k)\}$  is a minimizing sequence for (D). Moreover, if  $f$  is continuous on its domain and  $\text{val}(\text{P}) \in \mathbb{R}$ , then  $\text{val}(\text{P}) = \text{val}(\text{D})$ . Besides,  $x \in \Omega$  is a local minimizer (maximizer) of  $f$  if and only if  $\nabla g(x)$  is a local minimizer (maximizer) of  $\varphi_g$ .*

*Proof.* Note that by (12) and (13) the affirmation for the minimizing sequence holds true. Suppose that  $\text{val}(\text{P}) \in \mathbb{R}$ , then there exists a minimizing sequence  $\{x_k\}$  such that  $f(x_k) \rightarrow \text{val}(\text{P})$ . For every  $k \in \mathbb{N}$  we can find  $\tilde{x}_k \in \Omega$  so that  $f(\tilde{x}_k) \leq f(x_k) + \frac{1}{k}$ . Hence, noticing that  $f(\tilde{x}_k) \rightarrow \text{val}(\text{P})$  and using that  $\{\nabla g(\tilde{x}_k)\}$  is minimizing sequence for (D) we get the result.

Finally, from (12) and (13) the affirmation about the local optimizers is straightforward.  $\square$

**Remark 4.2.** *Consider the function*

$$f(x_1, x_2) := \begin{cases} \frac{1}{2} \log^2 x_1 + \exp(\tan x_2 \sec^2 x_2) & \text{if } x_1 > 0, x_2 \in (-\frac{\pi}{2}, \frac{\pi}{2}) \\ \frac{1}{2} \log^2 x_1 & \text{if } x_1 > 0, x_2 = -\frac{\pi}{2} \\ +\infty & \text{otherwise.} \end{cases}$$

Let  $\Omega = (0, +\infty) \times (-\frac{\pi}{2}, \frac{\pi}{2})$  and take  $g(x_1, x_2) = x_1 \log x_1 - x_1 + \frac{1}{2} \tan^2 x_2$ . We see that  $(\Omega, g)$  is a Legendre zone consistent pair satisfying (4).

Let us consider the mathematical program

$$(P_1) \quad \min \left\{ f(x_1, x_2) \mid 0 < x_1, -\frac{\pi}{2} \leq x_2 < \frac{\pi}{2} \right\}.$$

Hence, the dual problem is

$$(D_1) \quad \min \left\{ \frac{1}{2} y_1^2 + \exp y_2 \mid y_1, y_2 \in \mathbb{R} \right\}.$$

Notice that the solution of  $(P_1)$  is  $(1, -\frac{\pi}{2}) \in \text{bdry } \Omega$  and the solution of  $(D_1)$  is not attained, but  $\{(0, -k)\}_k$  is a minimizing sequences for the dual problem and consequently,  $\text{val}(P_1) = \text{val}(D_1)$ .

As one may expect, differentiability properties on the primal function are inherited by the dual function  $\varphi_g$ , and vice versa. This is summarized in the following statement.

**Proposition 4.1.** *Let  $(\Omega, g)$  be a Legendre zone consistent pair that satisfies (4). Suppose that  $f$  is Gâteaux differentiable at  $x \in \Omega$ , then  $\varphi_g$  is Gâteaux differentiable at  $y = \nabla g(x)$  as well with*

$$\varphi'_g(y, d) = f'(x, G(x)^{-1}d), \quad \forall d \in \mathbb{R}^n.$$

Moreover,  $f$  is Frechet differentiable at  $x \in \Omega$  if and only if  $\varphi_g$  is so at  $y = \nabla g(x)$ . In such a case we have

$$\nabla \varphi_g(y) = \nabla_G f(x) := G(x)^{-1} \nabla f(x).$$

*Proof.* It is enough to use the chain rule and the properties of  $g$  and  $g^*$ .  $\square$

Now this type of result can also be stated for non differentiable functions. Indeed, it is possible to characterize the viscosity subgradients (resp. supergradients) of  $\varphi_g$  when it is only lower (resp. upper) semicontinuous.

**Definition 4.1.** Let  $v : \mathbb{R}^k \rightarrow \mathbb{R} \cup \{+\infty\}$  be a given function and  $z \in \text{dom} v$ .

- a. If  $v$  is lower semicontinuous, a vector  $\zeta \in \mathbb{R}^k$  is called a viscosity subgradient of  $v$  at  $z$  if and only if there exists a continuous function  $\psi$  which is differentiable at  $z$  so that

$$\nabla \psi(z) = \zeta \text{ and } v - \psi \text{ attains a local minimum at } z.$$

The set of all the viscosity subgradients of  $v$  at  $z$  is called the viscosity subdifferential and it is denoted by  $\partial_D v(z)$ .

- b. If  $v$  is upper semicontinuous, a vector  $\zeta \in \mathbb{R}^k$  is called a viscosity supergradient of  $v$  at  $x$  if and only  $-\zeta \in \partial_D(-v)(z)$ . The set of all the viscosity supergradients of  $v$  at  $z$  is called the viscosity superdifferential and it is denoted by  $\partial^D v(z)$ .

The following result is going to be useful in section 6 where we use the viscosity sub and supergradients to investigate the Value Function of a Mayer problem in the application to Optimal Control.

**Proposition 4.2.** Let  $(\Omega, g)$  be a Legendre zone consistent pair that satisfies (4). Suppose that  $f$  is lower semicontinuous, then

$$\partial_D \varphi_g(\nabla g(x)) = G(x)^{-1} \partial_D f(x), \quad \forall x \in \Omega.$$

If on the other hand,  $f$  is upper semicontinuous, then

$$\partial^D \varphi_g(\nabla g(x)) = G(x)^{-1} \partial^D f(x), \quad \forall x \in \Omega.$$

*Proof.* Suppose  $f$  is lower semicontinuous, the other case is analogous. Let  $x_0 \in \Omega$ ,  $y_0 = \nabla g(x_0)$  and  $\zeta \in \partial_D \varphi_g(y_0)$ , then there exist  $\delta > 0$  and a continuous function  $\psi$ , differentiable at  $y_0$  such that  $\nabla \psi(y_0) = \zeta$  and

$$\varphi_g(y) - \psi(y) \geq \varphi_g(y_0) - \psi(y_0), \quad \forall y \in \mathbb{B}(y_0, \delta).$$

Set  $\mathcal{O} = \nabla g^*(\mathbb{B}(y_0, \delta))$  and consider  $\psi_g : \mathcal{O} \rightarrow \mathbb{R}$  given by

$$\psi_g(x) = \psi(\nabla g(x)), \quad \forall x \in \mathcal{O}.$$

This function is continuous on  $\mathcal{O}$  and differentiable at  $x$  by Proposition 4.1 with  $\nabla \psi(x) = G(x)^{-1} \nabla \psi_g(x)$ . Furthermore, by Proposition 4.1,  $x_0$  is a local minimizer of  $f - \psi_g$ . Hence  $\psi_g$  is an admissible test function and  $\nabla \psi_g(x_0) \in \partial_D f(x_0)$ , so  $\partial_D \varphi_g(\nabla g(x_0)) \subseteq G(x_0)^{-1} \partial_D f(x_0)$ . The other inclusion is similar, therefore the conclusion follows.  $\square$

As we showed in the previous section, if the Squared Hessian Riemannian manifold is complete then  $\Omega^* = \mathbb{R}^n$  and so, for any cost  $f$  we can associate a dual problem to (P) which has no constraints. In the incomplete case, the dual problem is constrained as well. However, under suitable hypotheses we can show that, if the minimum is achieved, it belongs to  $\Omega^*$  because  $\varphi_g$  blows up at the boundary of  $\Omega^*$ .

**Proposition 4.3.** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a coercive function on its domain and let  $(\Omega, g)$  be a Legendre zone consistent pair that satisfies (4). Then, for any  $\{y_k\} \subseteq \Omega^*$  with  $y_k \rightarrow y \in \partial\Omega^*$ ,*

$$\liminf_{k \rightarrow +\infty} \varphi_g(y_k) = +\infty.$$

*Proof.* Let  $\{y_k\} \subseteq \Omega^*$  with  $y_k \rightarrow y$  for some  $y \in \partial\Omega^*$  and let  $\ell = \liminf \varphi_g(y_k)$ . Let  $x_k = \nabla g^*(y_k)$ , then

$$\liminf_{k \rightarrow +\infty} \varphi_g(y_k) = \liminf_{k \rightarrow +\infty} f(\nabla g^*(y_k)) = \liminf_{k \rightarrow +\infty} f(x_k).$$

Recall that  $g^*$  is also a Legendre function, so  $|\nabla g^*(y_k)| = |x_k| \rightarrow +\infty$ . Hence, since  $f$  is coercive on its domain,  $\ell = +\infty$ .  $\square$

## 5. GRADIENT-LIKE ALGORITHMS FOR SOLVING OPTIMIZATION PROBLEMS.

**5.1. Geodesic search methods.** Inspired by the dual scheme proposed in the previous section, we study an interior point algorithm for the constrained optimization problem (P). This strategy can be seen as a line search algorithm on the dual space  $\Omega^*$  transported to the original space by means of the isometry given by Lemma 3.2.

In this section we assume

$$(H_0) \quad \begin{cases} (i) & (\Omega, g) \text{ is a Legendre zone consistent pair satisfying (4).} \\ (ii) & f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \text{ belongs to } \mathcal{C}^1(\Omega). \\ (iii) & \inf\{f(x) \mid x \in \bar{\Omega}\} = \inf\{f(x) \mid x \in \Omega\} \in \mathbb{R}. \\ (iv) & (\Omega, G^2) \text{ is a complete Riemannian manifold.} \end{cases}$$

The general scheme we propose is the following:

- a.** Choose  $x_0 \in \Omega$ .
- b.** If  $\nabla f(x_0) = 0$  then  $x_0$  solves the problem  $\rightarrow$  **stop**,  
**else**  $\rightarrow$  set  $X(t) = \nabla g^*(\nabla g(x_0) + td_0)$  for some  $d_0 \in \mathbb{R}^n$ .
- c.** Choose step length  $t_0 > 0$  such that  $f(x_0) > f(X(t_0))$ .
- d.** Set  $x_0 = X(t_0)$ , return to **b**.

We call this method the *Squared Hessian Riemannian Geodesic Search* algorithm (SHRGS for short) associated with the Legendre zone consistent pair  $(\Omega, g)$ . This strategy can be classified as a Geodesic search method; see for instance [28, 33]. However, the principal difference with those techniques, and also the main difficulty for the analysis, is that our Riemannian manifold

$\Omega$  is in reality an open set and so, solution of (P) may lie at the boundary of  $\Omega$  which is no longer a part of the manifold.

**Remark 5.1.** *The line search strategy in  $\mathbb{R}^n$  is a particular case of the SHRGS method. Indeed, if  $\Omega = \mathbb{R}^n$ , then  $g(x) = \frac{1}{2}|x|^2$  is an admissible Legendre function for (i) and (iv) in  $(H_0)$  to hold and furthermore,  $G(x)$  is the identity matrix, and so, the geodesics in this case are just line segments.*

For a given  $x_0 \in \Omega$ , the SHRGS method generates the following sequence:

$$(14) \quad x_{k+1} = \nabla g^*(\nabla g(x_k) + t_k d_k) \quad \forall k \geq 0,$$

where  $t_k > 0$  with  $X(t_k) \in \Omega$  and  $f(x_{k+1}) < f(x_k)$  whenever  $d_k \neq 0$ .

**5.2. Riemannian descent directions.** The interest of this algorithm relies in the dual lecture that we can give to (14). First of all, note that by Corollary 3.2 and  $(H_0)$  the dual space is the whole space  $\mathbb{R}^n$ . So the dual problem (D) is unconstrained. Moreover, if we transport the equation (14) to the dual space  $\Omega^*$  by means of the isometry  $\nabla g(\cdot)$  and we set  $y_k = \nabla g(x_k)$ ,

$$y_{k+1} = y_k + t_k d_k, \quad \forall k \geq 0.$$

This sequence correspond to a line search strategy on the dual space  $\Omega^*$ . So, considering this, we have a natural way to define descent directions and rules for choosing the step-size  $t_k$  based on the line search methods applied to the unconstrained problem (D). Recall that  $\nabla_G f(x) = G(x)^{-1} \nabla f(x)$ .

**Definition 5.1.** *Let  $(\Omega, g)$  be a Legendre zone consistent pair satisfying (4) and  $f$  differentiable at  $x \in \Omega$ . We say that  $d$  is a descent direction for the function  $f$  on  $(\Omega, G^2)$  at  $x$  provided  $\langle \nabla_G f(x), d \rangle < 0$ .*

The next table shows some suitable choices for the descent directions. By Proposition 4.1 we have that  $\nabla \varphi_g(y_k) = \nabla_G f(x_k)$ , and so we can also exhibit the expressions in terms of the dual variable  $y_k$ .

SHRGS method	$d_k$ in $\Omega$	$d_k$ in $\Omega^*$
Steepest descent	$-\nabla_G f(x_k)$	$-\nabla \varphi_g(y_k)$
Quasi-Newton	$-M_k^{-1} \nabla_G f(x_k)$	$M_k^{-1} \nabla \varphi_g(y_k)$

TABLE 1. Some descent directions for a given point  $x_k$ .

In Table 1,  $M_k$  is a real nonsingular matrix. Note that the directions of Table 1 yield the steepest descent and the quasi-Newton method for the dual problem (D), respectively. In particular, the way in which  $M_k$  is chosen can be based on conditions involving the dual function  $\varphi_g$ . We refer to the reader to [25, Chapter 6] for more details about the Quasi-Newton methods.

**Remark 5.2.** *Other choices for the descent directions, not based in the dual problem, are also possible; in fact, they have been already studied for convex mathematical programming. In [4] the authors considered the iteration*

$$x_{k+1} \in \arg \min \{ \lambda_k \langle \nabla f(x_k), x \rangle + D_g(x, x_k) : x \in \Omega \}, \quad \forall k \in \mathbb{N}.$$

In the light of Proposition 3.2, this algorithm corresponds exactly to a SHRGS iteration with  $d_k = -\nabla f(x_k)$ .

**Proposition 5.1.** *Let  $(\Omega, g)$  be a Legendre zone consistent pair satisfying (4) and  $f \in \mathcal{C}^1(\Omega)$ . Let  $x_0 \in \Omega$  and take  $d_0 \in \mathbb{R}^n$  a descent direction for the function  $f$  on  $(\Omega, G^2)$  at  $x_0$ . Then there exists  $t_0 > 0$  such that*

$$f(x_0) > f(\nabla g^*(\nabla g(x_0) + t_0 d_0)).$$

In particular, the sequence (14) is well-defined whenever  $d_k$  is a descent direction for  $f$  on  $(\Omega, G^2)$  at  $x_k$  and  $\{f(x_k)\}_{k \geq 0}$  is non-increasing.

*Proof.* By Proposition 3.1 there exist  $T_{min} < 0 < T_{max}$  such that  $X(t) = \nabla g^*(\nabla g(x_0) + t d_0)$  is for any  $t \in (T_{min}, T_{max})$ . Let  $\phi(t) = f(X(t))$ . Hence

$$\frac{d}{dt}\phi(t) = -\langle \nabla f(X(t)), \nabla^2 g^*(\nabla g(x_0) + t d_0) d_0 \rangle, \quad \forall t \in (T_{min}, T_{max}).$$

Recall that  $\nabla g^*(\nabla g(x)) = x$  for any  $x \in \Omega$ , this implies that

$$\nabla^2 g^*(\nabla g(x)) = [\nabla^2 g(x)]^{-1}, \quad \forall x \in \Omega.$$

Therefore,  $\frac{d}{dt}\phi(t) = \langle \nabla_G f(X(t)), d_0 \rangle$ . Consequently, since  $\phi(0) < 0$  and  $t \mapsto \nabla_G f(X(t))$  is continuous, there exists  $\delta > 0$  such that  $\frac{d}{dt}\phi(t) < 0$  for any  $t \in [0, \delta)$ . So the conclusion follows.  $\square$

**Remark 5.3.** *Thanks to Proposition 4.3, hypothesis (iv) in  $(H_0)$  can be replaced by requiring the coercivity of  $f$ . In this case, the sequence generated by (14) is well defined because the  $\varphi_g$  blows up at the boundary of  $\Omega^*$ .*

**5.3. Step length and convergence.** We can also define rules for choosing the step length  $t_k$  by transferring the information from the Euclidean case to the Riemannian one.

**Definition 5.2.** *Let  $(\Omega, g)$  be a Legendre zone consistent pair satisfying (4) and  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a differentiable function on  $\Omega$ . Let  $d_0$  be a descent direction for  $f$  on  $(\Omega, G^2)$  at  $x_0$ . Consider  $X(t) = \nabla g^*(\nabla g(x_0) + t d_0)$  for any  $t \in \mathbb{R}$ . Let  $t_0 > 0$  be a given step length. We say that  $t_0$  satisfies:*

(i) *the Riemannian exact line search or Cauchy rule if*

$$t_0 \in \arg \min\{f(X(t)) \mid t > 0\}.$$

(ii) *the Riemannian Wolfe rule if there exist  $0 < c_1 < c_2 < 1$  so that*

$$\begin{cases} f(X(t_0)) \leq f(x_0) + c_1 t_0 \langle \nabla_G f(x_0), d_0 \rangle, \\ \langle \nabla_G f(X(t_0)), d_0 \rangle \geq c_2 t_0 \langle \nabla_G f(x_0), d_0 \rangle. \end{cases}$$

Note that the Riemannian exact line search and the Riemannian Wolfe rules correspond to apply the exact line search and the Wolfe rules to the dual problem, respectively; see for instance [25, Chapter 3] for a concise exposition about these rules. Hence, it is not difficult to see that under  $(H_0)$  these rules are well defined; we refer the reader to [25, Lemma 3.1].

On the other hand, we can also present an adapted version of the Zou-tendijk theorem of convergence.



**Proposition 5.2.** *Suppose that  $(H_0)$  holds and let  $x_0 \in \Omega$ . Consider the sequence given by (14) with  $d_k$  a descent direction for  $f$  on  $(\Omega, G^2)$  at  $x_k$  and  $t_k$  is chosen with the Riemannian Wolfe rule. Then  $\nabla_G f(x_k) \rightarrow 0$  as  $k \rightarrow +\infty$  provided that*

a. *there exists  $\delta > 0$  such that*

$$(15) \quad \cos \theta_k = -\frac{\langle \nabla_G f(x_k), d_k \rangle}{|\nabla_G f(x_k)| |d_k|} > \delta, \quad \forall k \in \mathbb{N}.$$

b.  *$x \mapsto \nabla_G f(x)$  is Lipschitz in the following sense:  $\exists L > 0$  so that  $\forall x, x' \in \{x \in \Omega \mid f(x) \leq f(x_0)\}$  we have*

$$(16) \quad |\nabla_G f(x) - \nabla_G f(x')| \leq L \operatorname{dist}_\Omega(x, x').$$

*Proof.* By Proposition 4.1 and Corollary 3.1 we get that (16) implies that  $y \mapsto \nabla \varphi_g(y)$  is  $L$ -Lipschitz on  $\{y \in \mathbb{R}^n \mid \varphi_g(y) \leq f(x_0) = \varphi_g(y_0)\}$ . Since the Riemannian Wolfe rule corresponds to apply the Wolfe rule on the dual problem, we have that the so-called Zoutendijk condition holds for the dual function  $\varphi_g$ ; see for instance [25, Theorem 3.2]. Therefore,

$$\sum_{k=0}^{\infty} \left( \frac{\langle \nabla \varphi_g(y_k), d_k \rangle}{|\varphi_g(y_k)| |d_k|} \right)^2 |\nabla \varphi_g(y_k)|^2 < +\infty.$$

Finally, by (15) and Proposition 4.1 we get that

$$\sum_{k=0}^{\infty} |\nabla \varphi_g(y_k)|^2 < +\infty.$$

Hence, in particular  $\nabla_G f(x_k) = \nabla \varphi_g(y_k) \rightarrow 0$  as long as  $k \rightarrow +\infty$ .  $\square$

**5.4. Convexity properties and value convergence.** In this section we investigate the special case when the function  $f$  is convex on  $(\Omega, G^2)$  in the Riemannian sense as in Definition 2.3.

In the case of squared Hessian Riemannian manifolds, the convexity of  $f$  on  $(\Omega, G^2)$  yields to the convexity (in the classical sense) of dual function  $\varphi_g$ . Indeed, by (6) we have an expression for every geodesic on  $\Omega$

$$\begin{cases} \gamma_{xy}(t) = \nabla g^*(\nabla g(y) + t(\nabla g(x) - \nabla g(y))), \\ \gamma_{xy}(0) = x \in \Omega \quad \text{and} \quad \gamma_{xy}(1) = y \in \Omega. \end{cases}$$

Then, since  $f = \varphi_g \circ \nabla g$  and if we set  $u = \nabla g(x)$ ,  $v = \nabla g(y)$  we have

$$f \circ \gamma_{xy}(t) = \varphi_g(v + t(v - u)).$$

Moreover,

$$tf(y) + (1-t)f(x) = t\varphi_g(v) + (1-t)\varphi_g(u),$$

so  $\varphi_g$  is convex in classical sense on  $\Omega^*$ .

**Remark 5.4.** Consider  $f$  as in Remark 4.2 and  $\Omega = (0, +\infty) \times (-\frac{\pi}{2}, \frac{\pi}{2})$ . We also take  $g(x_1, x_2) = x_1 \log x_1 - x_1 + \frac{1}{2} \tan^2 x_2$  as well. In this case,  $f$  is convex on the Riemannian manifold  $(\Omega, G^2)$ , because its dual function is

$$\varphi_g(y_1, y_2) = \frac{1}{2} y_1^2 + \exp y_2, \quad (y_1, y_2) \in \Omega^* = \mathbb{R}^2.$$

In this case the theorem of convergence is stated as follows.

**Theorem 5.1.** Let  $f$  be convex on  $(\Omega, G^2)$  in the Riemannian sense. Let  $\{x_k\}$  the sequence induced by (14) starting from  $x_0 \in \Omega$ . Assume that  $(H_0)$  holds and that  $x \mapsto \nabla_G f(x)$  is Lipschitz on the level sets of  $f$ . Take  $t_k \in [0, h_{max}]$  with  $0 < h_{max} < \frac{2}{L}$  and  $\sum t_k = \infty$  where  $L$  is the Lipschitz constant of  $[f \leq f(x_0)]$  and consider the steepest descent direction in (14).

- a. Then  $f(x_k) \rightarrow \inf_{\Omega} f$  as long as  $k \rightarrow \infty$ .
- b. If  $\arg \min f \cap \Omega \neq \emptyset$ , then the sequence induced by (14) converges to  $x_{\infty} \in \arg \min f \cap \Omega$ .

To prove the previous theorem we need to recall a convergence result for the gradient method in the convex case; see, for instance [12].

**Lemma 5.1.** Let  $\Omega^* \subseteq \mathbb{R}^n$  be nonempty, open and convex. Let  $\varphi_g \in \Gamma_0(\mathbb{R}^n)$  be such that  $\varphi_g|_{\Omega^*} \in \mathcal{C}^1(\Omega^*)$ . Suppose that  $\varphi_g$  is bounded from below and that  $\nabla \varphi_g$  is Lipschitz on the level sets of  $\varphi_g$ . Let  $\{y_k\} \subseteq \Omega^*$  be a sequence such that  $\{\varphi_g(y_k)\}$  is nonincreasing and

$$y_{k+1} = y_k - h_k \nabla \varphi_g(y_k), \quad k \in \mathbb{N},$$

starting from  $y_0 \in \mathbb{R}^n$ ,  $h_k \in [0, h_{max}]$  with  $0 < h_{max} < \frac{2}{L}$ , where  $L$  is the Lipschitz constant of  $[\varphi_g \leq \varphi_g(y_0)]$  and  $\sum h_k = \infty$ . Then

- a.  $\varphi_g(y_k) \rightarrow \inf_{\Omega^*} \varphi_g$  as long as  $k \rightarrow \infty$ .
- b. If  $\arg \min_{\Omega^*} \varphi_g \neq \emptyset$ ,  $\exists y_{\infty} \in \arg \min_{\Omega^*} \varphi_g$  such that  $y_k \rightarrow y_{\infty}$ .

*Proof of Theorem 5.1.* Recall that by  $(H_0)$   $\Omega^* = \mathbb{R}^n$ . Hence, by the hypothesis on the statement,  $\varphi_g \in \mathcal{C}^1(\mathbb{R}^n)$  and is also convex. Moreover,  $\varphi_g$  is  $L$ -Lipschitz continuous on level sets of  $\varphi_g$ , indeed, we know that

$$\text{dist}_{\Omega}(x, y) = |\nabla g(x) - \nabla g(y)|,$$

and so

$$\begin{aligned} |\nabla \varphi_g(u) - \nabla \varphi_g(v)| &= |\nabla_G f(\nabla g^*(u)) - \nabla_G f(\nabla g^*(v))| \\ &\leq L \text{dist}_{\Omega}(\nabla g^*(u), \nabla g^*(v)) \\ &\leq L|u - v|. \end{aligned}$$

Therefore, by Proposition 5.1,  $\varphi_g(y_k) \rightarrow \inf \varphi_g$  that implies the first conclusion. Besides, if  $\arg \min f \cap \Omega \neq \emptyset$  then  $S := \arg \min \varphi_g \neq \emptyset$ . Hence, there exists  $y_{\infty} \in S$  such that  $y_k \rightarrow y_{\infty}$ , and consequently  $x_k$  is a minimizing sequence for (P), which in addition converges to  $x_{\infty} := \nabla g^*(y_{\infty})$ .  $\square$

## 6. APPLICATION TO OPTIMAL CONTROL PROBLEMS.

In this section we present an application to an infinite dimensional optimization problem with dynamics constraints. We consider an optimal control problem of Mayer type with convex state constraints.

It is well known that, unless some compatibility condition between constraints and dynamics holds, the Value Function has not enough regularity, or can fail to be the unique constrained viscosity solution of a Hamilton-Jacobi-Bellman equation. In this section explore the fact that the original optimal control problem can be transported into an unconstrained one, and we use this to characterize the original Value Function in terms of the Hamilton-Jacobi-Bellman equation. We want to emphasis that the class of systems of concern do not satisfy either the IPC or the OPC.

Consider the differential inclusion

$$(17) \quad \begin{cases} \dot{X}(t) \in F(X(t)) & \text{a.e. } t \in [\tau, T], \\ X(\tau) = x, \end{cases}$$

where  $T > 0$  is the final horizon,  $\tau \in [0, T]$  is the initial time and  $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is a set-valued map called the dynamics. Under mild hypothesis, for any  $x \in \mathbb{R}^n$ , the differential inclusion (17) admits solutions in the space of the absolutely continuous functions; see for instance [3, Chapter 2]

We are interested in optimal control problems where the arcs of (17) are confined to a closed state constraints set. We are specially concerned with sets that are the closure of a nonempty open convex set  $\Omega$ , that is,

$$(18) \quad X(t) \in \bar{\Omega}, \quad \forall t \in [\tau, T].$$

We denote by  $S_{\bar{\Omega}}^F(\tau, x)$  the set of absolutely continuous functions satisfying (17) and (18) with initial condition  $X(\tau) = x$ .

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a given function. Thus, the Mayer problem associated with the state-constrained control systems (17)-(18) is

$$(19) \quad V(\tau, x) := \inf \{f(X(T)) \mid X(\cdot) \in S_{\bar{\Omega}}^F(\tau, x)\}, \quad \forall (\tau, x) \in [0, T] \times \bar{\Omega}.$$

Where  $V : [0, T] \times \bar{\Omega} \rightarrow \mathbb{R} \cup \{+\infty\}$  is the Value Function related to the Mayer problem (19) that satisfies

$$V(T, x) = f(x), \quad \forall x \in \bar{\Omega}.$$

Let us see some properties of the Value Function.

**Proposition 6.1.** *Suppose that  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is lower semicontinuous on  $\bar{\Omega}$  and that  $F : [0, T] \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  satisfies*

$$(20) \quad \begin{cases} i) & F(x) \text{ is nonempty compact and convex, } \forall x \in \bar{\Omega}. \\ ii) & x \mapsto F(x) \text{ has closed graph on } \bar{\Omega}. \\ iii) & \exists c_F > 0 \text{ such that } \forall x \in \bar{\Omega} \\ & \max\{|v| \mid v \in F(x)\} \leq c_F(1 + |x|). \end{cases}$$

If  $V(\tau, x) \in \mathbb{R}$  then there exists an optimal trajectory  $\bar{x} \in S_{\Omega}^F(\tau, x)$ . Furthermore, the  $V : [0, T] \times \bar{\Omega} \rightarrow \mathbb{R} \cup \{+\infty\}$  is lower semicontinuous.

*Proof.* Let  $(\tau, x) \in [0, T] \times \bar{\Omega}$ . Suppose that  $V(\tau, x) \in \mathbb{R}$  and let  $\{\varepsilon_k\} \subseteq (0, 1)$  be a sequence such that  $\varepsilon_k \rightarrow 0$  as  $k \rightarrow +\infty$ . Thus for each  $n \in \mathbb{N}$ , there exists an absolutely continuous function  $X_k(\cdot) \in S_{\Omega}^F(\tau, x)$  for which

$$(21) \quad f(X_k(T)) \leq V(\tau, x) + \varepsilon_k.$$

By (20) and the Gronwall's Lemma,  $\{X_k\}$  is uniformly bounded on  $[\tau, T]$  and  $\{\dot{X}_k\}$  is dominated by a function in  $L^1([\tau, T]; \mathbb{R}^n)$ . Thus, by [3, Theorem 0.3.4], we derive that there exist a function  $\bar{X} \in W^{1,1}([\tau, T]; \mathbb{R}^n)$  and a subsequence, still denoted by  $X_k$ , which converges uniformly to  $\bar{X}$  on  $[\tau, T]$  and  $\dot{X}_k$  converges weakly to  $\dot{\bar{X}}$  in  $L^1([\tau, T]; \mathbb{R}^n)$ . Then, by (20) and the Convergence Theorem [3, Theorem 1.4.1] we get that  $\bar{X} \in S_{\Omega}^F(\tau, x)$ . Therefore, taking the limit inferior in (21), using the lower semicontinuity of  $f$  we get that  $\bar{X}$  is a minimizer.

Let  $\{(\tau_k, x_k)\} \subseteq [0, T] \times \bar{\Omega}$  converging to some  $(\tau, x)$ . Without loss of generality we can restrain our attention to the case  $V(\tau_k, x_k) \in \mathbb{R}$ . By the previous part, there exists  $\bar{X}_k \in S_{\Omega}^F(\tau_k, x_k)$  so that

$$(22) \quad V(\tau_k, x_k) = f(\bar{X}_k(T)).$$

Assume that  $\tau \geq \tau_k$  for every  $n \in \mathbb{N}$ . Otherwise, we can extend the trajectory  $\bar{X}_k$  on the interval  $[\tau, \tau_k)$ . Using the same argument as in the first part, there exists  $X \in S_{\Omega}^F(\tau, x)$ , such that  $\bar{X}_k \rightarrow X$  uniformly in  $[\tau, T]$ . Hence, taking the limit inferior in (22), using the lower semicontinuity of  $f$  and the definition of the Value Function we conclude the proof.  $\square$

**6.1. Dual optimal control problem.** Suppose that there exists  $g \in \Gamma_0(\mathbb{R}^n)$  for which  $(\Omega, g)$  is a Legendre zone consistent pair such that  $(\Omega, G^2)$  is a complete Riemannian manifold.

Take  $x \in \Omega$  and  $\tau \in [0, T]$ . Let  $X(t)$  be a solution of (17) that satisfies  $X(t) \in \Omega$  for any  $t \in [\tau, T]$ . Then the arc  $t \mapsto Y(t) := \nabla g(X(t))$  is a solution of the differential inclusion

$$(23) \quad \begin{cases} \dot{Y}(t) \in \Phi_g(Y(t)) & \text{a.e. } t \in [\tau, T], \\ Y(\tau) = y, \end{cases}$$

where  $y = \nabla g(x)$  and  $\Phi_g(y) = [\nabla^2 g^*(y)]^{-1} F(\nabla g^*(y))$ .

**Remark 6.1.** Let  $\Omega$  be given and suppose that there exists a Legendre function  $g \in \Gamma_0(\mathbb{R}^N)$ , together with two matrices  $A$  and  $B$  of dimensions  $n \times n$  and  $n \times m$ , respectively, for which

$$\nabla^2 g(x)F(x) = \{A\nabla g(x) + Bu : u \in [-1, 1]^m\}, \quad \forall x \in \Omega.$$

Then in this case we get that the dual differential equation (23) is a linear system because

$$\Phi_g(y) = \{Ay + Bu : u \in [-1, 1]^m\}, \quad \forall y \in \mathbb{R}^N.$$

Furthermore, if there exist two vector fields  $F_1, F_2 : \mathbb{R}^n \rightarrow \mathbb{R}^n$  so that

$$\nabla^2 g(x)F(x) = \{F_1(\nabla g(x)) + F_2(\nabla g(x))u : u \in [-1, 1]^m\}, \quad \forall x \in \Omega,$$

the dual differential equation (23) is a control affine system.

Similarly as done earlier, we denote by  $S^{\Phi_g}(\tau, y)$  set of absolutely continuous functions satisfying (23) with initial condition  $Y(\tau) = y$ .

**Remark 6.2.** Clearly, we can associate to any solution to (23) a solution to (17) by means of the isometry  $\nabla g(\cdot)$ . Moreover, that solution lies in  $\Omega$ , the interior of the state constraint. Note that the converse is also true.

Whenever  $(\Omega, G^2)$  is a complete Riemannian manifold, by Corollary 3.2,  $Y(\cdot)$  dwells in  $\Omega^* = \mathbb{R}^n$ , and so we can associate to (23) an unconstrained Mayer problem in the following way:

$$(24) \quad W_g(\tau, y) := \inf \{ \varphi_g(Y(T)) \mid Y(\cdot) \in S^{\Phi_g}(\tau, y) \}, \quad \forall (\tau, y) \in [0, T] \times \mathbb{R}^n,$$

where  $\varphi_g$  is given by (11) and  $W_g : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is the Value Function associated with the Mayer problem (24).

By remark 6.2, we see that  $V(\tau, x) \leq W_g(\tau, \nabla g(x))$  for any  $x \in \Omega$  and without additional hypotheses we can not expect the equality. However, under an interior approximation hypothesis it is possible to obtain it.

**Proposition 6.2.** Suppose that the Riemannian manifold associated with a Legendre zone consistent pair  $(\Omega, g)$  is complete. Assume that  $f$  is uniformly continuous of modulus  $\omega_f(\cdot)$  and that the following holds:

$$(H_1) \quad \begin{cases} \forall \varepsilon > 0, \forall \tau \in [0, T], \forall x \in \Omega, \forall X(\cdot) \in S_{\Omega}^F(\tau, x) \\ \exists X_{\varepsilon}(\cdot) \in S_{\Omega}^F(\tau, x) \text{ such that:} \\ X_{\varepsilon}(t) \in \Omega, \forall t \in [\tau, T] \text{ and } |X(T) - X_{\varepsilon}(T)| \leq \varepsilon. \end{cases}$$

Let  $V$  and  $W_g$  be the Value Functions of (19) and (24) respectively. Then

$$W_g(\tau, \nabla g(x)) = V(\tau, x), \quad \forall (\tau, x) \in [0, T] \times \Omega.$$

**Remark 6.3.** An interior approximation hypothesis is a quite natural hypothesis in the convex context where we are working in. This hypothesis does not require the existence of solutions for any initial state, so in particular,  $(H_1)$  may hold even if  $\bar{\Omega}$  is not a viable domain for the dynamics  $F$ . For a stronger sufficient condition for  $(H_1)$  to hold we refer to [7, Theorem 3.2].

*Proof.* If  $V(\tau, x) = +\infty$ , by Remark 6.2 we get that  $W_g(\tau, \nabla g(x)) = +\infty$ .

Assume that  $V(\tau, x) < +\infty$  and take  $\{\varepsilon_k\} \subseteq (0, 1)$  a sequence such that  $\varepsilon_k \rightarrow 0$  as  $k \rightarrow +\infty$ . Thus, for any  $n \in \mathbb{N}$ , there exists  $X_k(\cdot) \in S_{\Omega}^F(\tau, x)$  for which

$$f(X_k(T)) \leq V(\tau, x) + \varepsilon_k.$$

Besides, by  $(H_1)$ , for any  $n \in \mathbb{N}$  there exists  $\tilde{X}_k(\cdot) \in S_{\Omega}^F(\tau, x)$  such that

$$\tilde{X}_k(t) \in \Omega, \forall t \in [\tau, T] \quad \text{and} \quad |\tilde{X}_k(T) - X_k(T)| \leq \varepsilon_k.$$

Hence

$$f\left(\tilde{X}_k(T)\right) \leq f(X_k(T)) + \omega_f(\varepsilon_k) = V(\tau, x) + \varepsilon_k + \omega_f(\varepsilon_k).$$

Moreover, since  $\tilde{Y}_k(\cdot) := \nabla g\left(\tilde{X}_k(\cdot)\right) \in S^{\Phi_g}(\tau, \nabla g(x))$ ,

$$f\left(\tilde{X}_k(T)\right) = \varphi_g\left(\tilde{Y}_k(T)\right) \geq W_g(\tau, \nabla g(x)) \geq V(\tau, x).$$

So, letting  $k \rightarrow +\infty$ , we get the desired equality.  $\square$

The importance of the previous result relies in the fact that for optimal control without state constraints the Value Function can be characterized as the unique solution to a Hamilton-Jacobi-Bellman equation. In this case the corresponding equation related to (24) is written as

$$(25) \quad v_\tau + H(y, v_y) = 0, \quad (\tau, y) \in (0, T) \times \mathbb{R}^n,$$

$$(26) \quad v(T, y) = \varphi_g(y), \quad y \in \mathbb{R}^n.$$

where  $H : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  is given by:

$$H(y, p) := \sup \{ -\langle w, p \rangle \mid w \in \Phi_g(y) \}.$$

**Remark 6.4.** *The Hamiltonian can be written in term of the original dynamics  $F$  as follows*

$$H(y, p) := \sup \{ \langle \nabla^2 g(\nabla g^*(y))w, p \rangle \mid w \in F(\nabla g^*(y)) \}.$$

Solutions to (25) are understood in the so called viscosity sense introduced by Crandall and Lions; see for instance [5, Chapter 3]. We mention also that, solutions to (25) can be understood in other equivalent senses; we refer to [14, Chapter 4.7] or [32, Chapter 12] for a more detailed discussion.

Recall the definition of the viscosity subgradients and supergradients in Definition 4.1. We make precise the notion of viscosity solution.

**Definition 6.1.** *A continuous function  $v : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$  is a viscosity solution to (25) if*

$$\theta + H(y, \zeta) \leq 0, \quad \forall (\tau, y) \in (0, T) \times \mathbb{R}^n, \quad \forall (\theta, \zeta) \in \partial_D v(\tau, y),$$

$$\theta + H(y, \zeta) \geq 0, \quad \forall (\tau, y) \in (0, T) \times \mathbb{R}^n, \quad \forall (\theta, \zeta) \in \partial^D v(\tau, y).$$

We are interested in providing a characterization of the auxiliary Value Function  $W_g$  as the unique viscosity solution in the sense of the previous definition. For this purpose we are going to study a particular type of controlled systems that exhibits an absorbing property at the boundary.

**Proposition 6.3.** *Consider the state-constrained control system (17)-(18) and assume that (20) holds and that  $f$  is continuous.*

Suppose that there exists  $g \in \Gamma_0(\mathbb{R}^n)$  such that  $(\Omega, G^2)$  is a complete Riemannian manifold. We assume that the control system (17)-(18) is absorbing at  $\partial\Omega$  in the sense that

$$(27) \quad \max_{v \in F(x)} |v| \leq c_F \left( \frac{1 + |\nabla g(x)|}{|\nabla^2 g(x)|} \right) \quad \forall x \in \Omega.$$

We also assume that  $F$  satisfies a Lipschitz-like property for the riemannian distance, that is,  $\forall R > 0, \exists L_R > 0$  so that

$$(28) \quad \nabla^2 g(x)F(x) \subseteq \nabla^2 g(\tilde{x})F(\tilde{x}) + L_R \text{dist}_\Omega(x, \tilde{x})\mathbb{B}, \quad \forall x, \tilde{x} \in \Omega \cap \mathbb{B}(0, R).$$

Then  $W_g$  is continuous and it is the unique viscosity solution to (25)-(26).

**Remark 6.5.** We say that (27) is an absorbing property at the boundary of  $\Omega$  because we can extend the function  $\omega(x) = \frac{1 + |\nabla g(x)|}{|\nabla^2 g(x)|}$  in a continuous way up to  $\bar{\Omega}$  by setting  $\omega(x) = 0$  whenever  $x \in \partial\Omega$  and so  $F(x) = \{0\}$  on  $\partial\Omega$ . The extension is justified by the Fundamental Theorem of Calculus.

*Proof.* Since (20) holds, we can check that  $\Phi_g(\cdot)$  has nonempty convex compact images on  $\mathbb{R}^n$ . By the absorbing property (27), we have that  $\Phi_g(\cdot)$  has linear growth on  $\mathbb{R}^n$  and by (28) we have that it is locally Lipschitz. So by [14, Proposition 4.7.10] we get the desired result.  $\square$

**Theorem 6.1.** Suppose the same assumptions as in Proposition 6.3 with  $f$  uniformly continuous. Then  $V$  is uniformly continuous on  $\Omega$  and it is the unique viscosity solution of

$$V_\tau + \min_{v \in F(x)} \langle v, V_x \rangle = 0, \quad (\tau, x) \in (0, T) \times \Omega, \\ V(T, x) = f(x), \quad x \in \Omega.$$

*Proof.* First of all, note that  $(H_1)$  holds. Indeed, since under our assumptions  $\Phi_g(\cdot)$  has nonempty convex compact images on  $\mathbb{R}^n$ , it has linear growth and is locally Lipschitz, the set of trajectories  $S^{\Phi_g}(\tau, \nabla g(x))$  is nonempty for every  $(t, x) \in [0, T] \times \Omega$ . Let  $Y_0(\cdot) \in S^{\Phi_g}(\tau, \nabla g(x))$ , notice that

$$X_0(\cdot) = \nabla g^* \circ Y_0(\cdot) \in S_\Omega^F(\tau, x) \quad \text{and} \quad X_0(t) \in \Omega, \quad \forall t \in [\tau, T].$$

Let  $\varepsilon > 0$  and  $X(\cdot) \in S_\Omega^F(\tau, x)$ . Let  $\sigma = |X_0(T) - X(T)|$  and assume first that  $\sigma \geq \varepsilon$ . Consider  $X_\varepsilon : [\tau, T] \rightarrow \mathbb{R}^n$  given by

$$X_\varepsilon(t) = \frac{\varepsilon}{\sigma} X_0(t) + \frac{1 - \varepsilon}{\sigma} X(t), \quad \forall t \in [\tau, T].$$

Notice that  $X_\varepsilon$  is an absolutely continuous arc that dwells in  $\Omega$ . Furthermore, since  $F$  is convex-valued  $X_\varepsilon \in S_\Omega^F(\tau, x)$  and by construction,  $|X_\varepsilon(T) - X(T)| \leq \varepsilon$ , so, as stated earlier,  $(H_1)$  holds.

If  $\sigma < \varepsilon$  it is enough to take  $X_\varepsilon = X_0$ . Therefore,  $(H_1)$  holds as well.

By Proposition 6.2 we have that  $V(\tau, x) = W_g(\tau, \nabla g(x))$  on  $[0, T] \times \Omega$ . Furthermore, by Proposition 4.2 we get that  $(\theta, \nabla g(x)\zeta) \in \partial_D V(\tau, x)$  if and only if  $(\theta, \zeta) \in \partial_D W_g(\tau, \nabla g(x))$ . A similar relation also holds for the

viscosity superdifferential of  $V$  and  $W_g$ . Hence, the characterization of the Value Functions is a direct consequence of Proposition 6.3.  $\square$

**Acknowledgements.** The author would like to thanks to professor F. Álvarez for having introduced him into the topic of Squared Hessian Riemannian metrics and to professors J. Bolte and H. Zidani for their comments about this manuscript.

#### REFERENCES

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.
- [2] F. Álvarez, J. Bolte, and O. Brahic. Hessian riemannian gradient flows in convex programming. *SIAM J. Control and Optimization*, 43(2):477–501, 2004.
- [3] J.-P. Aubin and A. Cellina. *Differential inclusions: set-valued maps and viability theory*. Springer-Verlag New York, Inc., 1984.
- [4] A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization*, 16(3):697–725, 2006.
- [5] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Systems & Control: Foundations & Applications. Birkhäuser Boston Inc., Boston, MA, 1997. With appendices by Maurizio Falcone and Pierpaolo Soravia.
- [6] D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming. I. Affine and projective scaling trajectories. *Trans. Amer. Math. Soc.*, 314(2):499–526, 1989.
- [7] P. Bettiol, H. Frankowska, et al. Lipschitz regularity of solution map of control systems with multiple state constraints. *Discrete and Continuous Dynamical Systems-Series A*, 32(1):1–26, 2012.
- [8] O. Bokanowski, N. Forcadel, and H. Zidani. Deterministic state-constrained optimal control problems without controllability assumptions. *ESAIM: Control, Optimisation and Calculus of Variations*, 17(04):995–1015, 2011.
- [9] J. M. Borwein and J. Vanderwerff. Convex functions of legendre type in general banach spaces. *Journal of Convex Analysis*, 8(2):569–582, 2001.
- [10] J. M. Borwein and J. D. Vanderwerff. *Convex functions: constructions, characterizations and counterexamples*, volume 109. Cambridge University Press, 2010.
- [11] U. Boscaïn and B. Piccoli. *Optimal syntheses for control systems on 2-D manifolds*, volume 43. Springer, 2004.
- [12] R. Burachik, L. M. G. Drummond, A. N. Iusem, and B. F. Svaiter. Full convergence of the steepest descent method with inexact line searches. *Optimization*, 32(2):137–146, 1995.
- [13] I. Chrysoschoos and R. Vinter. Optimal control problems on manifolds: a dynamic programming approach. *Journal of mathematical analysis and applications*, 287(1):118–140, 2003.
- [14] F. Clarke, Y. Ledyev, R. Stern, and P. Wolenski. *Nonsmooth Analysis and Control Theory (Graduate Texts in Mathematics)*. Springer, 1 edition, 1997.
- [15] F. Clarke and R. Stern. Hamilton-jacobi characterization of the state constrained value. *Nonlinear Analysis: Theory, Methods & Applications*, 61(5):725–734, 2005.
- [16] M. P. do Carmo. *Riemannian Geometry*. Birkhäuser Boston, 1992.
- [17] A. V. Fiacco. Perturbed variations of penalty function methods. Example: projective SUMT. *Ann. Oper. Res.*, 27(1-4):371–380, 1990.
- [18] H. Frankowska and M. Mazzola. Discontinuous solutions of hamilton–jacobi–bellman equation under state constraints. *Calculus of Variations and Partial Differential Equations*, 46(3-4):725–747, 2013.



- [19] H. Frankowska and R. B. Vinter. Existence of neighboring feasible trajectories: Applications to dynamic programming for state-constrained optimal control problems. *Journal of Optimization Theory and Applications*, 104(1):20–40, 2000.
- [20] C. Hermosilla and H. Zidani. Infinite horizon problems on stratifiable state constraints sets. <http://hal.inria.fr/hal-00955921>, 2014.
- [21] A. N. Iusem, B. F. Svaiter, and J. X. Da Cruz Neto. Central paths, generalized proximal point methods, and Cauchy trajectories in Riemannian manifolds. *SIAM J. Control Optim.*, 37(2):566–588 (electronic), 1999.
- [22] J. M. Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer, 1997.
- [23] G. P. McCormick. The projective SUMT method for convex programming. *Math. Oper. Res.*, 14(2):203–223, 1989.
- [24] Y. Nesterov and M. Todd. On the riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math*, 2(4):333–361, 2002.
- [25] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 2000.
- [26] E. P. Quiroz and P. R. Oliveira. Proximal point methods for quasiconvex and convex functions with bregman distances on hadamard manifolds. *Journal of Convex Analysis*, 16(1):49–69, 2009.
- [27] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [28] S. T. Smith. Optimization techniques on riemannian manifolds. *Fields institute communications*, 3(3):113–135, 1994.
- [29] H. Soner. Optimal control with state-space constraint I. *SIAM Journal on Control and Optimization*, 24(3):552–561, 1986.
- [30] C. Udriste. *Convex Function and Optimization Methods on Riemannian Manifolds*. Kluwer Academic Publishers, 1994.
- [31] R. J. Vanderbei. Linear programming. *Foundations and extensions, International Series in Operations Research & Management Science*, 37, 2001.
- [32] R. Vinter. *Optimal control*. Springer, 2010.
- [33] Y. Yang. Globally convergent optimization algorithms on riemannian manifolds: Uniform framework for unconstrained and constrained optimization. *Journal of Optimization Theory and Applications*, 132(2):245–265, 2007.

(C. Hermosilla) PROJECT COMMANDS INRIA SACLAY & ENSTA PARISTECH, 828,  
BOULEVARD DES MARÉCHAUX, 91762 PALAISEAU CEDEX, FRANCE

*E-mail address:* `crisopher.hermosilla@inria.fr`